ISSN 2411-7390

jle

JOURNAL OF LANGUAGE & EDUCATION

Volume 7 Issue 1, 2021

HIGHER SCHOOL OF ECONOMICS
NATIONAL RESEARCH UNIVERSITY

Contents

Editorial

Lilia Raitskaya, Elena Tikhonova
Education for Sustainable Development: Glocal Implications for Universities 4

Research Articles

Minoo Alemi, Hessameddin Ghanbar, Atefeh Rezanejad
The Integration of Verbal Humor into EFL Classrooms: The Issues of Appropriateness and Relevance in Focus 14

Hesham Suleiman Alyousef
The Status of Theme in Research Article Abstracts in Seven Dentistry Subdisciplines: A Text-Based Study of Intradisciplinary Variations and Similarities in Thematic Choices and Thematic Progression Patterns..... 28

Aslihan Bagci, Belgin Aydin
Autonomy Orientations of English Language Teachers in Turkey..... 46

Tanju Deveci, Nader Ayish
Qualitative Adjectives in EFL Students' Reflective Writing Essays 64

Zoe Gavriilidou, Lydia Mitits
The Socio-linguistic Profiles, Identities, and Educational Needs of Greek Heritage Language Speakers in Chicago..... 78

Tahereh Heydarnejad, Azar Hosseini Fatemi, Behzad Ghonsooly
The Relationship between Critical Thinking, Self-regulation, and Teaching Style Preferences among EFL Teachers: A Path Analysis Approach 96

Ali Isik
Exploring How ELT Teachers Perceive and Practice English Language Assessment 109

Mohsine Jebbour
Self-Disclosure and Moroccan EFL Learners' Writing Development: Effects on Complexity, Accuracy, and Fluency 127

Maya Lambovska, Daniela Todorova
'Publish and Flourish' instead of 'Publish or Perish': A Motivation Model for Top-quality Publications 141

Tatiana Pospelova
The Collaborative Discussion Model: Developing Writing Skills through Prewriting Discussion.....156

Marta Segura Molla, Helena Roquet, Carmen Pérez-Vidl
The Effects of a CLIL Programme on Linguistic Progress at Two Different Points in Time.....171

Frankie Subon, Norseha Unin
The Efficacies of an Explicit Vocabulary Instruction Model on ESL Learners' Vocabulary Size and Writing Skills 190

Vu Van Tuan
The Impact of Social Networking Sites on Study Habits and Interpersonal Relationships among Vietnamese Students 206

Opinion Article

Irfan Ahmed Rind
Becoming an ESL Researcher: A Personal Monologue.....219

Reviews

Prodhan Mahbub Ibna Seraj
A Systematic Overview of Issues for Developing EFL Learners' Oral English Communication Skills 229

Awatif Alruwaili
Corpus linguistics for education-A guide for research by Pascual Perez-Paredes: A Book Review 241

National Research University Higher School of Economics
Journal of Language & Education Volume 7, Issue 1, 2021

Raitskaya, L., & Tikhonova, E. (2021). Education for Sustainable Development: Glocal Implications for Universities. *Journal of Language and Education, 7*(1), 4-13. https://doi.org/10.17323/jle.2021.11215

Education for Sustainable Development: Glocal Implications for Universities

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

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

[2] National Research University Higher School of Economics

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

Correspondence concerning this article should be addressed to Elena Tikhonova, National Research University Higher School of Economics, 11 Pokrovsky Bulvar, Pokrovka Complex, room S313, Moscow, 109028, Russian Federation. E-mail: etihonova@hse.ru

The editorial overviews the key research aspects of embedding sustainable development into university systems worldwide. The JLE editors dwell upon the pivotal role of higher education in transferring knowledge, skills, and underlying values in promoting Sustainable Development Goal No.4 (Quality Education for Sustainable Development). The editorial analysis is underpinned by the most cited Scopus-indexed articles (Top-50 as of March 2021) on sustainable development in higher education. JLE potential authors will find some recommendations on the subject field gaps and key directions to be published in the journal upcoming issues.

Keywords: sustainable development, sustainability, higher education, competencies, curriculum, teaching and learning environment, glocality

Introduction

Sustainability entered the global agenda in the 1970s with the UN Commission on Environment and Development report headlined "Our Common Future"[1]. The latter defined sustainable development as the process meeting "the needs of the present without compromising the ability of future generations to meet their own needs"[2].

The Millennium Development Goals (MDGs) were adopted in 2000. The eight goals were to be attained by 2015. All the UN member states committed to support the MDGs. Goal Two had universal primary education, including enrolment and completion for boys and girls, as its key target. In 2015, the UN Resolution headlined "Transforming Our World: the 2030 Agenda for Sustainable Development" (widely known as Agenda 2030) was passed. It set forth seventeen Sustainable Development Goals (SDGs). The goals were worded quite broadly to cover the global vital needs in economic, social, and environmental spheres. A special system of targets and indicators was developed to monitor the progress towards the SDGs.

Goal 4 – Quality Education – provides for "inclusive and equitable education" and "lifelong learning opportunities" for all[3]. Education for Sustainable Development (ESD) encompasses "economic, social and environmental dimensions of the formal and informal curricula"[4]. Though much of ESD is focused on primary and secondary education, the role of higher education is being reconsidered and highly appreciated. Since 2015, sustainable development and sustainability have been thoroughly discussed across economic, social, political, and environmental strata. The sustainability concepts were transformed into numerous practical models. At present, the prevailing view sticks to the three pillars of sustainable development, with education considered as raising awareness, generating transformation in mindsets. It 'aims to facilitate the development of skills to contribute to a more sustainable future" (Avila et al., 2018, p. 109). They all are addressed at both global and local levels, resulting in the glocality of the sustainable development as a phenomenon

Education is relevant to the other SDGs, as it prepares young people to live and work in the environment heavily relied on sustainable development principles, including the key competencies relating to SD (Systems

[1] Report of the World Commission on Environment and development: Our common future. https://sustainabledevelopment.un.org/content/documents/5987our-common-future.pdf

[2] World Commission on Environment and Development (WCED). (1991). *Our common future* (2nd ed.). Rio de Getúlio Vargas Foundation. P. 42.

[3] 4: Quality Education. http://www.undp.org/content/undp/en/home/sustainable-development-goals/goal-4-quality-education.html

[4] Advance HE. Education for Sustainable Development. https://www.advance-he.ac.uk/knowledge-hub/education-sustainable-development-0

thinking, Anticipatory, Normative, Strategic, Collaboration, Critical thinking, Self-Awareness, and Integrated problem-solving)[5].

As the JLE scope in education is mainly limited to higher education, we aim to scrutinize the prevailing research trends in the field of higher education for sustainable development. To complete the task, a review of the most cited articles indexed with the Scopus database was carried out to single out the top 50 most cited publications between 2015 (the adoption of Agenda 2030) and 2020. Thus, the research question in the review can be confined to the following: What are the prevailing trends within the field of higher education for sustainable development?

Materials and Methods

Rationale for the Review Criteria

To estimate the hot topics and find out the prevailing trends in the field of higher education for sustainable development, we focused of the publications indexed with the Scopus database. The Scopus database was selected as it covers the most influential sources on Sustainable Development. The search within the base brought authoritative results (publications with citations in high-ranking journals).

As Agenda 2030 was adopted in 2015, the review period spans from 2015 to 2020. The following inclusion/exclusion criteria were fixed to filter the publications (see Table 1).

Table 1
Review Inclusion and Exclusion Criteria

Criteria	Inclusion	Exclusion
Time Span	2015-2020	Beyond the stated period
Document Type	Article	All other types, but for articles
Keywords	Education for sustainable development Higher education for sustainable development Higher education institutions (HEIs) Sustainable development goals (SDGs) Education policy Lifelong learning	All other keywords
Research Area	Social Sciences	Sciences and research fields beyond social sciences
Languages	English	Other languages

The 50 most highly cited articles on higher education for sustainable development were sifted. The complete list forms Appendix 1 (See Appendix A).

Procedure

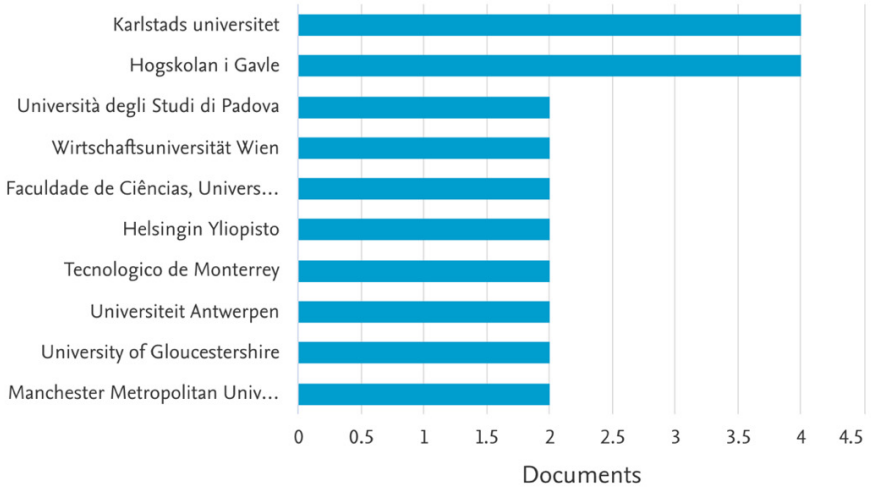
The initial search with "education for sustainable development" in the field covering titles, abstracts and keywords in the Scopus database, accessed March 1, 2021 and evaluated on the basis of the keyword inclusion criteria brought 24,177 documents. 10,223 indexed publications belonged to Social Sciences, partly overlapped by 1,234 in Economics, Econometrics and Finance, and 6,197 in Environmental Science. The search results were refined through the inclusion criteria. Then we singled out the first 50 most highly cited articles, with the article headlined "Connecting competences and pedagogical approaches for sustainable development in higher education: A literature review and framework proposal" (Lozano, Merrill, Sammalisto, Ceulemans, & Lozano, 2017) topping the search. The highest citation among the reviewed publications was recorded as high as 127 as of March 15, 2021. The lowest citation in the list reached 18. It was "Neoliberalism, pluralism and environmental education: The call for radical re-orientation" (Kopnina, 2015).

[5] UNESCO. 2017. Education for sustainable development goals learning objectives. United Nations Educational, Scientific and Cultural Organization. https://www.unesco.de/sites/default/files/2018-08/unesco_education_for_sustainable_development_goals.pdf

The bibliometric analysis of the 50 documents showed that the selected publications were distributed unevenly, with 14 in 2015; 8 in 2016; 11 in 2017; 9 in 2018; 8 in 2019. No publication reached comparatively high readings in 2020. The articles came out in the following sources: *Sustainability (Switzerland)* – 14 articles; *Environmental Education Research* – 9 articles; *International Journal of Management* – 3 articles; another three journals with two publications each; eleven journals with one article each.

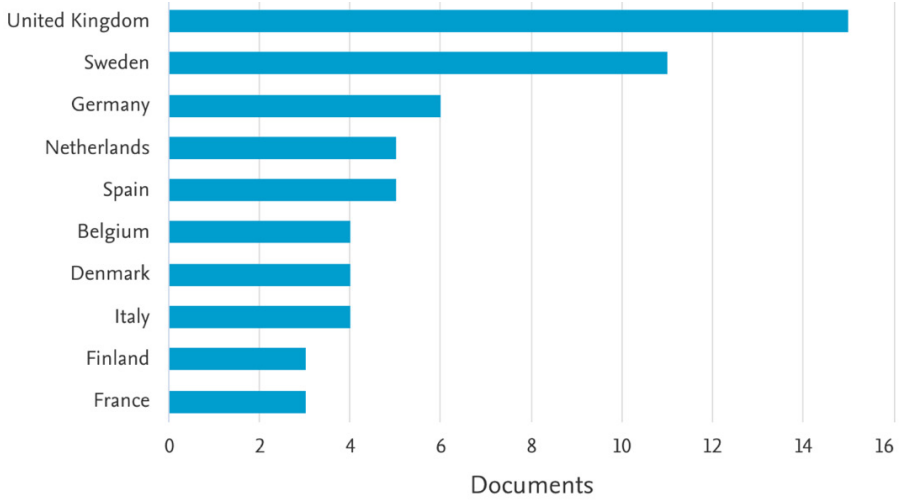
The research area was limited to social sciences (n=50, 100 %), though some of the articles in the review have more than one subject area: social science & environmental science (n=21); social science & energy (n=14); social sciences & business, management and accounting (n=4); social sciences & economics, econometrics and finance (n=4); social sciences & agricultural and biological sciences (n=1); social sciences & engineering (n=1); social sciences & mathematics (n=1). The most prolific authors in Top-50 are Gericke, N. (4 articles); Kopnina, H. (3 articles); Lozano, R. (3 articles); Olsson, D. (3 articles); Sammalisto, K. (3 articles).

Figure 1
Scopus Indexed Research on Higher Education for Sustainable Development (Social Sciences): Affiliation Breakdown. Source: Scopus Database as of March 1, 2021



Two universities were affiliation leaders, with four articles each (See Figure 1). They are Karlstads universitet and Högskolan i Gävle. Both are located in Sweden. As for the aggregate number of the articles in the top list, the highest scores belong to the UK (15 articles) and Sweden (11 articles) by a large margin (See Figure 2).

Figure 2
Scopus Indexed Research on Higher Education for Sustainable Development (Social Sciences): Leading Countries. Source: Scopus Database as of March 1, 2021



Results and Discussion

The thematic clusters within the field of higher education for sustainable development boiled down to eight: general issues of ESD; country-specific experiences; higher professional education; skills and competencies; attitudes and perceptions; assessment; curriculum; neoliberal aspects and ESD (See Table 2). The breakdown of the articles does not exclude their overlapping when their topic is complex.

Table 2
Thematic Clusters within the Field of Higher Education for Sustainable Development

Thematic Cluster	Number of Publications in Top 50	Cluster Description
General issues of ESD	10	Mechanisms of planning and implementing the SDGs, future prospects and expectations, effectiveness of SDG programmes, a framework for connecting the SDGs to educational outcomes, professional development initiatives for university educators to address the SD challenges
Country-specific experiences	8	The research of cases specific of a particular country (Portugal, Sweden, Canada, Malaysia, Italy, Germany, Spain, UK)
Higher professional education	8	This cluster covers specialized education (teacher education, engineering, science and environmental education, management education) in its efforts to promote sustainable development, incorporate the SD principles in their curriculum and learning/ teaching across disciplines
Skills and competencies	9	Pedagogical approaches in designing and delivering sustainability-oriented competencies, connections between pedagogical approaches and development of sustainability competencies, transformative learning
Attitudes and perceptions	8	Practical approaches to advancing education for sustainable development, hope and anticipation relating to ESD
Assessment	6	Assessment tools within the sustainability framework, assessment indicators to measure, audit, and benchmark efforts at embedding SD
Curriculum	5	Approaches and hurdles of SD integration into a university curriculum
Neoliberal aspects and ESD	3	The threat of neoliberal values to sustainability models in education

General Issues of ESD

Though Agenda 2030 was adopted over five years ago, there are still many research articles on the ESD foundations, on the ways the SD principles are incorporated into higher education systems. Researchers study the SD perspectives for higher education (Filho, Manolas & Pace, 2015; Sonetti, Brown & Naboni, 2019), sustainable educational policy (Van Poeck & Lysgaard, 2016); coherence between theory and praxis of ESD (Amador, Martinho, Bacelar-Nicolau, Caeiro & Oliveira, 2015); professional development initiatives for educators to efficiently embed SD into higher education (Mulà et al., 2017).

Country-Specific Experiences

As all UN member states are committed to Agenda 2030, each has its own experience in ESD. Top-50 encompasses several country-specific publications. They focus on the ESD implementation progress in Portugal (Aleixo, Azeiteiro & Leal, 2018); gender-related aspects of students' sustainability consciousness in Sweden (Olsson & Gericke, 2017); the national higher education sustainability policies in Canada (Lidstone, Wright & Sherren, 2015), a quantitative SD scale at an Italian university (Biasutti & Frate, 2017), etc.

Higher Professional Education

SD is relevant to most of spheres of human activities. But some professions are at the forefront, having more relevance and responsibility. Thus, teacher education (Álvarez-García, Sureda-Negre & Comas-Forgas, 2015), engineering education (Guerra, 2017), science (Eilks, 2015) and environmental education (Kopnina, 2015) can

promote SD not only within their segment of education, but via their graduates widely spread the SD values across their professional communities.

Skills and Competencies

The research on skills and competencies have been in focus since SDG No.4 is in place. In 2017, special skills were set forth as vital in attaining the SDGs (UNESCO). As a result, much attention is paid to studies on the pedagogical design and delivery of sustainability-oriented competencies (Lozano, Barreiro-Gen, Lozano & Sammalisto, 2019; Lozano, Merrill, Sammalisto, Ceulemans & Lozano, 2017; Cebrián & Junyent, 2015), critical thinking (Straková & Cimermanová, 2018; Hasslöf & Malmberg, 2015), and systems thinking (Waltner, Rieß & Mischo, 2019). Though many publications on competencies come out annually, the considered papers focus on the competencies that must transform the mindsets of the younger generations to efficiently address the sustainability challenges.

Attitudes and Perceptions

Some of the articles in this cluster enter into the other clusters, too. The publications highlight students', teachers', and educators' perceptions of sustainable development, attitudes to pedagogical approaches to fostering skills and competencies, hope and anticipation in ESD (Ojala, 2017; Berglund & Gericke, 2016; Olsson & Gericke, 2016).

Assessment

Assessment arose as separate sub-field in response to the pressing need to monitor SD implementation, including attaining targets within Goal 4. Indicators are being worked out to gauge the progress. The articles in this cluster encompass studies of sustainability assessment tools (Fischer, Jenssen & Tappeser, 2015); assessment of the impacts of higher education institutions on sustainable development (Findler, Schönherr, Lozano & Stacherl, 2018); assessment of sustainability literacy (Décamps, Barbat, Carteron, Hands & Parkes, 2017), etc.

Curriculum

Embedding SD into university curricula implies an all-round system of the SD implementation. The studies concentrate on curriculum change (Kolmos, Hadgraft & Holgaard, 2016), the project method in integrating SD into higher education curricula (Fuertes-Camacho, Graell-Martín, Fuentes-Loss & Balaguer-Fàbregas, 2019); SD curricula in degree programmes (Wyness & Sterling, 2015). Curricula change algorithm is not universal, as in some countries there are institutional hurdles. Curricula cannot easily be changed.

Neoliberal Aspects and ES

The neoliberal values are in contrast with the SD principles. Higher education in many countries is inclined to neoliberal philosophy. Students may absorb controversial sets of values. Thus, their commitment to SD may be considerable distorted (Kopnina & Cherniak, 2016; Kopnina, 2015).

Conclusion

The review analysis displays the eight key research trends in the subject area of ESD: general issues of ESD; country-specific experiences; higher professional education; skills and competencies; attitudes and perceptions; assessment; curriculum; neoliberal aspects and ESD, with a greater focus on all aspects of practical models of ESD across countries and sciences. The reviewed publications prove that all trends emerge and evolve glocally.

With its commitment to tertiary education, the JLE is going to bring out more articles and reviews in the light of the SDGs and ESD. Research on pedagogical approaches to efficient fostering skills and competencies for sustainability is certain to be prioritised. Studies of glocal models of imbedding sustainability in university curriculum and academic systems at large will also be highly appreciated.

References

Aleixo, A. M., Azeiteiro, U., & Leal, S. (2018). The implementation of sustainability practices in Portuguese higher education institutions. *International Journal of Sustainability in Higher Education, 19*(1), 146-178. https://doi.org/10.1108/IJSHE-02-2017-0016

Álvarez-García, O., Sureda-Negre, J., & Comas-Forgas, R. (2015). Environmental education in pre-service teacher training: A literature review of existing evidence. *Journal of Teacher Education for Sustainability, 17*(1), 72-85. https://doi.org/10.1515/jtes-2015-0006

Amador, F., Martinho, A. P., Bacelar-Nicolau, P., Caeiro, S., & Oliveira, C. P. (2015). Education for sustainable development in higher education: Evaluating coherence between theory and praxis. *Assessment and Evaluation in Higher Education, 40*(6), 867-882. https://doi.org/10.1080/02602938.2015.1054783

Avila, V.L., Facco, A.L.R., Bento, M.H.S., Arigony, M.M., Obregon, S.L., & Trevisan, M. (2018). Sustainability and Education for Sustainability: An Analysis of Publications from the Last Decade. *Environmental Quality Management*, 27, 107-118. https://doi.org/10.1002/tqem.21537

Berglund, T., & Gericke, N. (2016). Separated and integrated perspectives on environmental, economic, and social dimensions – an investigation of student views on sustainable development. *Environmental Education Research, 22*(8), 1115-1138. https://doi.org/10.1080/13504622.2015.1063589

Biasutti, M., & Frate, S. (2017). A validity and reliability study of the attitudes toward sustainable development scale. *Environmental Education Research, 23*(2), 214-230. https://doi.org/10.1080/13504622.2016.1146660

Cebrián, G., & Junyent, M. (2015). Competencies in education for sustainable development: Exploring the student teachers' views. *Sustainability, 7*(3), 2768-2786. https://doi.org/10.3390/su7032768

Décamps, A., Barbat, G., Carteron, J., Hands, V., & Parkes, C. (2017). Sulitest: A collaborative initiative to support and assess sustainability literacy in higher education. *International Journal of Management Education, 15*(2), 138-152. https://doi.org/10.1016/j.ijme.2017.02.006

Eilks, I. (2015). Science education and education for sustainable development - justifications, models, practices and perspectives. *Eurasia Journal of Mathematics, Science and Technology Education, 11*(1), 149-158. https://doi.org/10.12973/eurasia.2015.1313a

Filho, W. L., Manolas, E., & Pace, P. (2015). The future we want key issues on sustainable development in higher education after Rio and the un decade of education for sustainable development. *International Journal of Sustainability in Higher Education, 16*(1), 112-129. https://doi.org/10.1108/IJSHE-03-2014-0036

Findler, F., Schönherr, N., Lozano, R., & Stacherl, B. (2018). Assessing the impacts of higher education institutions on sustainable development-an analysis of tools and indicators. *Sustainability, 11*(1), 59. https://doi.org/10.3390/su11010059

Fuertes-Camacho, M. T., Graell-Martín, M., Fuentes-Loss, M., & Balaguer-Fàbregas, M. C. (2019). Integrating sustainability into higher education curricula through the project method, a global learning strategy. *Sustainability, 11*(3), 767. https://doi.org/10.3390/su11030767

Guerra, A. (2017). Integration of sustainability in engineering education: Why is PBL an answer? *International Journal of Sustainability in Higher Education, 18*(3), 436-454. https://doi.org/10.1108/IJSHE-02-2016-0022

Hasslöf, H., & Malmberg, C. (2015). Critical thinking as room for subjectification in education for sustainable development. *Environmental Education Research, 21*(2), 239-255. https://doi.org/10.1080/13504622.2014.940854

Kolmos, A., Hadgraft, R. G., & Holgaard, J. E. (2016). Response strategies for curriculum change in engineering. *International Journal of Technology and Design Education, 26*(3), 391-411. https://doi.org/10.1007/s10798-015-9319-y

Kopnina, H. (2015). Neoliberalism, pluralism and environmental education: The call for radical re-orientation. *Environmental Development, 15*, 120-130. https://doi.org/10.1016/j.envdev.2015.03.005

Kopnina, H. (2015). Sustainability in environmental education: New strategic thinking. *Environment, Development and Sustainability, 17*(5), 987-1002. https://doi.org/10.1007/s10668-014-9584-z

Kopnina, H., & Cherniak, B. (2016). Neoliberalism and justice in education for sustainable development: A call for inclusive pluralism. *Environmental Education Research, 22*(6), 827-841. https://doi.org/10.1080/13504622.2016.1149550

Lidstone, L., Wright, T., & Sherren, K. (2015). An analysis of Canadian STARS-rated higher education sustainability policies. *Environment, Development and Sustainability, 17*(2), 259-278. https://doi.org/10.1007/s10668-014-9598-6

Lozano, R., Barreiro-Gen, M., Lozano, F. J., & Sammalisto, K. (2019). Teaching sustainability in European higher education institutions: Assessing the connections between competences and pedagogical approaches. *Sustainability, 11*(6), 1602. https://doi.org/10.3390/su11061602

Lozano, R., Merrill, M. Y., Sammalisto, K., Ceulemans, K., & Lozano, F. J. (2017). Connecting competences and pedagogical approaches for sustainable development in higher education: A literature review and framework proposal. *Sustainability, 9*(10), 1889. https://doi.org/10.3390/su9101889

Mulà, I., Tilbury, D., Ryan, A., Mader, M., Dlouhá, J., Mader, C., Benayas, J., Dlouhý, J., Alba, D. (2017). Catalysing change in higher education for sustainable development: A review of professional development initiatives for university educators. *International Journal of Sustainability in Higher Education, 18*(5), 798-820. https://doi.org/10.1108/IJSHE-03-2017-0043

Ojala, M. (2017). Hope and anticipation in education for a sustainable future. *Futures, 94*, 76-84. https://doi.org/10.1016/j.futures.2016.10.004

Olsson, D., & Gericke, N. (2016). The adolescent dip in students' sustainability consciousness - implications for education for sustainable development. *Journal of Environmental Education, 47*(1), 35-51. https://doi.org/10.1080/00958964.2015.1075464

Olsson, D., & Gericke, N. (2017). The effect of gender on students' sustainability consciousness: A nationwide Swedish study. *Journal of Environmental Education, 48*(5), 357-370. https://doi.org/https://doi.org/10.1080/00958964.2017.1310083

Sonetti, G., Brown, M., & Naboni, E. (2019). About the triggering of UN sustainable development goals and regenerative sustainability in higher education. *Sustainability, 11*(1), 254. https://doi.org/10.3390/su11010254

Straková, Z., & Cimermanová, I. (2018). Critical thinking development-a necessary step in higher education transformation towards sustainability. *Sustainability (Switzerland), 10*(10). https://doi.org/10.3390/su10103366

Van Poeck, K., & Lysgaard, J. A. (2016). The roots and routes of environmental and sustainability education policy research. *Environmental Education Research, 22*(3), 305-318. https://doi.org/10.1080/13504622.2015.1108393

Waltner, E., Rieß, W., & Mischo, C. (2019). Development and validation of an instrument for measuring student sustainability competencies. *Sustainability (Switzerland), 11*(6). https://doi.org/10.3390/su11061717

Wyness, L., & Sterling, S. (2015). Reviewing the incidence and status of sustainability in degree programmes at Plymouth university. *International Journal of Sustainability in Higher Education, 16*(2), 237-250. https://doi.org/10.1108/IJSHE-09-2013-0112

Appendix A

Top-50 Highly Cited Articles on Higher Education for Sustainable Development

Albareda-Tiana, S., Vidal-Raméntol, S., & Fernández-Morilla, M. (2018). Implementing the sustainable development goals at university level. *International Journal of Sustainability in Higher Education, 19*(3), 473-497. https://doi.org/10.1108/IJSHE-05-2017-0069

Aleixo, A. M., Azeiteiro, U., & Leal, S. (2018). The implementation of sustainability practices in Portuguese higher education institutions. *International Journal of Sustainability in Higher Education, 19*(1), 146-178. https://doi.org/10.1108/IJSHE-02-2017-0016

Álvarez-García, O., Sureda-Negre, J., & Comas-Forgas, R. (2015). Environmental education in pre-service teacher training: A literature review of existing evidence. *Journal of Teacher Education for Sustainability, 17*(1), 72-85. https://doi.org/10.1515/jtes-2015-0006

Amador, F., Martinho, A. P., Bacelar-Nicolau, P., Caeiro, S., & Oliveira, C. P. (2015). Education for sustainable development in higher education: Evaluating coherence between theory and praxis. *Assessment and Evaluation in Higher Education, 40*(6), 867-882. doi:10.1080/02602938.2015.1054783

Annan-Diab, F., & Molinari, C. (2017). Interdisciplinarity: Practical approach to advancing education for sustainability and for the sustainable development goals. *International Journal of Management Education, 15*(2), 73-83. https://doi.org/10.1016/j.ijme.2017.03.006

Berglund, T., & Gericke, N. (2016). Separated and integrated perspectives on environmental, economic, and social dimensions – an investigation of student views on sustainable development. *Environmental Education Research, 22*(8), 1115-1138. https://doi.org/10.1080/13504622.2015.1063589

Bessant, S. E. F., Robinson, Z. P., & Ormerod, R. M. (2015). Neoliberalism, new public management and the sustainable development agenda of higher education: History, contradictions and synergies. *Environmental Education Research, 21*(3), 417-432. https://doi.org/10.1080/13504622.2014.993933

Biasutti, M. (2015). An intensive programme on education for sustainable development: The participants' experience. *Environmental Education Research, 21*(5), 734-752. https://doi.org/10.1080/13504622.2014.921805

Biasutti, M., & Frate, S. (2017). A validity and reliability study of the attitudes toward sustainable development scale. *Environmental Education Research, 23*(2), 214-230. https://doi.org/10.1080/13504622.2016.1146660

Borges, J. C., Ferreira, T. C., Borges de Oliveira, M. S., Macini, N., & Caldana, A. C. F. (2017). Hidden curriculum in student organizations: Learning, practice, socialization and responsible management in a business school. *International Journal of Management Education, 15*(2), 153-161. https://doi.org/10.1016/j.ijme.2017.03.003

Cebrián, G., & Junyent, M. (2015). Competencies in education for sustainable development: Exploring the student teachers' views. *Sustainability (Switzerland), 7*(3), 2768-2786. https://doi.org/10.3390/su7032768

Décamps, A., Barbat, G., Carteron, J., Hands, V., & Parkes, C. (2017). Sulitest: A collaborative initiative to support and assess sustainability literacy in higher education. *International Journal of Management Education, 15*(2), 138-152. https://doi.org/10.1016/j.ijme.2017.02.006

Eilks, I. (2015). Science education and education for sustainable development - justifications, models, practices and perspectives. *Eurasia Journal of Mathematics, Science and Technology Education, 11*(1), 149-158. https://doi.org/10.12973/eurasia.2015.1313a

Filho, W. L., Manolas, E., & Pace, P. (2015). The future we want key issues on sustainable development in higher education after Rio and the un decade of education for sustainable development. *International Journal of Sustainability in Higher Education, 16*(1), 112-129. https://doi.org/10.1108/IJSHE-03-2014-0036

Findler, F., Schönherr, N., Lozano, R., & Stacherl, B. (2018). Assessing the impacts of higher education institutions on sustainable development-an analysis of tools and indicators. *Sustainability, 11*(1), 59. https://doi.org/10.3390/su11010059

Fischer, D., Jenssen, S., & Tappeser, V. (2015). Getting an empirical hold of the sustainable university: A comparative analysis of evaluation frameworks across 12 contemporary sustainability assessment tools. *Assessment and Evaluation in Higher Education, 40*(6), 785-800. https://doi.org/10.1080/02602938.2015.1043234

Fuertes-Camacho, M. T., Graell-Martín, M., Fuentes-Loss, M., & Balaguer-Fàbregas, M. C. (2019). Integrating sustainability into higher education curricula through the project method, a global learning strategy. *Sustainability, 11*(3), 767. https://doi.org/10.3390/su11030767

Giangrande, N., White, R. M., East, M., Jackson, R., Clarke, T., Coste, M. S., & Penha-Lopes, G. (2019). A competency framework to assess and activate education for sustainable development: Addressing the UN sustainable development goals 4.7 challenge. *Sustainability, 11*(10), 2832. https://doi.org/10.3390/su11102832

Guerra, A. (2017). Integration of sustainability in engineering education: Why is PBL an answer? *International Journal of Sustainability in Higher Education, 18*(3), 436-454. https://doi.org/10.1108/IJSHE-02-2016-0022

Hartley, B. L., Pahl, S., Holland, M., Alampei, I., Veiga, J. M., & Thompson, R. C. (2018). Turning the tide on trash: Empowering european educators and school students to tackle marine litter. *Marine Policy, 96*, 227-234. https://doi.org/10.1016/j.marpol.2018.02.002

Hasslöf, H., & Malmberg, C. (2015). Critical thinking as room for subjectification in education for sustainable development. *Environmental Education Research, 21*(2), 239-255. https://doi.org/10.1080/13504622.2014.940854

Hoffmann, R., & Muttarak, R. (2017). Learn from the past, prepare for the future: Impacts of education and experience on disaster preparedness in the philippines and thailand. *World Development, 96*, 32-51. https://doi.org/10.1016/j.worlddev.2017.02.016

Kioupi, V., & Voulvoulis, N. (2019). Education for sustainable development: A systemic framework for connecting the SDGs to educational outcomes. *Sustainability, 11*(21), 6104. https://doi.org/10.3390/su11216104

Kolmos, A., Hadgraft, R. G., & Holgaard, J. E. (2016). Response strategies for curriculum change in engineering. *International Journal of Technology and Design Education, 26*(3), 391-411. https://doi.org/10.1007/s10798-015-9319-y

Kopnina, H. (2015). Neoliberalism, pluralism and environmental education: The call for radical re-orientation. *Environmental Development, 15*, 120-130. https://doi.org/10.1016/j.envdev.2015.03.005

Kopnina, H. (2015). Sustainability in environmental education: New strategic thinking. *Environment, Development and Sustainability, 17*(5), 987-1002. https://doi.org/10.1007/s10668-014-9584-z

Kopnina, H., & Cherniak, B. (2016). Neoliberalism and justice in education for sustainable development: A call for inclusive pluralism. *Environmental Education Research, 22*(6), 827-841. https://doi.org/10.1080/13504622.2016.1149550

Lambrechts, W., & Van Petegem, P. (2016). The interrelations between competences for sustainable development and research competences. *International Journal of Sustainability in Higher Education, 17*(6), 776-795. https://doi.org/10.1108/IJSHE-03-2015-0060

Leal Filho, W., Pallant, E., Enete, A., Richter, B., & Brandli, L. L. (2018). Planning and implementing sustainability in higher education institutions: An overview of the difficulties and potentials. *International Journal of Sustainable Development and World Ecology, 25*(8), 712-720. https://doi.org/10.1080/13504509.2018.1461707

Li, C., & Zhou, H. (2018). Enhancing the efficiency of massive online learning by integrating intelligent analysis into MOOCs with an application to education of sustainability. *Sustainability, 10*(2), 468. https://doi.org/10.3390/su10020468

Lidstone, L., Wright, T., & Sherren, K. (2015). An analysis of Canadian STARS-rated higher education sustainability policies. *Environment, Development and Sustainability, 17*(2), 259-278. https://doi.org/10.1007/s10668-014-9598-6

Lozano, R., Barreiro-Gen, M., Lozano, F. J., & Sammalisto, K. (2019). Teaching sustainability in European higher education institutions: Assessing the connections between competences and pedagogical approaches. *Sustainability, 11*(6), 1602. https://doi.org/10.3390/su11061602

Lozano, R., Merrill, M. Y., Sammalisto, K., Ceulemans, K., & Lozano, F. J. (2017). Connecting competences and pedagogical approaches for sustainable development in higher education: A literature review and framework proposal. *Sustainability, 9*(10), 1889. https://doi.org/10.3390/su9101889

Mulà, I., Tilbury, D., Ryan, A., Mader, M., Dlouhá, J., Mader, C., Benayas, J., Dlouhý, J., Alba, D. (2017). Catalysing change in higher education for sustainable development: A review of professional development initiatives for university educators. *International Journal of Sustainability in Higher Education, 18*(5), 798-820. https://doi.org/10.1108/IJSHE-03-2017-0043

O'Flaherty, J., & Liddy, M. (2018). The impact of development education and education for sustainable development interventions: A synthesis of the research. *Environmental Education Research, 24*(7), 1031-1049. https://doi.org/10.1080/13504622.2017.1392484

Ojala, M. (2017). Hope and anticipation in education for a sustainable future. *Futures, 94*, 76-84. https://doi.org/10.1016/j.futures.2016.10.004

Olsson, D., & Gericke, N. (2016). The adolescent dip in students' sustainability consciousness - implications for education for sustainable development. *Journal of Environmental Education, 47*(1), 35-51. https://doi.org/10.1080/00958964.2015.1075464

Olsson, D., & Gericke, N. (2017). The effect of gender on students' sustainability consciousness: A nationwide Swedish study. *Journal of Environmental Education, 48*(5), 357-370. https://doi.org/10.1080/00958964.2017.1310083

Pauw, J. B., Gericke, N., Olsson, D., & Berglund, T. (2015). The effectiveness of education for sustainable development. *Sustainability, 7*(11), 15693-15717. https://doi.org/10.3390/su71115693

Pihkala, P. (2017). Environmental education after sustainability: Hope in the midst of tragedy. *Global Discourse, 7*(1), 109-127. https://doi.org/10.1080/23269995.2017.1300412

Sammalisto, K., Sundström, A., Von Haartman, R., Holm, T., & Yao, Z. (2016). Learning about sustainability-what influences students' self-perceived sustainability actions after undergraduate education? *Sustainability, 8*(6), 510. https://doi.org/10.3390/su8060510

Schuler, S., Fanta, D., Rosenkraenzer, F., & Riess, W. (2018). Systems thinking within the scope of education for sustainable development (ESD)–a heuristic competence model as a basis for (science) teacher education. *Journal of Geography in Higher Education, 42*(2), 192-204. https://doi.org/10.1080/03098265.2017.1339264

Sonetti, G., Brown, M., & Naboni, E. (2019). About the triggering of UN sustainable development goals and regenerative sustainability in higher education. *Sustainability, 11*(1), 254. https://doi.org/10.3390/su11010254

Straková, Z., & Cimermanová, I. (2018). Critical thinking development-a necessary step in higher education transformation towards sustainability. *Sustainability, 10*(10), 3366. https://doi.org/10.3390/su10103366

Van Poeck, K., & Lysgaard, J. A. (2016). The roots and routes of environmental and sustainability education policy research. *Environmental Education Research, 22*(3), 305-318. https://doi.org/10.1080/13504622.2015.1108393

Vare, P., Arro, G., de Hamer, A., Gobbo, G. D., de Vries, G., Farioli, F., Kadji-Beltran, Ch., Kangur, M., Mayer, M., Millican, R., Nijdam, C., Réti, M., Zachariou, A. (2019). Devising a competence-based training program for educators of sustainable development: Lessons learned. *Sustainability, 11*(7), 1890. https://doi.org/10.3390/su11071890

Vesterinen, V. -., Tolppanen, S., & Aksela, M. (2016). Toward citizenship science education: What students do to make the world a better place? *International Journal of Science Education, 38*(1), 30-50. https://doi.org/10.1080/09500693.2015.1125035

Waldron, F., Ruane, B., Oberman, R., & Morris, S. (2019). Geographical process or global injustice? contrasting educational perspectives on climate change. *Environmental Education Research, 25*(6), 895-911. https://doi.org/10.1080/13504622.2016.1255876

Waltner, E., Rieß, W., & Mischo, C. (2019). Development and validation of an instrument for measuring student sustainability competencies. *Sustainability, 11*(6), 1717. https://doi.org/10.3390/su11061717

Wyness, L., & Sterling, S. (2015). Reviewing the incidence and status of sustainability in degree programmes at Plymouth university. *International Journal of Sustainability in Higher Education, 16*(2), 237-250. https://doi.org/10.1108/IJSHE-09-2013-0112

National Research University Higher School of Economics
Journal of Language & Education Volume 7, Issue 1, 2021

Alemi, M., Ghanbar, H., & Rezanejad, A. (2021). The Integration of Verbal Humor into EFL Classrooms: The Issues of Appropriateness and Relevance in Focus. *Journal of Language and Education, 7*(1), 14-27. https://doi.org/10.17323/jle.2021.10861

The Integration of Verbal Humor into EFL Classrooms: The Issues of Appropriateness and Relevance in Focus

Minoo Alemi[1], Hessameddin Ghanbar[2], Atefeh Rezanejad[3]

[1] Islamic Azad University, West Tehran Branch

[2] Islamic Azad University, Fereshtegan International Branch

[3] Allameh Tabataba'i University

Correspondence concerning this article should be addressed to Minoo Alemi, Applied Linguistics, Islamic Azad University, West Tehran Branch, Dadman Blv, Derakhti St, Saghafi St, No 16, Terhran, Iran.
E-mail: minooalemi2000@yahoo.com

The present study examined the manifestations of Iranian male and female EFL teachers' use of humor in the classroom environment. To this end, a qualitative study with 30 participants was implemented in two English language institutes in Iran. Equally, 15 male and female EFL teachers were selected by convenience sampling and their classes were audio-recorded and later transcribed for the examination of the types of humor they used and their frequency. Wanzer, Frymier, Wojtaszcyk, and Smith's (2006) method of humor analysis and categorization of appropriateness was exploited for the analysis of the types of humor collected from the participants of the study. The results suggested that the use of humor by male teachers was more frequent than that of female teachers. It was revealed that 57% of the humor production was by male EFL teachers and 43% was produced by female EFL teachers. The results revealed that the most frequent humor type in male teachers' classrooms was "funny comments" (27%), with "teasing students" (3%) being the least frequent one. In the case of appropriate humor use, similarly, female teachers used "funny comments" (52%) as the most frequent one, while there was no instance of "providing humorous examples". Considering inappropriate humor use, both male and female teachers used "funny comments" (45%) as the most frequent type. The findings of the present study can be of use to EFL teachers and suggests the need for workshops and training courses on the integration of humor into EFL classes.

Keywords: humor, appropriate humor, inappropriate humor, classroom interaction analysis

Introduction

The topic of humor and its impact on the performance of language learners has gained considerable attention during the past two decades (Heidari-Shahreza & Heydari, 2018). According to Heidari-Shahreza (2020), humor and language are both conceptually and practically interrelated and "tongueless humor is usually a helpless one" (p. 81). Humor or language play is believed to alleviate learners' anxiety (Booth-Butterfield & Booth-Butterfield, 1991), increase self-confidence and motivation to learn (Gorham & Christophel, 1992), help learners shape their own identity (Pomerantz & Bell, 2011), and lead to joyful class atmospheres (Bell, 2012; Carter & McCarthy, 2004; Cook, 2000; Heidari-Shahreza & Heydari, 2018; Wagner & Urios-Aparisi, 2011). As Heidari-Shahreza (2020, p. 79) posits, "humor-integrated language learning (HILL)" is flourishing in the domain of TESOL (Teaching English to Speakers of Other Languages) and different scholars have addressed this issue from different vantage points.

Different types of humor used by teachers have been investigated in terms of appropriate and inappropriate humor (Wanzer & Frymier, 1999). Additionally, there is a difference between appropriate and inappropriate humor. Appropriate humor leads to higher teacher evaluations, while inappropriate humor and sarcasm lead to lower teacher evaluations and can negatively impact student learning. Basically, humor constructs classroom cohesion, leads to more students' positive interaction, reduces criticism, and removes individual's stress (Banas, Dunbar, Rodriguez, & Liu, 2011). On the whole, entangling a complex and complicated construct like humor can facilitate teaching, and in particular, interaction in EFL classrooms. Against this backdrop, the present

study aimed to investigate the frequency and appropriateness of verbal humor used by Iranian non-native English language teachers.

Literature Review

Humor

The literature on humor suggests that different scholars have defined the term differently and there is no consensus on the preferred definition. According to Wagner and Urios-Aparisi (2011, p. 400), humor in language classrooms can be defined as "an act performed through linguistic or non-linguistic means by any of the participants (i.e., student(s) or teacher)". Moreover, as Bell (2011) argues, humor is a "specific communicative mode in which something is uttered with the intent to amuse" (p. 238). Attardo and Raskin (1991) posit that humor is instantiated by verbal jokes. Moreover, Wanzer, et al., (2006) categorized appropriate humor in the classroom as: (a) related humor, (b) humor unrelated to course material, (c) self-disparaging humor, and (d) unintentional humor. Although various definitions have been proposed with regard to the concept of humor, a leading similar observed feature remains to be a matter of *incongruity,* which finally leads to some joyful moments (Martin, 2007). In the same vein, Schmidt and Williams (2001) believed that incongruity in teaching materials can lead to "sustained attention and subsequent elaborative processes" (p. 311).

However, apart from the establishment of fun and amusement, other potential advantages of humor have been noticed in the educational realm (Berk, 2000; Heidari-Shahreza, 2018; Lovorn & Holaway, 2015; Mantooth, 2010; Wanzer & Frymier, 1999; Wanzer, Frymier, & Irwin, 2010). For instance, it has been argued (Garner, 2006; Wanzer, Frymier, & Irwin, 2010) that humor can improve and develop the learning outcome. According to Gönülal (2018), humor is regarded as a useful tool in developing a positive classroom atmosphere and bringing more fun into the learning context. Berk (2000) also reported cases where humor could decrease test anxiety among the learners. There is also a positive connection between teachers' use of humor and academic

achievement for students (Hickman & Crossland, 2004-2005). It might also be an effective relaxing, comforting, and stress-reducing device that can enhance the students' interest and enjoyment (Neuliep, 1991). While humor, as a tool, can be vital for the students' learning, its unbalanced use can be jeopardizing (Azizinezhad & Hashemi, 2011; Bell, 2009). Several empirical studies (e.g., Deneire, 1995; Schmitz, 2002; Wagner & Urios-Aparisi, 2011) have lent support to the advantages and ease that humor can bring to the classroom. According to Schmitz (2002), the appropriate use of humor in the classroom can lead to giving students ease, getting students' attention, and having a less formal classroom.

Gender and Humor

Humor has many functions and it may be viewed differently in various social contexts. Meanwhile, the important role of humor in bringing together people of different genders cannot be denied. In this regard, many previous studies have confirmed the leading role of gender in humor and its various production forms (e.g. Kazarian & Martin, 2006; Martin, Puhlik-Doris, Larsen, Gray, & Weir, 2003; Wu, Lin, & Chen, 2016; Yip & Martin, 2006).

The results of the research done by Yip and Martin (2006) further proved this gender difference and indicated that men use more aggressive humor than females. Wu, Lin, and Chen (2016) also asserted that males prefer aggressive humor in different styles; whereas females show more empathy in the understanding or production of humor. This fact is further confirmed by Christov-Moore, Simpson, Coudé, Grigaityte, Iacoboni, & Ferrari, (2014) who also argued that women generally show more empathy in every social interaction and context.

In the same vein, the results of self-reported questionnaires in the study by Chaplin and Aldao (2013) further demonstrated that females are generally more willing to show empathy in social interactions. Hampes (2010) also stated that there is a direct relationship between humor and empathy. Similarly, according to Wu, Lin, and Chen (2016) there is a correlation between empathy and the types of humor presented by people. To put it differently, people with more empathy show more understanding toward other people's feelings and are less likely to hurt their feelings, and accordingly they use less aggressive humor.

Culture and Humor

Although humor is a universal concept and observed around the world, it varies based on unique cultural settings. According to Chen and Martin (2007), people from East and West differ in their attitudes and views toward humor and how they use it. As humor is tightly linked to the mental and emotional well-being and status of people, it is necessary to figure out how humor can influence people and their perceptions toward the issue (Chen & Martin, 2007; Jiang, Li, & Hou, 2019; Martin, 2001; Martin & Ford, 2018).

This is more significant bearing in mind that there are controversial findings on the topic (Chen & Martin, 2007; Hiranandani &Yue, 2014; Kazarian & Martin, 2004). For instance, it is believed that Westerners accept humor more freely (Lefcourt, 2001; Martin & Ford, 2018), whereas Easterners are afraid of endangering their social status be accepting humor (Rudowicz & Yue, 2002; Yue, 2011). In this regard, Apte (1985) argues that:

> Humor is ubiquitous in American society and nothing escapes from becoming its target. Humor in its numerous techniques and forms is directed at the population through all conceivable channels – newsprint, magazines, books, visual and plastic arts, comedy performances, and amateur joke-telling contests, as well as many types of artifacts such as T-shirts, watches, bumper stickers, greeting cards, sculptures, toys, and so forth (p. 30).

Along the same lines, Mireault and Reddy (2016) also believe that contextual factors can greatly affect the understandability of humor. In fact, they emphasize "the context of humorous events does not generally refer to a specific physical setting, but instead is an emotional and social context with a dynamic ebb and flow" (p. 58) which can be either internal or external.

Previous Research on Humor in Classrooms

The topic of humor is a delicate topic to be discussed and accordingly different teachers of different experiences and educational backgrounds have viewed HILL and its benefits from different perspectives (Bell, 2011; Davies, 2015; Heidari-Shahreza, 2018). In one study, Sullivan (2000) explored a Vietnamese EFL teacher's use of humor in the language class, with an eye on the two techniques of storytelling and wordplay (e.g. repetition & puns) for over two months. The findings indicated that the impulsive occurrences of these two techniques could lead to some humorous atmospheres that later could facilitate the learners' motivation and willingness to engage actively in the class activities. Sterling and Loewen's (2015) observational research study investigated L2 Spanish classroom interactions and instances of teacher-initiated humor with a focus on the linguistic content. The data concerning the playful language-related episodes were collected through audio and video recordings. The results indicated that 6% of class interactions or 0.5% of total class time was devoted to humor and playful language.

Olajoke (2013) conducted a study on learners from a tertiary institute in Nigeria. Having controlled for gender in the research, he concluded that humor plays an important role in class activities and that the enhancement of class interactions and the appropriateness of the use of humor relies upon the students' perception and their provided opinions on the diversity and variety of classroom humor. Students' opinions are very influential regarding the appropriateness of humor use, which provides a kind of guidelines for English teachers to help them avoid using the wrong type of humor in their classrooms. As a result, he pointed out that teachers can find strategies to adjust their behaviors and use of humor in the classroom.

Later, in an ethnographic study, Petraki and Pham Nguyen (2016) explored thirty Vietnamese university teachers' perceptions, practices, and preferences with regard to the role of humor and humor types in English as a Foreign Language (EFL) teaching. The utilized instruments included observations, field notes, class video recordings, and interviews. The results showed that all of the teachers either favored or used humor in their language classes. They mainly used humorous comments, jokes, and funny stories to enliven the class atmosphere. In addition, Heidari-Shahreza (2018a) used a cross-sectional analysis of verbal humor and language play instigated by teachers in English as a foreign language (EFL) contexts among Iranian language teachers. The results of the study proved the use of several micro-level verbal humor techniques. The researcher found that puns, allusion, and irony were the three most used techniques applied by the EFL teachers, respectively. Humor forms, funny comments, visual humor, and physical humors were the three most common

forms used by teachers when teaching elementary levels. Additionally, in a more recent study, Heidari-Shahreza (2020) explored HILL in detail. He proposed that HILL is composed of two key components, namely humor competence and language competence. "The former is mainly concerned with enhancing learners' knowledge of humor or humor competency training, while the latter is basically aimed at developing learners' language proficiency" (p. 82).

Purpose of the Study

As Schmitz (2002, p. 96) asserts "there is, without any doubt, a need for research on the use of humor in language classrooms". However, according to Petraki and Nguyen (2016), there is a paucity of research on humor in language classrooms. Heidari-Shahreza (2020) also calls for more attention to the necessity of more research on teaching with and about humor. Moreover, most of the previous studies on humor in L2 classes have been devoted to the language learners and researcher have paid scant attention to EFL teachers in the field (Forman, 2011; Heidari-Shahreza, 2018). Along the same lines, Petraki and Nguyen (2016) point out that EFL teachers need to consider the relevance or appropriateness of the humor used in classrooms and more specifically in different language contexts (Schmitz, 2002). Also, to the best of the researchers' knowledge, no other study has investigated the impact of teacher gender on the application of humor in English language classrooms. Therefore, to contribute to the scant literature on language teacher humor in an EFL context, the present study set out to examine the frequency and appropriateness of humor and language play of 30 male and female Iranian EFL teachers. To this end, the following research questions were put forward:

1. What are the frequency and manifestations of appropriate humor in male Iranian EFL teacher talk?
2. What are the frequency and manifestations of inappropriate humor in male Iranian EFL teacher talk?
3. What are the frequency and manifestations of appropriate humor in female Iranian EFL teacher talk?
4. What are the frequency and manifestations of inappropriate humor in female Iranian EFL teacher talk?
5. Is there any significant difference between male and female teachers in terms of the number of appropriate, inappropriate, and total instances of humor?

Materials and Methods

Participants

A total of thirty Iranian non-native English language teachers (equally divided into males and females) with an average age of 26 took part in this study. These teachers taught English in different language institutes in Tehran and on average had seven years of teaching experience. More detailed information on the participants is depicted below in Table 1. As illustrated, the majority of the teachers, 40%, were from the TEFL (Teaching English as a Foreign Language) field of study. The majority of the participants (37%) had a master's degree. The participants' first language was Persian and all of them were under 30 years of age. In addition, although the classes were held in English, there were some instances where the teachers used the learners' first language, Persian, for different purposes, such as clarifying the meaning of unknown words or even for fun.

Table 1
Descriptive statistics for the participants of the study

Variables	Categories	Frequency	Percentage
Major	TEFL	12	40%
	Literature	2	7%
	Translation	1	3%
	Linguistics	1	3%
	Non-English	4	14%

Variables	Categories	Frequency	Percentage
Degree	B.A.	8	27%
	M.A.	11	37%
	Ph.D.	1	3%
Gender	Male	15	50%
	Female	15	50%
Years of teaching experience	1 – 5 years	9	30%
	6+	11	70%

Instruments

Non-participant observation was used in order to address the research questions of this study. Thus, a recording device was given to each teacher to record the intended session. This was done to avoid the observer effect, which could affect the participants' normal teaching style (Bogdan & Biklen, 2007; Petraki, & Nguyen, 2016; Vanderstoep & Johnston, 2009). All of the teachers provided 90 minutes of audio-recorded data including class interactions and discourse. Occurrences of humor were analyzed in the transcribed texts produced from the class conversations.

Data Collection Procedure

Having employed a case study, the data were collected from the two institutes, in which 15 male and 15 female teachers had been selected through random sampling. It should be noted that there was no significant difference between the two institutes in terms of the teachers, students, and the books taught. The classes were 90-minutes each and the core approach was communicative language teaching. One session from each teacher's class was recorded via a non-observant method, which is without the presence of the researcher. With the provided recorder, the teachers recorded a complete session of their classes.

Data Analysis

All 30 audio classes (45 hours of audio-recorded data) were precisely transcribed. Then, the conversations were analyzed from a broader omniscient point of view considering both the conversations and contexts. The overall analysis of the data was based on the coding scheme, borrowed from Frymier, Wanzer, and Wojtaszczyk (2008), which was selected as the main framework of the study (See the Appendix). The coding scheme included the categorization of various types of humor (funny comments, providing jokes, providing humorous examples, making stories, teacher performance, creative language, teasing students, and critical humor), their appropriateness (appropriate or inappropriate), and their relatedness to the content of the study (related or unrelated).

Moreover, based on the classification scheme developed by Wanzer et al. (2006), the humor instances were categorized into their functional categories. The humor types were pinpointed and their frequencies were counted to gain an understanding of the exact times when specific humorous features happened. In addition, through the comparison of the responses and the verbal feedback available from the learners, the researchers were able to understand the extent of each humor's appropriateness. Finally, it should be said that 30% of the recordings (12 recorded classes comprised of six male teachers and six female teachers) were re-coded by an expert whose expertise was on classroom discourse and two-way random interclass correlations were used for inter-rater reliability.

Results

RQ1: What are the Frequency and Manifestations of Appropriate Humor in Male Iranian Efl Teacher Talk?

This question inquired about the number of appropriate instances where male teachers used humor in the classroom. Table 2 clearly illustrates appropriate instances plus the total instances of humor used with regard to the male teachers. As can be seen in Table 2, the total number of appropriate instances of humor was 69. Furthermore, the

percentage of appropriate instances of humor produced by male teachers was 58% (n = 40) of the total numbers of appropriate instances of humor produced by all of the teachers. Of the total instances of humor used by male teachers (n = 49), 82% were appropriate to the context. Inter-rater reliability was also assessed using a two-way random average measure intra-class correlation coefficient, which was .92 and indicates excellent reliability.

Table 2
Humor instances across gender and type

	Appropriate	Inappropriate	Total
Male	40 (58%)	9 (53%)	49 (57%)
Female	29 (42%)	8 (47%)	37 (43%)
Total	69 (100%)	17 (100%)	86 (100%)

Based on the data, the mostly-used categories of humor for male teachers were *funny comments*, *creative language*, and *teacher performance*. The most important category of appropriately-used humor was "funny comments". Considering the instances of appropriate humor from male teachers (n=40), 11 instances were funny comments amounted were 27%. Of these, 91% were related to the context and the rest (9%) was unrelated. The second-most frequent humor type was creative language which occurred nine times (22%) in male utterances appropriately, with 77.7% of the instances related and 23.3% unrelated to the classroom. Teacher performance occurred five times (12%), among them 80% were related and 20% were unrelated. *Providing jokes* occurred five times (13%), of which 60% were related and 40% were unrelated. *Critical humor* was noted four times (10%), all of which were related; *making stories* three times (8%), all of which were related; *provided examples* two times (5%), 50% of which was related and the other half were unrelated; and *teasing students* happened one time (3%), all of which were related to the context. An example of an appropriate joke used five times (13%) by male teachers is depicted in the excerpt below:

S: Teacher, can we say they are? (Instead of Louis and Jack)
T: If you are very lazy and Shirazi yes! (In everyday jokes, Shirazi people are famous for being very lazy). When we have 'and' you can know that they are plural.

It can be understood that in this example the teacher used humor to teach the concept of *plural* by the use of *and.*

RQ2: What are the frequency and manifestations of inappropriate humor in male Iranian EFL teacher talk?

The second research question deals with the general number of occurrences of inappropriate humor by male teachers. As can be seen in Table 2, the total number of instances of inappropriate humor used by teachers were 17. Furthermore, the percentage of instances of inappropriate humor produced by male teachers was 53% (n = 9) of the total number of inappropriate instances of humors produced by both genders. Also, from that total instances of humor used by male teachers (n = 49), 18% were considered inappropriate to the context. Inter-rater reliability was assessed in this part of the analysis using a two-way random average-measures intraclass correlation coefficient, which was .87 and indicates good reliability.

According to the data, it was contended that the mostly-used categories considered to be inappropriate humor by male teachers were funny comments (n = 4, p = 45%), 50% of which were related and the other half were unrelated to the context, and teasing students (n = 2, p = 22%), both of which were related to the context of the classroom. Critical humor occurred only once (p = 11%) and was related. Similarly, providing examples and creative language both occurred one time (p = 11%) each and they were unrelated. There were no instances of teacher performance, making stories, or joke provision. Below is an example of a related example of inappropriate humor uttered by a male teacher:

S: I want to go to park and play with slide.
T: The phrase is "to go on a slide". You want to go on a slide! (with a surprising voice)
S: Yes I want to go to park and go on a slide.
T: You want to go on a slide and you call these two ladies babies?!!!
They should buy chocolate for you kid.
Students laugh.

As can be seen above, the first funny comment was added after giving feedback on the correct use of the phrase "to go on a slide" and the repetition used resulted in greater exposure to the phrase.

RQ3: What are the frequency and manifestations of appropriate humor in female Iranian EFL teacher talk?

Following the third research question, the two factors of appropriateness and gender were considering once again. As depicted in Table 2, it can be seen that the total number of appropriate instances of humor produced by female teachers was 42% (n = 29) of the total number of appropriate instances produced by both genders. Likewise, it should be mentioned that from the total number of instances used by female teachers (n = 37), 29 of them (78%) were appropriate to the context. Inter-rater reliability was calculated by means of a two-way random average-measures intraclass correlation coefficient, which proved to be .95 and indicates excellent reliability.

It was also found that the most-used categories for appropriate humor by female teachers were funny comments, providing jokes, teacher performance, and creative language. The most important category of appropriately-used humor was funny comments, which occurred 15 times (p = 52%), and among them 80% were related and 20% were unrelated. The second significant humor type was providing jokes, which occurred four times (p = 14%), 75% of which were related and 25% were unrelated. Teacher performance was observed three times (10%), 66.6% of which were related and 33.4% were unrelated. Creative language occurred three times (p = 10%), 66.6% related and 33.4% unrelated. Critical humor occurred two times (p = 7%), 50% related and 50% unrelated. Making stories happened once (p = 3%) and it was related. Finally, providing examples was not observed at all and teasing students was seen only one time (p = 4%) and it was related. To provide an example, the excerpt below manifests a type of unrelated appropriate creative language that was used by a female teacher.

T: And one thing about your phones… and if the authorities know, they will kill me… you can use it as a dictionary… Students laugh
T: But don't use it for checking Instagram, telegram or me and the class, spontaneously (یهویی، کلاس و من) Students laugh.

As can be seen, a concept from the first language (i.e., .. من و کلاس، یهویی a common situation among Iranian people when at the moment that something interesting is happening, they try to take a picture of that event or with a specific person and say: *me and the -----, spontaneously or just now*) was used in the second language, which actually may not be popular or used a lot in the L2. However, it was appropriately used to convey meaning.

RQ4: What are the frequency and manifestations of inappropriate humor in female Iranian EFL teacher talk?

The last research question deals with the frequency and manifestations of inappropriate humor by female teachers. As shown in Table 2, it can be said that the total instances of inappropriate humor produced by female teachers was 47% (n = 8) of the total instances of inappropriate humor produced by all teachers. Likewise, it should be mentioned that from that total instances of humor used by female teachers (n = 37), 22% were inappropriate to the context. Additionally, inter-rater reliability was estimated in this part of the study through a two-way random average-measures intra-class correlation coefficient, which was .82 and indicates good reliability.

The data showed that using funny comments was the mostly-used category resulting in inappropriate humor by the female teachers. With regard to female teacher inappropriate humor, four instances were funny comments (p = 45%), of which 50% were related and 50% were unrelated. The second-most prevalent humor type was joke provision, which occurred twice (p = 22%), and both of which were related to the context of the classroom. Teasing students occurred once (p = 11%) and it was unrelated. Example provision was observed one time (p = 11%) in an unrelated way as a realization of inappropriate female teacher humor use. There were no instances of teacher performance, critical humor, making stories, or creative language as inappropriate usage. Below is an example of unrelated inappropriate humor uttered by a female teacher.

T: Oh, you have a guitar. Is it yours?
S1: Yes.

T: I think you have to play for us.
S1: No.
T: Yesssssss.
S1: I can't play.
T: Why? What is the problem? I promise that we`ll close our eyes.
S2: Because she went to class just two sessions.
T: So after two sessions you can make just some noise. It`s ok. We will enjoy it. Please play dalang dalang (which means an unrelated funny noise) for us.
Students laugh.

In this instance, not only was it inappropriate for the students, but it was also unrelated to the context of the lesson as well. It was just a funny comment about something that popped up.

RQ5: Is there any significant difference between male and female teachers in terms of the number of appropriate, inappropriate, and the total instances of humor?

To answer this research question, three Mann–Whitney U tests (hereafter MWT) were utilized. We used this non-parametric test instead of independent-samples *t* tests, a parametric test, given that the distributional assumption of normality, the most important statistical assumption of independent-samples *t* tests, was not satisfied in any of the distributions except one (see Table 3 for the results of two tests of normality, Kolmogorov-Smirnov test and Shapiro-Wilk test).

Table 3
Tests of normality

Humor	Gender	Kolmogorov-Smirnov			Shapiro-Wilk		
		Value	df	Sig.	Value	df	Sig.
Appropriate	Male	.31	15	.00	.81	15	.00
	Female	.33	15	.00	.82	15	.01
Inappropriate	Male	.33	15	.00	.76	15	.00
	Female	.33	15	.00	.74	15	.00
Total	Male	.23	15	.03	.83	15	.01
	Female	.18	15	.20	.91	15	.13

As depicted in Table 4, The first MWT test showed that there were no significant differences between male (Mdn = 2, M = 2.66) and female (Mdn = 2, M = 1.93) teachers with regard to the number of instances of appropriate humor, U = 98.5, z = -.62, p = .56. The second MWT also revealed no significant differences between male (Mdn = 0, M = .6) and female (Mdn = 0, M = .53) teachers with regard to the number of instances of inappropriate humor, U = 109, z = -.16, p = .90. Ultimately, the third MWT also illustrated that there were no significant differences between male (Mdn = 2, M = 3.26) and female (Mdn = 2, M = 2.46.) teachers with regard to the total instances of humor, U = 103, z = -.40, p = .71.

Table 4
Descriptive statistics of humor instances across type and gender

Humor	Gender	N	Mean Rank	Sum of Ranks	Mean	Median
Appropriate	Male	15	16.43	246.50	2.66	2
	Female	15	14.57	218.50	1.93	2
Inappropriate	Male	15	15.73	236.00	0.6	0
	Female	15	15.27	229.00	0.53	0
Total	Male	15	16.13	242.00	3.26	2
	Female	15	14.87	223.00	2.46	2

Discussion

The issue of humor and its appropriateness in the EFL context has remained understudied, despite its dramatic impact on the process of teaching and learning a second language. The present research was conducted with the participation of multiple teachers in a foreign language context in order to illustrate the various types of humor used in EFL classrooms with regard to gender. The results showed interesting facts with regard to the appropriateness of humor used by male and female Iranian EFL teachers.

In terms of the first research question, the results revealed that the use of appropriate instances was nearly four times greater than inappropriate ones. From among the appropriate humor examples, male teachers used 40 instances of appropriate humor, which were mainly focused on funny comments, creative language, and teacher performance. The results of the study supported the fact that funny comments were the main tool of humor for male teachers, which is in line with other previous studies (e.g., Heidari-Shahreza, 2018a). Male teachers performed better using different varieties of humor types. This can be justified bearing in mind that in Iranian Persian culture, men are free and less likely to be judged by their audience. Furthermore, they are usually the starters of humor in conversations, even in their own language and culture; therefore, they tend to be good humor initiators. However, this result runs counter to a number of studies (e.g., Abdullah & Akhter, 2015; Garner, 2006), which showed that half of the instances of humor used by the male teachers were appropriate. To put it another way, in the present study less than one-fourth of the humor used by the male teachers was considered inappropriate. It may be contended that male teachers felt free to use these types of humor. This finding goes in tandem with Yip and Martin (2006) and Wu, Lin, and Chen (2016) who also asserted that men use more aggressive and inappropriate humor than women.

With regard to the second research question, it was revealed that the male teachers did not have any inappropriate instances of humor use in categories such as "teacher performance", "joke provision", and "story provision". This suggests that, although the use of these types of humor might be less frequent, they were perfectly used as a means of humor provision without any inappropriateness. This finding is in consonance with that of Heidari-Shahreza (2018b), who found that cases of relevant humor outnumbered the irrelevant ones in nearly all observed classrooms. Moreover, the most inappropriate humor type by males was funny comments. One reason that might justify this occurrence is the fact that the frequency of use brought about the frequency of inappropriate use. Therefore, it was perceived that funny comments were actually positive humor types implemented by male teachers but the reason for inappropriate occurrences was that they were being used more than the other types of humor. Other types of humor only had one or two inappropriate instances. By contrast, with regard to the types of inappropriate humor observed from the male teachers, the current study is incompatible with some previous studies (e.g., Abdullah & Akhter, 2015; Garner, 2006; Wanzer, Frymier, Wojtaszczyk, & Smith, 2006), who found that inappropriate humor consisted of the teachers sarcastically and humorously talking about sex, race, religion, or sexual orientation.

Regarding the third research question, female teachers also used "funny comments" as their main category of humor provision, to the extent that the difference between their productions was eye-catching considering the paucity of their second and third frequent types of humor. It can be contended that funny comments played an influential role when considering Iranian EFL teachers' humor production, especially in the case of female teachers. It appears that using funny comments is the dominant form of humor provision in Iranian EFL teachers' mother tongue, regardless of their gender. Another significant factor in the practice of funny comments was the difference between male and female teachers. As the results of the current study revealed, female teachers had greater tendencies toward using funny comments as their main tool of humor provision. The examples provided by female teachers were greater in number in comparison with the male teachers. This finding can be justified as a response to the boundaries and limitations faced by female Iranian Muslim EFL teachers. It seems that they preferred to use funny comments as a safer expression of humor provision. This may remind us of the cross-cultural manifestations of humor. It is argued that people around the world (e.g. Westerners and Easterners) act differently in the production and perception of humor (Yue, Jiang, Lu, & Hiranandani, 2016). For instance, just like Iranian women, Chinese women are also encouraged to laugh quietly (in comparison to men) and in some cases even cover their mouths when laughing (Lin, 1934).

Another important source of humor production was "*creative language*" for male teachers. However, considering female teachers, "*provision of jokes*" was a more frequent form of humor provision in comparison with male teachers. In addition, it was found that most of the humor types used by male teachers were related to the context of the study and the teaching materials. In comparison, female teachers had more unrelated humor, although both male and female teachers used related humor more than the unrelated ones. As mentioned earlier, divergence could be seen in the second most frequently observed humor category. Although, male teachers had a tendency toward using creative language, the usage of jokes seemed to be more consistent with female teachers. However, it was shown that instances of creative language for male teachers were nearly as high as the first one, funny comments. Nine instances of creative language were used by male teachers, which were as frequent as funny comments. On the other hand, female teachers only used jokes four times, which was not comparable with their 15 instances of funny comments. Therefore, a simple but significant conclusion can be made, which is the fact that the usage of funny comments was the main tool of humor provision for females. A thorough analysis and comparison of the current study's results with the related literature on the appropriateness of female EFL teachers' humor indicates that the Iranian female teachers lacked variety in their humor production. Funny comments were basically the only tool of humor production used by female teachers in the present study, which is caused by caused by the insecurity that other humor types may bring to the teaching context. However, Petraki and Nguyen (2016) found that "media/ external objects", "jokes", "examples", and "stories" were among the most frequent methods of humor provision among female teachers.

With regard to the fourth research question on inappropriate humor use, the female teachers in the present research were different from teachers in previous studies. Usually, female inappropriate instances of humor have been proved to be targeting one learner based on a variety of reasons. Actually, disparaging humor was used in eighty percent of the occurrence of such inappropriate instances. By and large, some of the findings in the literature were related to the comparison of the performances and productions of humor from both male and female teachers and the extent of appropriateness of the categories. For instance, a comparison of the relatedness of appropriate humor use was emphasized in Alatalo and Poutiainen's (2016) study. Comparable to our study, their study's results proved the efficiency of the use of related humor in the educational context as a motivational technique used by EFL teachers. In another study, Petraki and Nguyen (2016) reported more uses of humor by female teachers; this is not in harmony with the results of the present study, which indicated more frequent usage of humor by male teachers, apparently due to their freedom of humor initiation. Moreover, along with the findings of the present study, the main tool for humor presentation in their study was funny comments, which may have arisen from the adjacency of cultural factors. Additionally, another factor is the presentation of linguistic elements in the case of puns, allusion, and irony. These can be investigated in the present study considering the presence of linguistic elements according to Wanzer et al.'s (2006) categorization, which had been implemented in the forms of providing jokes, providing examples, and teasing students. Finally, the relevance of the present findings can be pinpointed in the study by Heidari-Shahreza (2018) whose results showed that funny comments were the most frequent type of humor used by teachers.

Finally, regarding the last research question, it was observed that male and female teachers did not show any significant differences in the use of humor in English language classrooms. Although, the male and female teachers were varied in their use of appropriate and inappropriate humor, this difference was not significant. As Holmes (2006) argues, part of the identity of men and women in the workplace is shaped by their use of humor. In fact, it seems that "humor can provide insights into the gender stereotypes to which participants relate, or even subscribe" (Holmes, 2006, 41). According to Crawford (2003, p. 1427), men and women utilize humor in their speech and "in same-gender and mixed-gender settings as one of the tools of gender construction. Through it and other means, they constitute themselves as masculine men and feminine women". Teachers in the current study also may have used humor differently in various classes with different gender distributions. To put it differently, language teachers may use different types of humor when they have a class with male or female students of different age levels or backgrounds. Our results in this study are in line with Abel (1998) who reported no significant difference between the use of humor by males and females in general. Some other researchers however (e.g., Johnson, 1991; Rim, 1988) found differences between males and females in terms of their perceptions and attitudes toward humor. As mentioned previously, no specific study could be pinpointed focusing on the male and female English language teachers' use of humor in the class and their variations.

Conclusion

The main objective of the present study was to investigate the appropriateness and relevance of verbal humor among Iranian EFL language teachers. The data were collected through non-observant class observations of thirty male and female EFL teachers. The results observed the frequency of different instances of humor used by male and female teachers and indicated that funny comments were the main type of humor used in Iranian EFL classrooms. Furthermore, male teachers were observed to use more humor more often in comparison with female teachers. In addition, from the 58 instances produced by male teachers, nine of them were inappropriate to the classroom environment. However, from the 40 instances observed by female teachers, eight of them were deemed inappropriate, which was a higher percentage than their male counterparts. Generally, with regard to the overall findings of the study, the role of religion and culture cannot be denied. The male and female teachers in the current study came from an Islamic Iranian cultural background. Undoubtedly, this background could have an effect on the teachers' attitudes and inclinations toward the use of different forms of humor.

The results of our study have a number of pedagogical implications. First and foremost, we think that studies of this kind can raise the awareness of language teachers toward different pedagogical issues. This study brings to the fore the significance of humor and the prevalence of its various types among male and female EFL teachers in language classrooms. It needs to be noted that there are numerous positive points for using humor in the class (e.g. increased learner motivation and lower anxiety levels) and we hope this paper could bring the general significance of humor into the limelight. Moreover, current and prospective EFL teachers, whether in Iran or around the world, can benefit from our findings with regard to the different types of humor used by language teachers. We believe this can raise awareness and make teachers more aware of the significant role of humor in language learning and the learners' success. This would lead to teachers who consider the probable advantages of humor in their classrooms and try to come up with practical techniques to enliven the classroom atmosphere, increase learner motivation and self-confidence, diminish their anxiety, and increase their willingness to communicate. Finally, the present study's results emphasize the necessity of training and education for EFL teachers on the integration of humor into EFL classrooms. The results indicated that Iranian EFL teachers were not totally aware of the different types of humor, their appropriateness in different contexts, and their relevant use in language classes. That is to say, teachers need to be educated on how to practically use different kinds of humor in different situations in EFL classes. This further calls for more attention to the content of teacher education courses. Teacher educators need to assist teachers who are not confident enough in using humor in language classes through conducting workshops on the proper integration of humor into the classroom context.

It needs to be acknowledged that just like any research study, this study had some limitations that may affect the generalizability of the findings. We investigated the use of verbal humor produced by thirty Iranian EFL teachers through observation. Further studies may explore the topic more meticulously through a diversity of research instruments (e.g. questionnaires & interviews) and with a larger number of participants. Moreover, interested researchers may also investigate teachers' attitudes toward different types of humor and in different educational contexts namely, schools, private institutes, and universities. Also, to shed more light on the use of humor in language classrooms, it seems necessary to investigate language learners' views with regard to humor and its different types along with how they respond to humor initiated by teachers. Finally, one remaining question is whether the various forms of humor can have any impact on the acquisition of a second language among learners of different age levels and cultures in different contexts.

Conflict of interests

The authors declare that they have no conflict of interest.

References

Abdullah, S., & Akhter, J. (2015). Uses of humour in an English language class. *International Journal of Humanities and Cultural Studies, 2*(2), 16-21.

Abel, M. H. (1998). Interaction of humor and gender in moderating relationships between stress and outcomes. *The Journal of Psychology*, *132*(3), 267-276.

Apte, M. L. (1985). *Humor and laughter: An anthropological approach*. Cornell University Press.
Attardo, S. (1994). *Linguistic theories of humor*. Mouton de Gruyter.
Bell, N. D. (2011). Humor scholarship and TESOL: Applying findings and establishing a research agenda. *TESOL Quarterly, 45*(1), 134-159. https://doi.org/10.5054/tq.2011.240857
Bell, N. (2012). Comparing playful and nonplayful incidental attention to form. *Language Learning, 62*(1), 236-265. https://doi.org/10.1111/j.1467-9922.2011.00630.x
Berk, R. (2000). Does humour in course tests reduce anxiety and improve performance? *College Teaching, 48*(4), 151-159. https://doi.org/10.1080/87567550009595834
Bogdan, R. C., & Biklen, S. K. (2007). *Qualitative research for education: An introduction to theories and methods* (5th ed.). Pearson.
Booth-Butterfield, S., & Booth-Butterfield, M. (1991). Individual differences in the communication of humorous messages. *Southern Communication Journal, 56*, 205-218.
Carter, R., & McCarthy, M. (2004). Talking, creating: Interactional language, creativity, and context. *Applied Linguistics, 25*(1), 62-88. https://doi.org/10.1093/applin/25.1.62.
Chaplin, T. M., & Aldao, A. (2013). Gender differences in emotion expression in children: A meta-analytic review. *Psychological Bulletin, 139*(4), 735-765. https://doi.org/10.1037/a0030737
Christov-Moore, L., Simpson, E. A., Coudé, G., Grigaityte, K., Iacoboni, M., & Ferrari, P. F. (2014). Empathy: Gender effects in brain and behavior. *Neuroscience and Bio Behavioral Reviews, 46*, 604-62 7. https://doi.org/10.1016/j.neubiorev.2014.09.001
Cook, G. (2000). *Language play, language learning*. Oxford University Press.
Chen, G., & Martin, R. A. (2007). A comparison of humor styles, coping humor, and mental health between Chinese and Canadian university students. *Humor: International Journal of Humor Research, 20*, 215-234. https://doi.org/10.1515/HUMOR.2007.011
Crawford, M. (2003). Gender and humor in social context. *Journal of Pragmatics*, *35*(9), 1413-1430. https://doi.org/10.1016/S0378-2166(02)00183-2
Davies, C. E. (2015). Humor in intercultural interaction as both content and process in the classroom. *Humor: International Journal of Humor Research, 28*(3), 375-395. https://doi.org/10.1515/humor-2015-0065
Forman, R. (2011). Humorous language play in a Thai EFL classroom. *Applied Linguistics, 32*(5), 541-565. https://doi.org/10.1093/applin/amr022
Garner, R. L. (2006). Humour in pedagogy: How ha-ha can lead aha! *College Teaching, 54*(1), 177-180. https://doi.org/10.3200/CTCH.54.1.177-180
Gönülal, T. (2018). Investigating the potential of humour in EFL classrooms: An attitudinal study. *European Journal of Humour Research, 6*(1), 141-161. https://hdl.handle.net/20.500.12432/1211
Gorham, J., & Christophel, D. M. (1990). The relationship of teachers' use of humor in the classroom to immediacy and student learning. *Communication Education, 39*, 46-62.
Hampes, W. P. (2010). The relation between humor styles and empathy. *Europe's Journal of Psychology, 6*(3), 34-45. https://doi.org/10.5964/ejop.v6i3.207
Heidari-Shahreza, M. A. (2018a). A cross-sectional analysis of teacher-initiated verbal humor and ludic language play in an English as a foreign language (EFL) context. *Cogent Education, 5*(1), 1-20. https://doi.org/10.1080/2331186X.2018.1430474
Heidari-Shahreza, M. A. (2018b). A proficiency-based analysis of EFL learners' humorous language play. *International Journal of Applied Linguistics, 28*(3), 406-423. https://doi.org/10.1111/ijal.12212
Heidari-Shahreza, M. A., & Heydari, S. (2019). Humor-integrated language learning: A teacher's and student's perspective. *TESOL Journal, 10*(1), 1-4. https://doi.org/10.1002/tesj.376
Heidari-Shahreza, M. A. (2020). Humor-intergrated language learning (HILL): Teaching with and about humor. In J. Rucynski, & C. Prichard (Eds.), *Bridging the humor barrier: Humor competence training in English language teaching* (pp. 79-106). Lexington Books, Rowman & Littlefield Publishers.
Hiranandani, N. A., & Yue, X. D. (2014). Humour styles, gelotophobia and self-esteem among Chinese and Indian university students. *Asian Journal of Social Psychology*, *17*, 319-324. https://doi.org/ 10.1111/ajsp.12066
Holmes, J. (2006). Sharing a laugh: Pragmatic aspects of humor and gender in the workplace. *Journal of Pragmatics*, *38*(1), 26-50. https://doi.org/10.1016/j.pragma.2005.06.007
Johnson, A. M. (1991). Sex differences in the jokes college students tell. *Psychological Reports, 68*, 851-854.
Kazarian, S. S., & Martin, R. A. (2004). Humour styles, personality, and well-being among Lebanese university students. *European Journal of Personality, 18*, 209-219. https://doi.org/10.1002/per.505.
Kazarian, S. S., & Martin, R. A. (2006). Humor styles, culture-related personality, well-being, and family adjustment among Armenians in Lebanon. *Humor International Journal of Humor Research, 19(*4), 405-423. https://doi.org/10.1515/HUMOR.2006.020

Lefcourt, H. M. (2001). *Humor: The psychology of living buoyantly*. Kluwer Academic. https://doi.org/10.1007/978-1-4615-4287-2

Lin, Y. T. (1934). On humor. *Analects Fortnightly, 33*, 434-438.

Lovorn, M., & Holaway, C. (2015). Teachers' perceptions of humour as a classroom teaching, interaction, and management tool. *European Journal of Humour Research, 3*(4), 24-35. http://dx.doi.org/10.7592/EJHR2015.3.4.lovorn

Mantooth, J. D. (2010). *The effects of professor humour on college students' attention and retention* [Unpublished doctoral dissertation]. Auburn University.

Martin, R. A. (2001). Humor, laughter, and physical health: Methodological issues and research findings. *Psychological bulletin, 127*, 504-519. https://doi.org/10.1037/0033-2909.127. 4.504

Martin, R. A. (2007). *The psychology of humor: An integrative approach*. Cambridge University Press.

Martin, R. A., & Ford, T. (2018). *The psychology of humor: An integrative approach*. Elsevier Academic Press.

Martin, R. A., Puhlik-Doris, P., Larsen, G., Gray, J., & Weir, K. (2003). Individual differences in uses of humor and their relation to psychological well-being: Development of the humor styles questionnaire. *Journal of Research in Personality, 37*, 48-75. https://doi.org/10.1016/S0092-6566(02)00534-2

Mireault, G. C., & Reddy, V. (2016). *Humor in infants: Developmental and psychological perspectives*. Springer.

Morrison, M. K. (2008). *Using humor to maximize learning: The links between positive emotions and education*. Rowman & Littlefield Education.

Petraki, E., & Nguyen, H. (2016). Do Asian EFL teachers use humor in the classroom? A case study of Vietnamese EFL university teachers. *System, 61*, 98-109. https://doi.org/10.1016/j.system.2016.08.002

Pomerantz, A., & Bell, N. D. (2011). Humor as safe house in the foreign language classroom. *The Modern Language Journal, 95*,148-161. https://doi.org/10.1111/j.1540-4781.2011.01274.x

Rim, Y. (1988). Sense of humour and coping styles. *Personality and Individual Differences, 9*, 559-564. https://doi.org/10.1016/0191-8869(88)90153-5

Rudowicz, E., & Yue, X. D. (2002). Compatibility of Chinese and creative personalities. *Creativity Research Journal, 14*, 387-394. https://doi.org/10.1207/S15326934CRJ1434_9

Schmitz, J. B. (2002). Humor as a pedagogical tool in foreign language and translation courses. *Humor: International Journal of Humor Research, 15*, 89-113. https://doi.org/10.1515/humr.20 02.0 07

Schmidt, S., & Williams, A. (2001). Memory for humorous cartoons. *Memory and Cognition, 29*, 305-311. https://doi.org/10.3758/BF03194924

Sterling, S., & Loewen, S. (2015). The occurrence of teacher-initiated playful LREs in a Spanish L2 classroom. *System, 53*, 73-83. https://doi.org/10.1016/j.system.2015.07.002

Sullivan, P. N. (2000). Playfulness as mediation in communicative language teaching in a Vietnamese classroom. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 115-131). Oxford University Press.

Tamblyn, D. (2003). *Laugh and learn - 95 ways to use humor for more effective teaching and training*. Amacom.

Vanderstoep, S. W., & Johnston, D. D. (2009). *Research methods for everyday life: Blending qualitative and quantitative approaches.* Jossey-Bass.

Wagner, M. & Urios-Aparisi, E. (2011). The use of humor in the foreign language classroom: Funny and effective. *Humour: International Journal of Humour Research, 24*(4), 399-434. https://doi.org/10.1515/humr.2011.024

Wanzer, M., & Frymier, A. (1999). The relationship between student perceptions of instructor humor and students' report of learning. *Communicative Education*, *48*, 48-62. https://doi.org/10.1080/03634529909379152

Wanzer, M., Frymier, A., Wojtaszczyk, A., & Smith, T. (2006). Appropriate and inappropriate uses of humor by instructors. *Communication Education, 55*, 178-196. https://doi.org/10.1080/03634520600566132

Wanzer, B. M., Frymier, A. B., & Irwin, J. (2010). An explanation of the relationship between instructor humour and student learning: Instructional humour processing theory. *Communication Education, 59*(1), 1-18. https://doi.org/10.1080/03634520903367238.

Wu, C. L., Lin, H. Y., & Chen, H. C. (2016). Gender differences in humour styles of young adolescents: Empathy as a mediator. *Personality and Individual Differences, 99*, 139-143.

Yip, J. A., & Martin, R. A. (2006). Sense of humor, emotional intelligence, and social competence. *Journal of Research in Personality, 40(*6), 1202-1208. https://doi.org/10.1016/j.jrp.2005.08.005

Yue, X. D. (2011). The Chinese ambivalence to humor: Views from undergraduates in Hong Kong and China. *Humor: International Journal of Humor Research, 24*, 463-480. https://doi.org/10.1515/humr.2011.026.

Yue, X., Jiang, F., Lu, S., & Hiranandani, N. (2016). To be or not to be humorous? Cross cultural perspectives on humor. *Frontiers in Psychology, 7*, 1495-1512. https://doi.org/10.3389/fpsyg.2016.01495

Appendix

The coding scheme for the categorization of various types of humor (Frymier, Wanzer, & Wojtaszczyk, 2008):

- **Funny comments**: use of some humorous words or phrases in order to add some fun to the talk
- **Providing jokes**: telling some famous jokes about common subjects where both the teachers and the students are aware of
- **providing humorous examples**: funny and humorous examples of are mentioned of when someone did the funny thing
- **making stories**: short funny stories are recalled to make someone laugh
- **teacher performance**: use of some actions by the teacher to make the students laugh, which may include some physical activity
- **creative language**: includes the creative use of language and use of funny words in humor along with play with words
- **teasing students**: use of some funny comments relating to the students and their actions to enliven the class atmosphere
- **critical humor**: a humor which may include harsh or offensive jokes or satire

National Research University Higher School of Economics
Journal of Language & Education Volume 7, Issue 1, 2021

Alyousef, H. S. (2021). The Status of Theme in Research Article Abstracts in Seven Dentistry Subdisciplines: A Text-Based Study of Intradisciplinary Variations and Similarities in Thematic Choices and Thematic Progression Patterns. *Journal of Language and Education, 7*(1), 28-45. https://doi.org/10.17323/jle.2021.10574

The Status of Theme in Research Article Abstracts in Seven Dentistry Subdisciplines: A Text-Based Study of Intradisciplinary Variations and Similarities in Thematic Choices and Thematic Progression Patterns

Hesham Suleiman Alyousef
King Saud University

Correspondence concerning this article should be addressed to Hesham Suleiman Alyousef, Department of English Language and Literature, Faculty of Arts, King Saud University, Riyadh, Saudi Arabia.
Email: hesham@ksu.edu.sa

A research article (RA) abstract provides an overview or summary of the whole research. It is one of the most important sections in an RA since it is the first section researchers read to decide if the article is relevant to their research or not. Researchers need to know the intradisciplinary (within the same discipline) variations and similarities in the choice of Theme[1] and thematic progression (TP) patterns in RA abstracts in their discipline. Several studies have investigated variations and similarities in the use of Theme across disciplines. To the best of our knowledge, there is a lack of studies investigating intradisciplinary variations in the use of Theme in RA abstracts of dentistry subdisciplines. As epistemological differences exist between the various dentistry subdisciplines, it is pertinent to examine if there are intradisciplinary variations in the construction of Theme in the subfields of dentistry. The present study aims to investigate if there are any intradisciplinary variations and similarities in the use of Theme and TP patterns across seven dentistry subdisciplines: oral sciences, periodontics, endodontics, operative dentistry, prosthodontics, oral and maxillofacial surgery, and orthodontics. The study is framed by Halliday's systemic functional linguistics' approach to language and Daneš's model for TP patterns. The findings revealed intradisciplinary significant differences between the subfields of dentistry in terms of the use of Theme types at $p < .05$ (p-value=0.0294), while there were no significant differences in the use of TP patterns and thematic markedness. Various interesting linguistic features characterizing the subdisciplines were found, although no significant interdisciplinary differences were found between dentistry RA abstracts and the findings reported in the literature of other disciplines. Finally, implications for novice dental researchers attempting to write an RA abstract are presented.

Keywords: Research article (RA) abstract, Theme choice, thematic progression patterns, dentistry subdiscipline, intradisciplinary and interdisciplinary variations

Introduction

Analysis of the thematic organization of discourse reveals the way texts are created and organized coherently, thereby facilitating the smooth readability of texts. The application of this analysis to the research article (RA) abstract is necessary since the findings may provide invaluable insights for writers of an RA abstract. Since the abstract summarizes the whole RA, it is the first section researchers read to decide whether or not to read the rest of the RA.

Whereas few studies investigated thematic choices and thematic progression (TP) patterns in RA abstracts of one discipline (Lorés, 2004; Nwogu & Bloor, 1991), others conducted interdisciplinary (across disciplines) studies of variations and similarities in the thematic organization of RA abstracts across a range of disciplines

[1] As the word theme has other meanings (or homonyms), Theme and Rheme are capitalized in this paper to exclusively refer to Theme and Rheme parts of a clause.

(Alotaibi, 2020; Ebrahimi, 2016; Ebrahimi & Khedri, 2011; Ghadessy, 1999; Leong, 2016).The study of Theme in dentistry RA abstracts does not appear to have attracted this kind of research. Moreover, as epistemologies (or demands) are unique to each dentistry subdiscipline (or dental specialty), it is essential to investigate if there are intradisciplinary (within the same discipline) variations in the use of Theme in RA abstracts of various subfields of dentistry. For example, whereas the subdiscipline of periodontics is concerned with "oral hygiene, oral health, dental health, dental plaque, dental calculus, dental debris, dental caries, dental fluorosis, gingivitis, periodontitis, and periodontal treatment" (Shamim, 2018, p. 102), the discipline of oral and maxillofacial surgery is concerned with the treatment of cases, defects, and injuries in the head, face, neck, jaws, and the hard and soft tissues of the oral and maxillofacial region. To the best of our knowledge, there is a lack of studies investigating intradisciplinary variations in the construction of Theme in RA abstracts of dentistry subdisciplines.

The present study aims to investigate if there are any intradisciplinary variations or similarities in the use of Theme and TP patterns across seven dentistry subdisciplines. The study addressed the following research question: Are there any significant intradisciplinary variations and similarities in the choice of Theme and TP patterns in RA abstracts from seven dentistry subdisciplines: oral sciences, periodontics, endodontics, operative dentistry, prosthodontics, oral and maxillofacial surgery, and orthodontics?

Theoretical Framework

The present study is framed by Halliday's (2014) systemic functional linguistics' (SFL) approach to language and Daneš's (1974) model for TP patterns. SFL holds that language serves three main functions: experiential meanings represented by our experience of the world, interpersonal meanings represented by our use of the interactive features, and textual meanings that create cohesive and coherent texts. An analysis of the textual metafunction of a text, therefore, can reveal its organization. One of the main structural systems within the textual metafunction is Theme.

Theme is always clause-initial in the English language, it is "the element that serves as the point of departure of the message; it is that which locates and orients the clause within its context" (Halliday, 2014, p. 89). It is the first nominal, adverbial, or prepositional group or phrase that provides experiential content. Rheme, on the other hand, occurs later in a clause, it is "the part in which the Theme is developed" (ibid). For example, "half of the specimens of each group (n = 10)" in "half of the specimens of each group (n = 10) [Theme] underwent no preloading [Rheme]" (Journal of Prosthodontics, 2018) is the topical Theme and the rest of the clause is the Rheme.

The initial elements of a clause can be topical (i.e., experiential), interpersonal, and textual. Whereas the latter two elements are optional, the topical Theme is obligatory, and it is realized by a participant (nominal group), a circumstance (prepositional or adverbial phrase), or even a process (verb). For example, "17% (7) of RCTs" in "but only 17% (7) of RCTs [Theme] reported all 4 parameters [Rheme]" (Journal of Prosthodontics, 2018) is a topical Theme and the rest of the clause "reported all 4 parameters" is the Rheme. The two elements 'but' and 'only' are, respectively, textual and interpersonal Themes. Interpersonal Themes express a writer's opinion, while textual Themes develop grammatical and logico-semantic links within and beyond a clause (Halliday, 2014).

The second classification of thematic choices is the unmarked/marked contrast. The unmarked Theme is the typical way of starting the message of a clause, while the marked Theme is the atypical ('unusual') choice rarely employed by writers. At the clause level, the unmarked Theme in declarative statements is realized by the Subject[2] (e.g. "the results [Theme] indicate that … [Rheme]") and in marked cases by fronted complements and circumstantial adjuncts. Complements are less commonly used in academic writing than circumstantial adjuncts, and they are realized by a nominal/adjectival group and have the potential to be Subjects but they are not (e.g. "this responsibility [Theme] we accept wholly [Rheme]") (Halliday, 2014, p. 99). Circumstantial adjuncts, on the other hand, are more commonly employed, and they are realized by an adverbial or prepositional group/phrase. For example, the marked Theme "for the type of preloading" in "for the type of preloading [Theme] significant differences were observed (p = 0.02; f = 5.24) between the Lava Plus group and

[2] The subject is the constituent about which something is predicated, i.e. the grammatical subject when the verb is in active form and the object in passive.

the other groups [Rheme]" (Journal of Prosthodontics, 2018) is a circumstantial adjunct realized by a prepositional phrase. One function of marked Themes is announcing a change of topic, guiding readers through the text, and showing them how to interpret the ideas being presented (Mauranen, 1996). Marked Themes then have a facilitative effect as they render the message unambiguous by setting the scene for the clause carrying the message; most writers of RA abstracts, however, minimally employ this thematic choice type due to word limit constraints.

Whereas thematic choices establish cohesion within a clause, TP patterns of Theme-Rheme can develop cohesion beyond the clause. Daneš (1974) was the first to introduce the TP model adopted by many researchers. Theme-Rheme patterns contribute to maintaining a naturally flowing text in three possible ways. One is through linear (or zig-zag/sequential) development in which the Rheme of one clause is introduced as the Theme of the subsequent clause, e.g., The sample [Theme] included 134 subjects [Rheme] .../// All subjects [Theme] had six intact maxillary anterior teeth [Rheme] (Journal of Prosthodontics-Implant Esthetic & Reconstructive Dentistry, 2019). Authors choose to present ideas in this way to create a text that seems to flow logically. A second TP pattern, constant, is where the same Theme is reiterated (or repeated) in subsequent Themes, while Rhemes change (e.g., "A total of 44 adult patients with high, moderate and low caries risk with 516 incipient caries [Theme] took part in the study. [Rheme]// These patients [Theme] were assessed for caries with International Caries Detection and Assessment System (ICDAS) criteria [Rheme]" (Operative Dentistry, 2019). Maintaining the same entity as the Theme of more than one clause develops a topic-focused text if it is moderately employed. A third TP pattern is the split Rheme (or multiple-Theme) pattern. It is when the Rheme, mostly of a paragraph-initial clause, carries more than one idea that is then developed one by one in subsequent Themes (or clauses), as shown in Table 1.

Table 1
An example of the split Rheme pattern

Theme	Rheme
The trial	included 68 periodontal patients with dental anxiety scores of ⩾13
who	were randomly assigned to either an intervention group or a control group (n = 34 per group).
The intervention group	was administered progressive muscle relaxation therapy for 20 min and oral health education for 15 min before periodontal treatment once per week for 4 wks.
The control group	was provided with oral health education only, for the same duration.

Note. Progressive muscle relaxation therapy to relieve dental anxiety: a randomized controlled trial," by E. S. Park, H. W. Yim, and K. S. Lee, 2019, *European Journal of Oral Sciences*, 127(1), p. 45 (https://doi.org/10.1111/eos.12585). Copyright 2019 by Wiley.

The split Rheme pattern is considered "the most difficult for students as different pieces of information are packed or listed in the Rheme position and then picked up and used as Themes in the following clauses" (Alyousef, 2020, p. 276)

Literature Review

Whereas few studies investigated thematic choices and TP patterns in RA abstracts from one discipline (Lorés, 2004; Nwogu & Bloor, 1991), others conducted interdisciplinary studies of variations and similarities in the thematic organization of RA abstracts across a range of disciplines (Alotaibi, 2020; Ebrahimi, 2016; Ebrahimi & Khedri, 2011; Ghadessy, 1999; Leong, 2016).

Nwogu and Bloor (1991) investigated TP in medical RA abstracts and the findings revealed that constant and linear TP patterns were evenly employed. Lorés (2004) analyzed the rhetorical structure and thematic organization in thirty-six applied linguistics RA abstracts and the findings indicated that RA abstracts tend to combine linear and constant TP patterns, with a tendency for the Results/Discussion sections within the abstract to employ linear pattern.

Several studies investigated interdisciplinary variations in the thematic organization of RA abstracts across a range of disciplines. For example, Ghadessy (1999) conducted an interdisciplinary study of Theme

in one hundred and fifty RA abstracts written by native and non-native students from thirty disciplines at Hong Kong Baptist University. The selection of five RA abstracts in each discipline, however, is too small to draw reliable conclusions. Moreover, the dentistry discipline was not among the thirty disciplines. The study of Theme in dentistry RA abstracts does not appear to have attracted this kind of research. Ghadessy's (1999) findings showed that unmarked topical Theme was the most frequently used type in all of the sample abstracts, with sociology having the most (100%) and film and cinema the fewest (70.6%). As stated in Section 2, this finding is not surprising since the unmarked Theme is the typical way of starting the message of a clause. Temporal circumstantial adverbials were more frequently used in thematic position than location adverbials. Elliptical Themes were found in most of the abstracts. In their small-scale study, Ebrahimi and Khedri (2011) investigated thematicity in ten RA abstracts from two disciplines: chemical engineering and applied linguistics. The findings revealed that the unmarked topical Theme was over-used in the RA abstracts of both disciplines. Chemical engineering writers employed more interpersonal Themes (%14) than did those of applied linguistics (%5). Besides, the chemical engineering RA abstracts lacked instances of marked Themes. This finding could be attributed to the researchers' use of a small data set, only five RA abstracts in each discipline. In another study, Ebrahimi (2016) conducted a cross-disciplinary study of variations in the use of Theme in one hundred and twenty RA abstracts from four disciplines: applied linguistics, economics, agriculture, and applied physics. Unmarked topical Theme was also the most frequently used type, followed by the textual Theme. The marked topical Themes were more used in soft disciplines than in hard disciplines. Ebrahimi (2016, p. 108) attributes the lack of interpersonal Theme to the "argumentative nature and impersonal tone of the RA abstract genre." The findings confirm the presence of disciplinary variations in the use of Theme. Alotaibi (2020) argues that interpersonal Themes are more favored in soft disciplines than in hard disciplines. This indicates that the sciences favor a more direct style of writing. Employing Halliday's framework, Leong (2016) examined if there were any differences in the thematic organization of two hundred RA abstracts from two disciplines: sciences and the humanities. The sciences, however, did not include dentistry. The linear TP pattern was the general structure in both disciplines; the two disciplines, however, differed in the clausal distribution of topical Themes. The authors of the science RA abstracts employed more topical Themes and fewer embedded-clause topical Themes than those of the humanities. Alotaibi (2020) investigated interdisciplinary variations in the use of thematic choices in one hundred and eighty RA abstracts from six disciplines: business administration, applied linguistics, accounting, physics, chemistry, and computer science. The findings indicated that the unmarked topical Theme was the most frequently used type in all sets of RA abstracts, while the interpersonal Themes were the least frequent. Some interdisciplinary variations occurred only in textual Theme type, such as the common use of adversative conjunctions in all the disciplines. Although conjunctive adjuncts were used more than conjunctions in all six disciplines, computer science heavily employed conjunctive adjuncts (91.8%) and Applied Linguistics employed conjunctions and conjunctive adjuncts almost equally.

The findings in the aforementioned studies indicate that a discipline's requirements (i.e. rules and conventions) influence the choice of a particular Theme type or pattern. What motivated the present study is the lack of similar investigations of intradisciplinary variations in thematic choices and TP patterns in RA abstracts from dentistry subdisciplines.

Materials and Methods

Data

The sample comprised 119 RA abstracts (28,847 words) of empirical studies published by dental practice scholars (Table 2) in seven dentistry subdisciplines. As the researcher was not a member of this community of practice, a practicing dentist provided support for the selection of the major dentistry subdisciplines and the selection of reputable journals with high impact. Also, since most intradisciplinary linguistic studies of RA abstracts select 12-20 abstracts in each subdiscipline, 17 RA abstracts in each subdiscipline were deemed sufficient to provide invaluable insights for dentistry academic practice. Data sampling was thus based on the idea that larger data sets would yield more reliable findings. The data selection criterion of whether the RAs were written by native or non-native authors was not taken into consideration since the RAs were published in high-ranking journals.

Table 2
Summary of data and journals' impact factor

	Dentistry Subdiscipline	Journal	Impact Factor	No. of RAs	Word Count
1	Oral Sciences	European Journal of Oral Sciences	2.220 (2019)	17	3,668
2	Periodontics	International Journal of Periodontics & Restorative Dentistry	1.513 (2019)	17	2,688
3	Endodontics	Journal of Endodontics	3.118 (2019)	17	4,524
4	Operative Dentistry	Operative Dentistry	2.213 (2019)	17	4,716
5	Prosthodontics	Journal of Prosthodontics- Implant Esthetic & Reconstructive Dentistry	2.187 (2019)	17	4,912
6	Oral & Maxillofacial Surgery	British Journal of Oral & Maxillofacial Surgery	1.061 (2019)	17	3,448
7	Orthodontics	European Journal of Orthodontics	2.202 (2019)	17	4,891
			Total	**119**	**28,847**

The 119 RAs were published during the years 2018-2019. The abstracts were downloaded from the databases and then converted from pdf format into txt format. The abstract sections were then extracted and saved in a separate Microsoft Word file. The abstracts were carefully checked if there were any conversion errors. It should be noted that one journal covers more than one subdiscipline, the International Journal of Periodontics and Restorative Dentistry.

Procedures and Instrumentation

The unit of analysis in this study was the T-unit (Forey & Sampson, 2017; McCabe, 1999), rather than the clause or the sentence since it has been theoretically established that the T-unit is the optimal unit for capturing TP patterns. The choice was also prompted by the fact that previous studies of thematic development in RAs selected the T-unit as their unit of analysis (Ebrahimi, 2016; Jalilifar, 2010; Leong, 2015; Leong, Toh, & Chin, 2018; Williams, 2009).

The T-unit includes an independent clause plus one or more dependent clauses. Dependent clauses include embedded clauses, such as noun clauses and relative clauses. Young (1995) gives some examples of what a T-unit is and is not:

> The following elements were counted as one T-unit: a single clause, a matrix plus subordinate clause, two or more phrases in apposition, and fragments of clauses produced by ellipsis. Co-ordinate clauses were counted as two t-units. Elements not counted as t-units include backchannel cues such as mhm and yeah, and discourse boundary markers such as okay, thanks or good. False starts were integrated into the following t-unit (Young, 1995, p. 38).

Fries (1995, p. 318) defines a T-unit as "a clause complex which contains one main independent clause together with all the hypotactic clauses which are dependent on it." Thus, there will be two T-units if a sentence includes two independent clauses.

The steps in analyzing the data were as follows. T-units were calculated using the Web-based L2 Syntactic Complexity Analyzer (Ai & Lu, 2013; Lu, 2010, 2011; Lu & Ai, 2015), Single Mode. Each independent clause in the RAs was identified and recorded and the Theme of each independent clause was classified. Each topical Theme was coded as marked or unmarked. Next, the TP patterns were identified. All of the features identified in the dentistry RA abstracts were reported both qualitatively and quantitatively. Since the RA abstracts for dentistry subdiscipline had different lengths in relation to their effect on the frequency count of the T-units, the frequency counts for Themes based on T-units can be inflated by RA text length. To account for the differing number of T-units, the total number of occurrences of a Theme type/TP pattern in a given subdiscipline was divided by the total number of occurrences of all Theme types/TP patterns in

each subdiscipline and then the result was multiplied by 100. This in turn will facilitate comparability across the seven dentistry subdisciplines regarding variations or similarities in the choice of Theme and TP patterns. Furthermore, a chi-square test was run, employing Preacher's online chi-square Test Calculator[3] to investigate if there were significant differences between the frequencies of occurrence of the three thematic aspects (Theme types, thematic markedness, and TP patterns) in the seven dentistry subdisciplines.

Recognition of Theme-Rheme and TP Patterns: Delimitations

The annotation of Theme-Rheme is not always straightforward as there are issues in the delimitation of Theme, such as the cases of marked and multiple topical Themes in clause complexes: What counts as a Theme? Where does Theme stop and Rheme begin? Halliday's (2014) and other scholars' (Hasan & Fries, 1995; McGregor, 1992) positions in this regard were adopted since such analysis renders the method of development of the message clearer as the clause identifies the point of departure.

The identification of a Theme depends on the position of the dependent clause in the complex T-unit. If it occurs initially, the entire clause is considered the Theme; if the independent clause occurs initially, the grammatical subject is the Theme. For example, the circumstantial prepositional clause in the following sentence is the Theme of the clause complex:

After 6 months,	all groups showed similar total MMP activity, dry mass loss, and HYP release, // (*European Journal of Oral Sciences, 2019, 127*, 1–9)
Theme	Rheme

Since the clause "after 6 months" has the ideational function of temporal circumstance, it is a marked topical Theme. A Theme is realized here by the elements preceding the grammatical subject. According to Hasan and Fries (1995, p. xxxvi), "everything up to and including the element Subject is Theme so long as there is no marked Topical Theme." Thus, the Subject is excluded in the case of marked topical Theme since it has a Mood function other than Subject.

Elliptical and structural-topical Themes, such as 'that' in relative clauses and 'which/who,' conflate with the Subject. In the case of empty grammatical subjects ('it' and existential clauses), the finite (verb) is regarded as the topical Theme. This contradicts the Hallidayan approach, which marks empty subjects as the Theme. Leong (2004, pp. 202-203) argues that since empty subjects such as "there" and "it" do not carry experiential content, it would be "unacceptable, both theoretically and intuitively, since this amounts to saying that the rheme develops nothing."

Following Thompson (2004), anticipatory "it" and the projecting clause are treated as thematized comments (or interpersonal projections) of a writer's claims. The dummy "it" (or place-holder) and the projecting clause "it is very important" in "it is very important to prevent tooth decay from eroding" are annotated as the Theme since the method of development will be obscured if only "it" is selected.

Results and Discussion

This section attempts to investigate and discuss if there are any significant intradisciplinary variations or similarities in the choice of Theme and TP patterns in the RA abstracts from seven dentistry subdisciplines: oral sciences, periodontics, endodontics, operative dentistry, prosthodontics, oral and maxillofacial surgery, and orthodontics. Table 3 shows the frequencies and percentages and the chi-square distribution of the three thematic aspects (Theme types, thematic markedness, and TP patterns) in the seven dentistry subdisciplines at both local and global levels.

[3] Preacher, K. J. (2001, April). Calculation for the chi-square test: An interactive calculation tool for chi-square tests of goodness of fit and independence [Computer software]. http://www.quantpsy.org/chisq/chisq.htm.

Table 3
The frequency and percentage of three thematic aspects in the seven subdisciplines

		Oral Sciences		Perio-dontics		Endo-dontics		Operative Dentistry		Prostho-dontics		Oral & Maxillofacial Surgery		Ortho-dontics		Chi-square	P-Value
		Freq	%	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%		
	Words	3668		2688		4524		4716		4912		3448		4891			
	T-Units	194	100	132	100	230	100	229	100	242	100	176	100	263	100		
Theme Types	Topical	194	85.46	132	86.84	230	90.55	229	89.11	242	87.68	176	87.56	263	93.59		
	Textual	33	14.54	19	12.50	24	9.45	28	10.89	30	10.87	24	11.94	15	5.34		
	Interpers.	0	0.00	1	.66	0	0.00	0	0.00	4	1.45	1	.50	3	1.07		
		227	100	152	100	254	100	257	100	276	100	201	100	281	100	22.806	0.0294
Thematic Markedness	Un-marked	179	92.27	113	85.60	204	88.69	195	85.15	216	89.26	154	87.50	244	92.78		
	Marked	15	7.73	19	14.40	26	11.31	34	14.85	26	10.74	22	12.50	19	7.22		
		194	100	132	100	230	100	229	100	242	100	176	100	263	100	11.419	0.0762
Thematic progression	Linear	33	63.46	22	64.71	31	50	33	57.89	26	45.62	18	52.94	15	45.46		
	Constant	16	30.77	12	35.29	30	48.39	22	38.60	28	49.12	14	41.18	16	48.48		
	Split Rheme	3	5.77	0	0.00	1	1.61	2	3.51	3	5.26	2	5.88	2	6.06		
		52	100	34	100	62	100	57	100	57	100	34	100	33	100	10.436	0.5777

A detailed presentation and discussion of the findings is presented below in terms of local and global levels organization of Theme in the seven dentistry subdisciplines.

Local Organization

The chi-square test finding for the use of Theme types (df=12) indicates significant differences between the seven dentistry subdisciplines at $p < .05$ (p-value=0.0294), whereas there were no significant differences in the use of thematic markedness between the seven subdisciplines since the p-value was larger than .05 (p-value=0.0762, df=11.419).

Topical Theme

The results showed that topical Theme was the most frequently used Theme type in the RA abstracts of all seven dentistry subdisciplines (Table 3). Between 85.46% and 90.55% of all the Theme types occurring in the dentistry RA abstracts were topical Themes. This indicates the authors' awareness that maintaining the same entity (or topic) as the Theme of more than one clause develops a topic-focused text. This finding is in line with many studies (Alotaibi, 2020; Ebrahimi, 2016; Ebrahimi & Khedri, 2011; Ghadessy, 1999; Leong, 2016). For example, Leong (2016) investigated 200 RA abstracts from science and humanities disciplines and found that topical Themes were used more in scientific abstracts. As stated in Section 2, this finding is not surprising since the unmarked Theme is the typical way of starting the message of a clause.

Another interesting finding is the minimal occurrence of what I call a stand-alone (or one-word) topical Theme (1.31-5.32) in three of the seven dentistry subdisciplines (Table 4). The authors preferred to highlight the topical Theme by using the heading "Objectives." These instances mostly occurred at the beginning of the RA abstracts where the authors presented the 'background/objectives' of their study without repeating the whole Theme: "Objectives [Theme] To compare Oral health-related quality of life (OHRQoL) among 9-year-old children with excessive overjet (EO) to children with unilateral posterior crossbite (UPC) and children with normal occlusion (NO)" [Rheme] // (European Journal of Orthodontics, 2019).

Table 4
The frequency and percentage of complete and stand-alone topical Themes in the dentistry RA abstracts

Subdiscipline	Complete Theme		One-word Theme		Topical Themes	
	Freq	%	Freq	%	Freq	%
Oral Sciences	194	100	0	0.00	194	100
Periodontics	132	100	0	0.00	132	100
Endodontics	230	100	0	0.00	230	100
Operative Dentistry	226	98.69	3	1.31	229	89.11
Prosthodontics	231	95.46	11	4.54	242	87.68
Oral & Maxillofacial Surgery	176	100	0	0.00	176	100
Orthodontics	249	94.68	14	5.32	263	93.59

The required word limit is maintained in the dentistry RA abstract through the use of the stand-alone topical Theme "Objectives" that replaced the phrase "the objectives of the present study." As Tullu (2019, p. S14) states, since "abstracts are the only sections of the research paper that are often freely available to the readers on the journal websites," they need to be "simple, direct, specific, functional, clear, unbiased, honest, concise, precise, self-sufficient, complete, comprehensive, scholarly, balanced, and should not be misleading."

All the authors employed passive voice in their RA abstracts (Table 5) since the use of this construction is a characteristic feature of medical discourse (Amdur, Kirwan, & Morris, 2010). Amdur, Kirwan, and Morris (2010, p. 102) state that medical writers employ passive constructions for five reasons: "to reflect objectivity, to avoid first-person pronouns, to appear scholarly and sophisticated, to avoid responsibility, and to conform to the established writing style." Writers employed passive constructions to background their role as agents (or doers) of the action. The topical Themes in Table 5 are conflated with the subject although they are the object at the syntactic level. For example, although the clause "Demineralized dentin beams" has the participant role of Goal rather than the Actor, it is the subject at the semantic level, whereas it is the object at the syntactic level.

Table 5
Instances of topical Themes in the dentistry RA abstracts

Subdiscipline	Textual Theme	Topical Theme	Rheme
Oral Sciences		Demineralized dentin beams	were divided into groups (n = 10 in each group) //
	and -		incubated in artificial saliva (AS, control) ... //
Orthodontics		The study	would be strengthened
	if	a longitudinal design	had been performed. //
Prosthodontics		CATIA 3D evaluation software	was used to assess the digital STL files ... //
Operative Dentistry		One hundred sixty-five NCCLs	were selected and randomly assigned to three groups. //
Oral & Maxillofacial Surgery		Patients who had palliative procedures	were excluded. //

One of the authors in the Oral Sciences subdiscipline employed an elliptical topical Theme by using the structural conjunct 'and.'

Halliday (2014, p. 87) argues that "all deictic elements are characteristically thematic." A closer look at the elements selected as topical Themes in the dentistry RA abstracts reveals the occurrence of clause-initial deictic determiners (or demonstratives) 'this/these' in the subdisciplines periodontics, restorative dentistry, oral sciences, and endodontics (Frequency, f, = 11-17)), whereas the frequency of occurrence of initial deictic determiners in the other subdisciplines ranged 3-8 (Table 6).

Table 6
Examples of initial deictic determiner "this" as topical Theme[4]

Topical Theme	Rheme
"This recognition	leads to a significant increase in bacteria phagocytosis." // (Journal of Endodontics, 2019)
"This study	aimed to evaluate the effect of different interface designs on the loadbearing capacity of bilayered composite structures (BLS)." // (European Journal of Oral Sciences, 2019)
"This case series	aimed to clinically and histologically evaluate porcine-derived membrane …". // (International Journal of Periodontics & Restorative Dentistry, 2019)

These deictic elements are the most commonly employed discourse deixis in RAs as they contribute to cohesion and coherence of a text. Most of the instances of the demonstrative (or deictic determiner) "this" co-occurred with the Themes "study," "case," "findings," and "results" when presenting the aims, as in "This retrospective cohort study [Theme] aimed to observe … [Rheme]," "This case series [Theme] aimed to … [Rheme]" and "In this observational series, [Theme] three patients … [Rheme]." Other deictic elements that occurred in the dentistry subdisciplines included the deictic relatives "that," "which," and "who," particularly in the dentistry subdisciplines of orthodontics, prosthodontics, and oral and maxillofacial surgery (f = 18-49), whereas its occurrence in the other subdisciplines ranged 6-8.

Another topical Theme observed in the dentistry RA abstracts is the first-person plural pronouns 'we/our', which were employed in five out of the seven dentistry subdisciplines. The use of this form indicates an authorial stance and increased author visibility (Ebrahimi & Chan, 2015; Ebrahimi, Chan, & Ain, 2014; Leong et al., 2018; McKenna, 1997). Whereas this form occurred frequently in the oral and maxillofacial surgery subdiscipline (47 instances), it rarely occurred in the other four subdisciplines (2-8 instances). This suggests oral and maxillofacial surgery scholars' preference to exhibit greater visibility than their counterparts in the other dentistry subdisciplines. This could also be attributed to the peculiar epistemologies of oral and maxillofacial surgery that differentiate it from the other subdisciplines. Researchers argue that a further function of thematized "we" is that it enables the authors to emphasize the novelty of their proposed work (Ebrahimi & Chan, 2015; Martínez, 2003), which seems to be the case in the following excerpts.

(1) "We therefore suggest that it is used when discussing surgical outcomes, taking consent from newly-diagnosed patients, and when calculating the costs of head and neck oncological surgery."/ "We evaluated the degree of autophagy and cytoplasmic HMGB1 in radioresistant oral squamous cell carcinoma."/ "In this prospective study, [Theme] we meticulously searched for more examples using both formalin-fixed cadavers and neck dissections. [Rheme]"/ "We retrospectively reviewed the impact of repair of cleft lip on marriage in adult patients in our institution by questioning a group of patients treated from July 2009-June 2017." (British Journal of Oral & Maxillofacial Surgery, 2019)

The use of the first-person plural pronoun "we" was more frequent than the possessive "our" in the other six subdisciplines. This contradicts Behnam, Mirzapour, and Mozaheb's (2014) study of Chemistry RAs, which showed that writers in hard sciences use the first-person plural possessive "our" more than "we" because they are trying to reduce the responsibility of their claim. The dental authors used the first-person plural pronoun "we" to show their contributions to their field of research. This result also refutes the claim that personal pronouns are very rarely employed in hard disciplines (Hyland, 2005) as quantitative research must be objectives and impersonal. Kuo (1999) states that writers use the exclusive "we" to express their role and contribution to their field of research. The frequent use of the first-person plural pronoun "we" converges with several studies of personal pronouns in linguistics (Molino, 2010), applied linguistics (Dobakhti & Hassan, 2017), and computer science, electronic engineering, and physics (Kuo, 1999). For example, Dobakhti and Hassan (2017) investigated authorial presence in Applied Linguistics RAs and found that plural pronouns are more common in quantitative than qualitative RAs. These studies, however, attributed the use of first-person plural pronouns not only to multiple authorship (exclusive "we") but also to involve the author and the reader (inclusive "we") in the argument. The RA abstracts of the periodontics, restorative dentistry, and prosthodontics subdisciplines lacked instances of the first-person plural pronoun

[4] Double slashed lines indicate the end of a T-unit.

"we" since members of this discourse community seem to have preferred employing the traditional conventions of academic writing by retaining an objective stance through the use of passive constructions so that their abstracts appear more professional. As all the RA abstracts were written by multiple authors, none of the seven dentistry subdisciplines included instances of first-person singular pronouns. The use of first-person plural pronouns in five out of the seven dentistry subdisciplines reflects the collaborative nature of these subdisciplines where multiple authors are involved (or foregrounded as agents). It also indicates the authors' confidence in the propositions or the claims they are making.

Textual Theme

The results showed that textual Theme was the second most frequently used Theme type in the RA abstracts of all the seven dentistry subdisciplines (Table 3). The frequency of this Theme type ranged from 5.34% to 14.54% A textual Theme is employed to set grammatical and/or semantic relations within a clause using conjunctions (e.g. so, but, or), while logico-semantic relationships between a clause and the preceding text were established using conjunctive adjuncts of addition, apposition, and other semantic relations (e.g. however, while, on the other hand). Over half of the instances of the textual elements were 'and' followed by 'however.' Conjunctive adjuncts are employed in RA abstracts to explicitly advance arguments and highlight the importance of the study. This finding is in line with Alotaibi's (2020) study of Theme in computer science, which indicated the heavy use of conjunctive adjuncts. Table 7 shows instances of the authors' ways of realizing causal (therefore, consequently), adversative (however), and additive (additionally, moreover) meanings through conjunctive adjuncts.

Table 7

Examples of textual Theme in the dentistry RA abstracts

Subdiscipline	Textual Theme	Interpersonal	Topical Theme	Rheme
Oral Sciences	However,		high concentrations of tHA	stimulate production of reactive oxygen species (ROS), resulting in cell injury and apoptosis. //
Periodontics	However,	very little	evidence of long-term results	is available. //
Orthodontics	Therefore,		in regular orthodontic cases,	the choice between anchoring devices may rely on other factors, such as costs, patient comfort, personal preferences, familiarity with the device, and insertion procedures. //
Endodontics	Additionally,		in the CBCT images,	the number of apical foramina was also evaluated. //
Operative Dentistry	Moreover,		the 1.5% hydrogen peroxide– or essential oil-based mouth rinses	affected the bleached enamel properties, promoting an alteration in morphologic surface and mineral loss in depth. //

Similar conclusions were drawn in previous research showing that retrospective cohesive devices tend to occur in argumentative RA sections to persuade readers of the validity of claims (Gosden, 1992; Jalilifar, 2010).

Interpersonal Theme

The use of interpersonal Theme in dentistry RA abstracts was negligible. This finding converges with several studies of Theme in RA abstracts (e.g. Alotaibi, 2020; Ebrahimi, 2016; Ghadessy, 1999). The low percentage of the interpersonal Theme type is in line with what would be expected in objective, non-evaluative scientific RA discourse (Gray & Biber, 2012; Hyland, 2005). Except for the orthodontics subdiscipline, the other six dentistry RA abstracts included instances of impersonal projections that expanded the meaning potential in a clause, as they created a logical relation between Theme-Rheme (e.g., 'show,' 'suggest,' and 'indicate'): "Our results show that …." These projections yield a clear and coherent line of argument. Halliday (2004, p. 155) argues that "as scientific discourse has come to be depersonalized, during the past hundred years or so, impersonal projections have tended to be increasingly hedged around."

Rather than using interpersonal Themes, the dentistry scholars employed modality (e.g. should, must), as in "Clasps composed of zirconia *should* improve aesthetics without inducing the risk of metal allergy //" (European

Journal of Oral Sciences, 2019). Stotesbury (2003) believed that while the humanities and social sciences employ more evaluative attributes (e.g. Interpersonal Themes), natural sciences more often resort to modality to express authorial stance.

Marked & Unmarked Themes

The findings revealed that unmarked Themes are far more prevalent than marked Themes in the RA abstracts of all the seven dentistry subdisciplines (Table 3). The distribution of thematic markedness was almost equal across the seven dentistry subdisciplines, as evidenced by the chi-square test finding that showed no significant differences in the use of thematic markedness across the seven subdisciplines (p-value=0.0762, df=12 & 6). The frequency of unmarked Themes ranged between 85.15% and 92.78%, whereas that of marked ranged 7.22%- 14.85% (Table 3). The high frequency of unmarked Theme indicates the authors' preference for maintaining the same topic rather than shifting focus through the frequent use of marked Themes (e.g., "The remake rate for crowns [Unmarked Theme] was 3.8% [Rheme]" (Journal of Prosthodontics- Implant Esthetic & Reconstructive Dentistry, 2019). The authors' preference for employing the typical(or usual) way of starting the message in a clause (i.e. using the unmarked Theme) is expected since it saves text space. The frequent use of unmarked Themes could be attributed to the relatively compact nature of the RA abstract. This finding is in line with the majority of studies (e.g. Alotaibi, 2020; Ebrahimi, 2016; Ebrahimi & Khedri, 2011; Ghadessy, 1999; He & Tian, 2018; Leong, 2016) which found that unmarked Themes outnumbered the marked Themes. For example, Ebrahimi and Khedri (2011) found that unmarked Themes were over-used in RA abstracts of both Chemical Engineering and Applied Linguistics.

Marked Themes are employed in the prosthodontics subdiscipline to validate claims and to provide information related to time or space, as shown in Excerpts 2-5 below.

(1) "According to the results of this study, [Marked Theme] the digital implant impression technique had the least accuracy." [Rheme] //
(2) "After that, [Marked Theme] the splinted assemblies were attached to implant analogs (DANSE) …" [Rheme] //
(3) "After bonding, [Marked Theme] the shear bond strength (SBS) of the ceramic to dentin was measured." [Rheme] //
(4) "In the CRC group, [Marked Theme] 75% of the failures were of the adhesive type, compared with 66.7% and 83.3% in the LRC and LGC groups, respectively." [Rheme] // (Journal of Prosthodontics, 2019)

The adverbial group and the prepositional phrases (or circumstantial adjuncts) function as Theme as they can be moved into Rheme position (backgrounded); both the Theme and Rheme are called agnate clauses since they are reversible. The circumstances in the above excerpts conflate with the Subject (or take its position) to make texts more elaborate. Fronting the circumstantial elements allows authors to place the study within a wider context and thereby motivate the study. Davies (1989) states that a marked Theme is mostly employed in scientific RAs to validate some evidence (e.g. according to our findings), a writer's viewpoint (Surprisingly), and time/space (in brief, in figure x). Alyousef (2020) argues that a Theme is emphasized through intonation when it is moved into the Rheme position and replaced with adverbial or prepositional phrases.

The first part of this section examined within-clause thematic choices. What follows is the global thematic analysis of the dentistry RA abstracts to reveal the authors' preferred ways of sequencing information across the clause.

Global Organization

A combination of linear and constant Theme patterns constituted over 90% of the TP patterns in the seven dentistry subdisciplines (Table 3). This finding is in line with the majority of studies on RA abstracts in other disciplines (e.g. Lorés, 2004; Nwogu & Bloor, 1991). The chi-square test finding for the use of TP patterns indicated no significant differences across the seven dentistry subdisciplines (p-value=0.5777, df=10.436).

Linear TP pattern

The linear (or zig-zag) TP pattern was the most dominant pattern in five out of the seven subdisciplines, as the prosthodontics and the orthodontics RA abstracts had slightly more instances of constant TP patterns than linear. The occurrence of the linear TP pattern in the seven subdisciplines comprised 45.46% - 64.71% of all the

TP patterns. The dominance of this pattern corresponds with studies of RA abstracts in sciences and humanities (Leong, 2016) and applied linguistics (Ebrahimi, 2016; Lorés, 2004). The high frequency of this pattern can be explained by the fact that this pattern contributes to a hierarchically structured cohesive text when information in the Rheme position is linked and developed in the subsequent Theme, thereby creating a clear and coherent line of argument that acts as a lead for readers. As Halliday (1993, p. 131) states, scientific texts tend to develop a linear "chain of reasoning (ultimately based on experiments) in which each step led on to the next." The skillful employment of this pattern is illustrated in Table 8.

Table 8
Examples of linear TP pattern in the seven dentistry RA abstracts

Subdiscipline	Theme	Rheme
Oral Sciences	The marginal signal near 4q31 detected in previous genome-wide association studies	might be caused by an interaction between the IRF6 and GLRB genes.
	This interaction	needs to be further validated by experimentation in follow-up studies.
Periodontics & Restorative Dentistry	Microcomputed tomography and histologic and histomorphometric analyses	were performed to determine bone density.
	Bone density	increased significantly over time (P < .001), with stabilization between 12 and 24 weeks (P = 1.000). //
Endodontics	Our aim	was to determine their relevance to infected pulp and how their components affect human dental pulp cell (HDPC) responses. //
	Diseased pulp tissue	was stained for the presence of extracellular DNA and elastase to detect the presence of NETs. //
Prosthodontics	The sample	included 134 subjects—students of the University of Rijeka, Croatia (65% female) aged 19 to 28 years (median 21). //
	All subjects	had six intact maxillary anterior teeth without restorations or severe malocclusions and healthy gingiva with no signs of inflammation. //
Oral & Maxillofacial Surgery	Between December 2015 and June 2018 44	patients had skull SPECT (15 male and 29 female patients, whose ages ranged from 4-33 years). //
	They	were having bone scans … //
Orthodontics	Objectives	To investigate the changes in oral health-related quality of life (OHRQoL) before, during, and after orthodontic treatment, … //
	Materials and methods OHRQoL questionnaires	were used in an ongoing observational prospective cohort study at baseline (T0), 1 year after start (T1), and 1 month after the end of active orthodontic treatment (T2). //
Operative Dentistry	At T3,	group CU exhibited the lowest ΔE values (p=0.008). //
	The ΔE values	increased from T4 to T5 in the CW and CU groups, //

Each bolded Theme in Table 8 is derived from the Rheme in the previous clause, giving a strong sense of topic continuity and in turn sustaining the argument being made. The clause-initial deictic determiner 'this' in one of the oral sciences RA abstracts refers back to the clause "an interaction between the IRF6 and GLRB genes." The author uses this element instead of repeating the whole clause to save text space. Similarly, the author in the prosthodontics subdiscipline employs the Theme (head noun) "All subjects" to refer to back to the clause "134 subjects—students of the University of Rijeka, Croatia (65% female) aged 19 to 28 years (median 21)." Structural-topical Themes, such as 'that' in relative clauses and 'which / who,' conflate with the Subject to produce cohesive and elaborated discourse through the use of linear TP pattern, as in "Polydopamine-templated hydroxyapatite (tHA) [Theme] is a type of nano-biomaterial [Rheme] that [Theme] can promote osteogenesis in bone tissue engineering. [Rheme]//" (Oral Sciences Journal, 2019). It should be noted that the demonstratives 'this', 'these' and 'that' seem most often to be used anaphorically referring back to a proposition in the Rheme position (See also Table 6). This pattern aids scholars in producing a characteristically coherent argumentative RA abstract, as they have to convince their readers of the relevance of their proposed work (Swales, 2004).

Constant TP Progression

The second most common TP pattern in the dentistry RA abstracts was the constant pattern, comprising 30.77%-49.12% of all the TP patterns the authors employed as one of the methods to develop information. While all seven dentistry subdisciplines recognized the usefulness of linking clauses by repeating the Theme of a previous clause, the qualitative analysis indicated that they skillfully used it for slightly different purposes. There was no quantitative difference in the use of the constant TP pattern across the seven dentistry subdisciplines; the qualitative analysis, however, revealed a difference in the functions that underlie the use of this pattern (Table 9).

Table 9

Examples of constant TP pattern in the dentistry RA abstracts

Subdiscipline	Textual Theme	Topical Theme	Rheme
Oral & Maxillofacial Oral and Maxillofacial Surgery		We	evaluated the degree of autophagy and cytoplasmic HMGB1 in radioresistant oral squamous cell carcinoma (SCC) ... //
		We	then compared the degree of autophagy between non-irradiated control and radioresistant cancer cells ....//
		We	also compared the total and cytoplasmic concentrations of HMGB1 ... //
		Many indices and scoring systems	exist for assessing skeletal patterns and malocclusion//
	but	none	have been universally adopted by teams providing orthognathic surgery in the UK. //
Oral Sciences		Ninety-six samples of RNHBM	were prepared and exposed for 24 h to 17%, 42.5%, or 85% glycerol, or to [sic] distilled H2O (control). //
		Sections	were stained with haematoxylin and eosin (H&E) ... //
Periodontics		At 24 weeks,	the HA50 group exhibited more NB than the HA100 (P < .001) and control (P = .066) groups. //
		At 24 weeks,	the HA100 and HA75 groups showed high RM percentages. //
Endodontics		One thousand retrospective cone-beam computed tomographic scans	were randomly selected from a private oral radiologic clinic database. //
		The images	were acquired with a limited field of view ... //
		The pain severity	was similar in both groups at 48 hours, //
	and	it	was less severe in the long-acting betamethasone group compared with the dexamethasone group after 72 hours and 7 days. //
Operative Dentistry		Class V composite resin restorations	were performed (Z350, 3M ESPE, St Paul, MN, USA). //
		The restorations	were evaluated at baseline and at six, 12, 24, and 36 months according to the World Dental Federation criteria. //
Prosthodontics & Implant		According to the results of this study,	the digital implant impression technique had the least accuracy. //
		Based on the study outcomes,	distance and angulation errors associated with the intraoral digital implant impressions were too large to fabricate well-fitting restorations ... //
Orthodontics		Miniscrews	provide better anchorage reinforcement at a higher price. //
		They	should be used in cases where anchorage loss cannot be accepted. //

The thematic element "we" (referring to the authors) was successively reiterated across several clauses to recursively present the aims and the methods (Alyousef, in press). The reiterated clause "ninety-six samples of RNHBM (reconstructed normal human buccal mucosa)" in the oral sciences subdiscipline is elided (or omitted) in the following clause and replaced by the word "sections" to refer to the presupposed clause "sections of the RNHBM samples." In the oral and maxillofacial surgery RA abstracts, the word "none" was employed instead of reiterating the clause "Many indices and scoring systems." The use of this complex pattern demonstrates the author's advanced writing skills where each aim is followed by the procedure which keeps the reader focused. Rather than repeating the same Theme "cone-beam computed tomographic scans," one of the authors in the endodontics subdiscipline uses the synonym "images." Similarly, one of the authors in the prosthodontics subdiscipline used the synonym 'outcomes' in "the study outcomes" instead of reiterating the same phrase "the results of this study." Finally, the classifiers (or the words) "Class V composite resin" accompanying the head noun "restorations" in the operative dentistry RA abstract was deleted in the following clause and replaced with "the restorations" to avoid redundancy. The identified functions of this pattern offer authors a strategy to give additional information about a topic closely related to their work. The dental authors demonstrated their advanced writing skills by not only reiterating Theme but also using various writing strategies that cope with word limit constraints, such as omitting the classifiers (or words) accompanying the head noun, the use of synonyms, and the pronouns "it" and "none."

Split Rheme TP Pattern

The split Rheme pattern occurred the least in all seven subdisciplines, comprising 0.00% - 6.06% of all the TP patterns. This finding, however, was expected due to the concise nature of an RA abstract, which requires an author to shift from one proposition to another. The split TP pattern unfolds across many clauses because it carries several pieces of information revolving around one major topic, with each piece taking a thematic position in a separate consequent clause, forming a cluster of topically related clauses. Although there was a quantitative difference in the use of the split pattern across the seven dentistry subdisciplines, the qualitative analysis found a similarity in the functions of this pattern (Table 10).

Table 10

Examples of split Rheme pattern in the dentistry RA abstracts

Subdiscipline	Theme	Rheme
Oral Sciences	Wistar albino rats	were randomized into four groups: control; ALA; AP; and ALA + AP. //
	Rats of the control and ALA groups	were not endodontically treated, //
	but saline and ALA (100 mg kg_1)	were administered. //
	In rats of the AP and ALA + AP groups,	the pulp chambers of mandibular first molar teeth were exposed and left open for 30 d to induce AP. //
Endodontics	Methods: In this double-blind randomized clinical trial,	242 healthy patients with irreversible pulpitis undergoing 1-visit endodontic treatment were included. //
	Forty-five patients	were lost during the follow-up, //
	and the remaining 197 patients	were followed for 7 days (67 patients in the placebo group, 66 in the long-acting betamethasone group, and 64 in the dexamethasone group). //
Operative Dentistry	Participants in each group	were further divided randomly into two subgroups. //
	In subgroups A1, B1, and C1,	an intensive preventive protocol was applied, //
	while in subgroups A2, B2, and C2,	the protocol consisted only of instructions in oral hygiene. //

The authors used a split Rheme pattern to link clauses containing a list of points (or aspects) such as the four groups "control; ALA; AP; and ALA + AP" in the oral sciences subdiscipline.

The skillful use of this pattern establishes cohesion in the text as the elements in the Rheme signal connections with the successive Themes by highlighting the subsequent propositional content. The split Rheme pattern thus functions prospectively as a guide to signal the information coming next. The lack of significant intradisciplinary differences in the use of TP patterns across the seven dentistry subdisciplines could be attributed to the conventional stylistic practice in the dentistry discipline that requires the non-random use of

constant and linear TP patterns. This finding corresponds with Lorés's (2004) and Nwogu and Bloor's (1991) studies which indicated that RA abstracts tend to combine these two patterns. Table 11 summarizes the significant intradisciplinary similarities and differences in the use of thematic choices and TP patterns in the seven subdisciplines.

Table 11

Similarities and differences in the use of thematic choices and TP patterns in the seven subdisciplines

Linguistic Aspect	Function & notes	Example
Topical Theme (most frequently used Theme type in all the seven dentistry subdisciplines).	- Developing a topic-focused text - Most of the instances of the demonstrative (or deictic determiner) "this" co-occurred with the Themes "study," "case," "findings," and "results" when presenting the aims. - Use of first-person plural pronouns "we/our" as Topical Themes to indicate an authorial stance and increased author visibility (highly employed in the Oral and Maxillofacial Surgery subdiscipline).	"The purpose of this retrospective cohort study [Theme] was to radiographically evaluate the outcome of Resilon... [Rheme]" (Journal of Endodontics, 2019).
Textual Theme (second most frequently used Theme type in all the seven dentistry subdisciplines although its occurrence in the Orthodontics subdiscipline was the least).	- Develops grammatical and logico-semantic links within and beyond a clause. - Over half of the instances of the used textual elements were 'and' followed by 'however.' - Conjunctive adjuncts explicitly advance arguments and highlight the importance of the study.	"However, [Textual Theme] high concentrations of tHA [Topical Theme] stimulate production of reactive oxygen species (ROS), resulting in cell injury and apoptosis. [Rheme] //" (European Journal of Oral Sciences, 2019).
Interpersonal Theme (minimally employed in four subdisciplines- e.g. only, most- but all the seven dentistry subdisciplines employed modality, e.g. should, must).	- Expresses a writer's opinion. - Modality expresses an author's authorial stance.	"Clasps composed of zirconia should improve aesthetics without inducing the risk of metal allergy" // (European Journal of Oral Sciences, 2019).
Unmarked Themes (far more prevalent than the marked Themes in all the seven dentistry subdisciplines, 85.15%-92.78%).	- Maintaining the same topic. - Saving text space through the use of Unmarked Themes.	"We [Unmarked Theme] used novel analytical approaches to identify inflammatory response patterns to plaque accumulation in experimental gingivitis studies in humans. [Rheme]" (European Journal of Oral Sciences, 2019).
Marked Theme (minimally employed in all subdisciplines, 7.22%-14.85%).	- To validate claims (e.g. according to our findings), - To provide information related to time and space.	"According to the results of this study, [Marked Theme] the digital implant impression technique had the least accuracy [Rheme]." // "After that, [Marked Theme] the splinted assemblies were attached ... [Rheme]" // (Journal of Prosthodontics, 2019).
Linear (or zig-zag) TP pattern (most dominant pattern in all subdisciplines, 45.46% - 64.71%).	- Creates a clear and coherent line of argument that acts as a lead for readers. - Saving text space through the use of head nouns and demonstratives 'this', 'these', and 'that' to refer back to a proposition in Rheme position.	"In the entire sample, [Theme] 115 (51.57%) patients were females and 108 (48.43%) were males [Rheme]. They [Theme] ranged in age from 28 to 81 years old. [Rheme]" (Journal of Prosthodontics- Implant Esthetic & Reconstructive Dentistry, 2019).
Constant TP pattern (second most common pattern in all subdisciplines, 30.77%-49.12%; highest were Prosthodontics 49.12%, Orthodontics 48.48%, and Endodontics 48.39%).	- Maintain readers' focus on the topic. - Save text space through the use of the pronouns "it" and "none" as well as omitting the words accompanying the head noun. - The use of synonyms (results/ outcomes, tomographic scans/images) to avoid repetition and make the text more cohesive.	"Many indices and scoring systems [Theme] exist for assessing skeletal patterns and malocclusion [Rheme]// but none [Theme] have been universally adopted by teams providing orthognathic surgery in the UK [Rheme]"// (British Journal of Oral & Maxillofacial Surgery, 2019).
Split Rheme pattern (minimally used by six subdisciplines due to word limit constraints, only 1-3 instances in each subdiscipline, except for Periodontics, n=0).	- Signal information coming in subsequent clauses, thereby forming a cohesive text. - Several pieces of information (or ideas/ aspects) revolving around one major topic are introduced in Rheme position, usually at the beginning of a paragraph, and then each piece takes a thematic position in a separate consequent clause.	"Participants in each group [Theme] were further divided randomly into two subgroups. [Rheme]// In subgroups A1, B1, and C1, [Theme 1] an intensive preventive protocol was applied, [Rheme] // while in subgroups A2, B2, and C2, [Theme 2] the protocol consisted only of instructions in oral hygiene [Rheme]"// (Operative Dentistry, 2019).

Conclusion

Several studies have conducted interdisciplinary investigations on variations and similarities in the thematic organization of RA abstracts across a range of disciplines; however, similar investigations of RA abstracts of dentistry subdisciplines are lacking. This research paper is significant as it is the first to investigate the local and global organization of Theme in RA abstracts of seven dentistry subdisciplines. The results highlighted quantitative and qualitative intradisciplinary similarities and differences. The qualitative results showed that even though there were no significant interdisciplinary differences between the dentistry RA abstracts and the findings reported in the literature of other disciplines, various interesting linguistic features characterizing the subdisciplines were employed: elliptical topical Themes, one-word topical Themes (in three subdisciplines), passive constructions to reflect objectivity and professionalism, and clause-initial deictic demonstratives 'this/these' (in three subdisciplines). Oral and maxillofacial surgery authors regularly used first-person pronouns to express their role and contributions to the field of research and indicate increased author visibility.

The study has practical implications for novice dental researchers attempting to write an RA abstract as well as for tutors of English for Academic Purposes courses. One important implication of the results is that thematicity helps academic writers orient their writing at the text level rather than the sentence level through the skillful use of TP patterns. An awareness of the available thematic choices and TP patterns yields a well-structured cohesive and coherent message both within and across the clause. This in turn will facilitate smooth readability and could increase the readership of an abstract. Dental researchers need to consider the intradisciplinary significant similarities and differences in the use of Theme types as well as the key linguistic features when writing an RA abstract in one of the seven subdisciplines. Having observed these important aspects, the flow and organization of a dentistry RA abstract will be more effective and the chances of having a paper accepted in a high-ranking journal will increase.

Future studies could further investigate thematic choices and TP patterns in other sections of dentistry RAs to compare their findings with those in the present study. It is also recommended that future research investigate and compare interdisciplinary variations in the use of Theme and TP patterns across subfields of other disciplines.

Acknowledgments

The author is indebted to the two anonymous reviewers for their insightful and helpful comments. He is also grateful to the editor who read the manuscript and provided invaluable comments as well as the dentist who extracted the terms and the sentences that are incomprehensible to a layperson. The author wishes to express his gratitude to both the Deanship of Scientific Research at King Saud University and the Research Centre at the Faculty of Arts for funding the present study.

References

Ai, H., & Lu, X. (2013). A corpus-based comparison of syntactic complexity in NNS and NS university students' writing. In A. Díaz-Negrillo, N. Ballier, & P. Thompson (Eds.), *Automatic treatment and analysis of learner corpus data* (pp. 249-264). John Benjamins.

Alotaibi, H. S. (2020). The thematic structure in research article abstracts: Variations across disciplines. *Cogent Arts & Humanities, 7*(1), 1756146. https://doi.org/10.1080/23311983.2020.1756146

Alyousef, H. S. (2020). A multimodal discourse analysis of English dentistry texts written by Saudi undergraduate students: A study of theme and information structure. *Open Linguistics, 6*, 267–283. https://doi.org/10.1515/opli-2020-0103

Alyousef, H. S. (in press). A move-structure model for dentistry research article abstracts: A genre-based study of variations and similarities in seven dentistry subdisciplines. *Discourse & Interaction, 14*(1).

Amdur, R. J., Kirwan, J., & Morris, C. G. (2010). Use of the passive voice in medical journal articles. *AMWA Journal: American Medical Writers Association Journal, 25*(3), 98-104. https://doi.org/10.2302/kjm.2014-0009-RE

Behnam, B., Mirzapour, F., & Mozaheb, M. A. (2014). Writer's presence in English native and non-native speaker research articles. *Procedia-Social & Behavioral Sciences, 98*(6), 369-374. https://doi.org/10.1016/j.sbspro.2014.03.428

Daneš, F. (1974). Functional sentence perspective and the organization of the text. In F. Daneš (Ed.), *Papers on functional sentence perspective* (pp. 106-128). Academia.

Davies, F. (1989). *Developing competence in academic discourse: The role of language awareness (ELU Working Paper No. 1)*. University of Liverpool.

Dobakhti, L., & Hassan, N. (2017). A Corpus-based study of writer identity in qualitative and quantitative research articles. *3L: The Southeast Asian Journal of English Language Studies, 23*(1), 1-14. http://doi.org/10.17576/3L-2017-2301-01

Ebrahimi, S. F. (2016). Theme types and patterns in research article abstracts: A cross disciplinary study. *International Journal of English Language & Translation Studies, 4*(3), 104-115.

Ebrahimi, S. F., & Chan, S. H. (2015). Research article abstracts in Applied Linguistics and Economics: Functional analysis of the grammatical subject. *Australian Journal of Linguistics, 35*(4), 381-397. https://doi.org/10.1080/07268602.2015.1070660

Ebrahimi, S. F., Chan, S. H., & Ain, N. A. (2014). Discourse functions of grammatical subject in results and discussion sections of research article across four disciplines. *Journal of Writing Research, 6*(2), 125–140. https://doi.org/10.17239/jowr-2014.06.02.2

Ebrahimi, S. F., & Khedri, M. (2011). Thematicity in research article abstracts: A cross-disciplinary study. *Educational Quest: An International Journal of Education and Applied Social Sciences, 2*(3), 287-292.

Forey, G., & Sampson, N. (2017). Textual metafunction and Theme: What's "it" about? In T. Bartlett & G. O'Grady (Eds.), *The Routledge handbook of systemic functional linguistics* (pp. 131–145). Routledge.

Fries, P. (1995). Themes, methods of development, and texts. In R. Hasan & P. Fries (Eds.), *On subject and theme: A discourse functional perspective* (pp. 317-360). John Benjamins.

Ghadessy, M. (1999). Thematic organization in academic article abstracts. *Estudios ingleses de la universidad complutense, 7*, 141-161.

Gosden, H. (1992). Discourse functions of marked theme in scientific research articles. *English for Specific Purposes, 11*(3), 207-224. https://doi.org/10.1016/S0889-4906(05)80010-9

Gray, B., & Biber, D. (2012). Current conceptions of stance. In K. Hyland, & C. S. Guinda (Eds.), *Stance and voice in written academic genres* (pp. 15-33). Palgrave Macmillan.

Halliday, M. A. K. (1993). The analysis of scientific texts in English and Chinese. In M. A. K. Halliday & J. R. Martin (Eds.), *Writing science: Literacy and discursive power* (pp. 124-132). Falmer Press.

Halliday, M. A. K. (2004). The Language of Science. In J. Webster (Ed.), *Collected works of M.A.K.Halliday* (vol. 5, pp. 162-178). Continuum.

Halliday, M. A. K. (2014). *Introduction to Functional Grammar. Revised by Christian M. I. M. Matthiessen* (4th ed.). Taylor & Francis.

Hasan, R., & Fries, P. H. (1995). Reflections on subject and theme: An introduction. In R. Hasan & P. H. Fries (Eds.), *On subject and theme : A discourse functional perspective* (pp. xiii-xlv). John Benjamins.

He, Y., & Tian, J. (2018). A Comparative study of thematic progression in abstracts of scientific papers: A case study of science and science bulletin. *International Journal of Arts and Commerce, 7*(7), 1-7.

Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies, 7*(2), 173-192. https://doi.org/10.1177/1461445605050365

Jalilifar, A. R. (2010). The status of theme in applied linguistics articles. *Asian ESP Journal, 6*(2), 7–39.

Kuo, C.-H. (1999). The use of personal pronouns: Role relationships in scientific journal articles. *English for Specific Purposes, 18*(2), 121-138. https://doi.org/10.1016/S0889-4906(97)00058-6

Leong, A. P. (2004). *Theme and rheme: An alternative account* (vol. 266). Peter Lang.

Leong, A. P. (2015). Topical themes and thematic progression: The "picture" of research articles. *Text & Talk, 35*(3), 289-315. https://doi.org/10.1515/text-2015-0001

Leong, A. P. (2016). Thematic density of research-article abstracts: A systemic-functional account. *WORD, 62*(4), 209-227. https://doi.org/10.1080/00437956.2016.1248668

Leong, A. P., Toh, A. L. L., & Chin, S. F. (2018). Examining structure in scientific research articles: A study of thematic progression and thematic density. *Written Communication, 35*(3), 286–314. https://doi.org/10.1177/0741088318767378

Lorés, R. (2004). On RA abstracts: From rhetorical structure to thematic organization. *English for Specific Purposes, 23*(3), 280-302. https://doi.org/10.1016/j.esp.2003.06.001

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics, 15*(4), 474-496. https://doi.org/10.1075/ijcl.15.4.02lu

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., & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing, 29*, 16-27. https://doi.org/10.1016/j.jslw.2015.06.003

Martínez, I. A. (2003). Aspects of theme in the method and discussion sections of biology journal articles in English. *Journal of English for Academic Purposes, 2*(2), 103-123. https://doi.org/10.1016/S1475-1585(03)00003-1

Mauranen, A. (1996). Discourse competence: Evidence from thematic development in native and non-native texts. In E. Ventola, & A. Mauranen (Eds.), *Academic Writing: Intercultural and textual issues* (pp. 195-230). John Benjamins.

McCabe, A. M. (1999). *Theme and thematic patterns in Spanish and English history texts* [Unpublished doctoral dissertation]. University of Aston in Birmingham.

McGregor, W. (1992). The place of circumstantials in systemic-functional grammar. In A. Davies, & L. Ravelli (Eds.), *Advances in systemic linguistics: Recent theory and practice* (pp. 136-149). Pinter.

McKenna, B. (1997). How engineers write: An empirical study of engineering report writing. *Applied Linguistics, 18*(2), 189-211. https://doi.org/10.1093/applin/18.2.189

Molino, A. (2010). Personal and impersonal authorial references: A contrastive study of English and Italian Linguistics research articles. *Journal of English for Academic Purposes, 9*(2), 86-101. https://doi.org/10.1016/j.jeap.2010.02.007

Nwogu, K. N., & Bloor, T. (1991). Thematic progression in professional and popular medical texts. In E. Ventola (Ed.), *Functional and systemic linguistics: Approaches and uses* (vol. 55, pp. 369-384). Mouton de Gruyter.

Shamim, T. (2018). A working classification proposed for indices based on dental specialties in dental research. *Mustansiriya Medical Journal, 17*(2), 100-103. https://doi.org/10.4103/MJ.MJ_26_18

Stotesbury, H. (2003). Evaluation in research article abstracts in the narrative and hard sciences. *Journal of English for Academic Purposes, 2*(4), 327-341. https://doi.org/10.1016/S1475-1585(03)00049-3

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

Thompson, G. (2004). *Introducing functional grammar* (2nd ed.). Arnold.

Tullu, M. S. (2019). Writing the title and abstract for a research paper: Being concise, precise, and meticulous is the key. *Saudi Journal of Anesthesia, 13*(1), 12-17. https://doi.org/10.4103/sja.SJA_685_18

Williams, I. A. (2009). Discourse style and theme–rheme progression in biomedical research article discussions: A corpus-based contrastive study of translational and non-translational Spanish. *Languages in Contrast, 9*(2), 225-266. https://doi.org/10.1075/lic.9.2.03wil

Young, R. (1995). Conversational styles in language proficiency interviews. *Language Learning, 45*(1), 3-42. https://doi.org/10.1111/j.1467-1770.1995.tb00961.x

National Research University Higher School of Economics
Journal of Language & Education Volume 7, Issue 1, 2021

Bagci, A., & Aydin , B. (2021). Autonomy Orientations of English Language Teachers in Turkey.. *Journal of Language and Education, 7*(1), 46-63. https://doi.org/10.17323/jle.2021.11244

Autonomy Orientations of English Language Teachers in Turkey

Aslihan Bagci[1],Belgin Aydin[2]

[1] Bilecik Şeyh Edebali University

[2] TED University

Correspondence concerning this article should be addressed to Bilecik Şeyh, Edebali University, School of Foreign Languages, 11100 Bilecik, Turkey. E-mail: aslihan.bagci@bilecik.edu.tr

The present study aimed to explore autonomy orientations of English language teachers' work at an intensive language school in Turkey. The autonomy orientations of the teachers were analysed through self-determination theory. The study also investigated if these orientations vary according to the teachers' gender, years of experience, and department of graduation. The investigation was carried out through an autonomy orientations questionnaire. Data were collected from 111 language teachers, 11 of whose opinions were utilized for further analysis. The results showed that teachers had moderate autonomy supportive orientation, which was reported to be risky. Moreover, gender was the only variable that had a significant effect on the autonomy orientations. The teachers mainly blamed the education system for restricting their adoption of the autonomous orientations. The results imply the necessity of explicit training on how teachers can be encouraged to have more autonomy supporting orientations.
Keywords: self-determination theory, intrinsic motivation, autonomy orientations of teachers, autonomy-supportive teaching, foreign language teaching

Introduction

Motivation is the moving power behind every decision that human beings take, and it is widely recognized for its importance in the education research field (Gardner, 1985). It has been put forward that motivation is a very critical factor in determining the effort and energy that learners put into learning activities (Csikszentmihalyi & Nakamura, 1979; Maehr, 1984; Pintrich et al., 1993).Deci and Ryan's (1985) self-determination theory (SDT) focuses on human motivation and its sources as well as the environment's effects on human motivation.Self-determined people are known to take responsibility and make choices about their lives, set their own goals, and do whatever is necessary to reach those goals without being forced by anyone else. They are also more likely to take part in difficult tasks when extrinsic rewards are not accessible, encounter lower levels of performance-related anxiety, and show more noteworthy levels of learning compared to those with a more extrinsic motivation (Deci & Ryan, 1985).

In an education context, there is a known positive correlation between learners' intrinsic motivation and the support they get to act autonomously as well as achieve positive learning outcomes (Black & Deci, 2000; Vandergift, 2005). Considering that students get most of the support from their teachers, teacher behaviors towards students deserve cautious attention. As claimed by Deci, Betley, Kahle, Abrams, and Porac (1981), teachers' instructing habits can range from highly autonomy supportive to highly controlling. Autonomy-supportive teachers tend to provide their students with enough time and resources and give them opportunities to express their ideas about their learning process. However, controlling teaching styles hinder students' chances to express themselves (Assor, Kaplan, Kanat-Maymon, & Roth, 2005).It is known that while studying in a high-pressure environment, students are not able to provide their own solutions and choose what to do or how to do (Reeve, 2002; Reeve, Bolt, & Cai, 1999; Reeve & Jang, 2006).

As in all areas of education, having self-determined learners is very important for increasing the effectiveness of foreign language learning. Turkey, as one of the foreign language contexts gives high priority to teaching English to its students in order to enable them to interconnect with the rest of the world. Turkish students start learning English starting at the second grade in the state schools. At the university level, a foundation year is provided to students at schools of foreign languages to help them cope with the requirements of their academic

fields. Yet, students' motivation levels are reported to drop immediately as soon as they enrol in these intensive language schools because they consider this one-year long education to be an obstacle to the start of their training in their academic fields (Aydın, 2017; Aygün, 2017).

Research focusing on the relationship between teachers' autonomy orientations and their relation to students' motivation is scarce. What is more, it is seen that most studies focus only on the Western societies and this results in a lack of data about Eastern societies, especially in terms of their autonomy orientations. Specifically, in a foreign language teaching and learning environment in Turkey, the teacher is often the main source of input so their behaviours towards their learners gain great importance. In addition, in collectivist societies, teachers are expected to act more controlling in parallel with their cultural norms, as stated by Reeve, et al. (2014). Turkey was found to be the third most collectivist country in the index of 39 countries Oishi, Diener, Suh, and Lucas (1999) created. In studies that are conducted not only with small children but also with older high school or university students, the ones taught by autonomy-supportive teachers were found to have higher academic achievement, higher perceived competence, higher self-esteem, and greater conceptual understanding (Reeve, 2002).Since learners might become more or less autonomous depending on their teachers' supporting or controlling orientations, studying Turkish language teachers might help us reveal the picture for the other collectivist societies as well. Thus, the teachers' autonomy orientations are worthy of investigation because learners might become more autonomous depending on their teachers' orientations. It is believed that by raising Turkish foreign language instructors' awareness of autonomy orientations and the possible effects of those orientations, this study will help create a better learning environment.

SDT focuses on human motivation and its sources as well as the environment's effects on human motivation. SDT claims that every human being is innately prone to continuously develop oneself towards a better self.As argued by Deci and Ryan (1985), self-determination is a universal need for all human beings. When people do not feel free, they find themselves desperate, even falling into amotivation.

In education contexts, motivation has been said to have a boosting effect on the effort learners put into their learning activities, regulating whether they are involved in the task eagerly or apathetically and reluctantly (Csikszentmihalyi & Nakamura, 1979; Maehr, 1984; Pintrich et al., 1993). The teachers' role is argued to be one of the most critical factors affecting students' learning (Deci & Ryan, 2002; Richard & Lockhart, 1994; Wigfield, Tonks, & Klauda, 2009). If learners have a good relationship with their teachers and if they are motivated, they will be more prone to apply tasks that can possibly help them learn better, namely, paying full attention to what is being instructed, practicing the learning material after the lessons, taking notes, asking for help if necessary, etc. (Yeşilyurt, 2008; Zimmerman, 2000). As referred by Deci & Ryan (2000), intrinsic motivation is "a natural wellspring of learning and achievement that can be systematically catalysed or undermined by teacher practices (p.55)."

Studies in literature, almost without exception, show that teachers' autonomy-supportive behaviours have positive effects on students as they can enjoy a high level of motivation, sense of involvement, productive learning, and psychological well-being, all resulting in higher academic performance in all their school years (Assor, et al., 2005; Chirkov & Ryan, 2001; Deci et al., 1981; Roth et al., 2007). In Jang, Reeve, and Deci's study (2010), the student engagement rates were positively predicted by the teachers' autonomy support rates, for which they had observers who rated both the teachers' and students' behaviours at the high school level. Similarly, in a study conducted with 7th graders' by Hofferber, Eckes, and Wilde (2014), focusing on rote learning and conceptual learning, the results revealed that the learners who were educated in an autonomy-supportive setting established a higher level of conceptual knowledge compared to the ones educated in a controlling setting. Yet, learners' rote learning did not change in either case. Griffin (2016) in his study with university students found that not only intrinsic motivation but also autonomy support was positively related to the ratings that the students gave about the instructions they received. In fact, intrinsic motivation had a balancing role, meaning that the higher the intrinsic motivation was the less predictive autonomy support was or vice versa.

In another study conducted with preparatory year students' by Hazaea and Alzubi (2018), it was shown that learners' autonomy increased to a greater degree as the teachers allowed the students to make their own choices of external reading materials. In addition, gaining more autonomy helped the students focus on the lesson objectives more deeply. Baranovskaya and Shaforostova (2018) underlined in their study that even though the significance of autonomy is widely recognized, it is not easy for the instructors to figure out effective strategies

that can cause a real change in the students' autonomy levels and it is teachers who are responsible for creating a learning environment where students can grow their own autonomy skills. Furthermore, Şakrak-Ekin and Balçıkanlı (2019) found a link between their students' language learning achievement and their autonomy levels, showing that learners who take responsibility of their own learning can be more successful in their language learning. It was concluded that the importance of learner autonomy should be stressed and different strategies that can help to promote it inside as well as outside the class must be examined.

When the literature related to foreign language teaching and autonomy orientation of teachers was examined, similar results were obtained. Dörnyei (2005) claimed that teachers whose orientations tend to be autonomy supporting and non-controlling increase foreign language learners' intrinsic motivation and their self-determination. Supporting this claim, Wu's study (2003) with young learners of English suggested a positive relationship between higher L2 intrinsic motivation and perceived autonomy of students. Similarly, in Pae and Shin's (2011) study with Korean university students, intrinsic motivation was found to have a relation to EFL achievement only in a classroom where students were given the chance to express themselves freely. Reeve and Jang (2006) showed that pre-service teachers in autonomy-supportive learning environment could reach higher levels of academic achievement compared to those in controlling environments.

Studies have proved that controlling factors are effective for reducing intrinsic motivation and even creating amotivational consequences (Deci, et al., 198; Deci & Ryan, 1985). Having rewards and punishments, setting a deadline, goals that have been created without asking the opinions of individuals, and undesired competition can be listed as examples of controlling factors. Pelletier et al. (2002) pointed out a close relationship between teachers' autonomous motivation and autonomous teaching environments. In their study, they found that as the teachers received pressure from their management and parents, they became less self-determined about their teaching, which resulted in being more controlling with their students.

There are some studies focusing on teachers' autonomy orientations or students' motivation related to various factors including teachers' gender and experience. The results of these studies, however, showed conflicting results. For example, Reeve et al. (2004) did not find significant differences in students' engagement based on the teachers' gender. Similarly, although Opdenakker and Van Damme (2007) found that teacher gender was not associated with differences in autonomous motivation, their study results revealed that teacher gender can predict successful classroom management, with male teachers maintaining their class management better than female teachers who were perceived to be stricter. On the other hand, analysing the relationship between student learning and female teachers in India, Chudgar and Sankar (2008) revealed a positive relationship between language learning and being taught by a female. Being in a female teacher's classroom was found to be advantageous for language learning in the 2nd through 6th grades. Klassen and Chiu (2010) investigated the effects of teachers' gender and their years of experience on self-efficacy. Years of experience appeared to have a nonlinear relationship with self-efficacy, since it increased as one reached the middle of his/her career and then fell after that point. In terms of gender, female teachers were found to have lower classroom management self-efficacy. In Aelterman et al.'s study (2014), neither teachers' years of teaching experience nor their gender had an association with autonomy support.

There have been a limited number of studies conducted on the autonomy orientations of teachers in the Turkish context. In their study, Güvenç and Güvenç (2014) examined how math and science teachers' autonomy support and classroom management styles varied in terms of their subject and years of experience. It was revealed that teachers' autonomy support was at a medium level and that their autonomy support levels did not vary according to their subject matters or years of experiences. Similarly, in a study conducted by Güvenç (2011) with class teachers, no variance in the teachers' autonomy supports according to their years of experience was found. On the other hand, one difference observed between novice teachers and more experienced teachers was that the novice ones generally had a more controlling approach rather than being autonomy supportive. A positive relationship was also found between academic achievement and intrinsic motivation as well as autonomous learning and academic achievement of Turkish undergraduate students (Karataş, et al., 2015).

In a report the British Council prepared in 2015 after investigating Turkish higher education in detail, it was pointed out that the students had stated that they knew they needed to learn English but they did not want to learn it. For this reason, they considered their one year of foreign language preparatory education to be like a

holiday. As Aydın (2017) underlined, students who feel helpless due to the curriculum repeated each year and still not making progress start to lose their motivation, which leads to a great problem when they must start university and continue their education related to their profession where they will use the English language. As soon as they step onto the university campus, their negative attitudes about foreign language learning and feelings of failure from their past affect them during their preparatory year (Aydın, 2017). The belief that it is difficult to learn a foreign language and that they cannot learn a foreign language after a certain age affects the motivation of the students even more negatively (Arslan & Akbarov, 2010).

Studies in the literature prove that autonomy-supportive teaching leads to better learning in many aspects. It is also argued that teachers might be encouraged to adapt a more autonomy-supportive orientation in their classes (Leroy et al., 2007).Hence, it is important to determine the autonomy orientations of the teachers as the first step. Learning English is very important in Turkey especially for those studying in higher education institutions because if they are not successful in their intensive English program, they cannot continue their academic lives. Turkish students are known to be very reluctant towards learning English even though they know that they must follow the literature in both their education and professional lives through English and that they cannot fully watch the media, which is an important part of their lives, without knowing English; this is the core reason why the approach of the English language instructors was specifically investigated in this study. Since this is still an underinvestigated field in the Turkish context, this study aims to reveal how teachers' autonomy orientations vary according to their gender, years of experience, and department of graduation. How Turkish EFL teachers perceived their autonomy orientations was another focus of the study. In Turkey, both graduates of English Language Teaching (ELT) departments and graduates from language-related departments can work as a language teacher at the university level. While ELT graduates have pedagogical courses and are equipped with methods and techniques for language teaching, non-ELT graduates do not always have such a pedagogical background. This was thought to be a variable in teachers' autonomy orientations in this study.

The following research questions were asked for these aims:

1. What are the autonomy orientations of the language teachers at Anadolu University School of Foreign Languages?
2. Do the orientations vary according to the teachers'
   a) gender? b) years of experience? c) department of graduation?
3. What are the teachers' perceptions regarding their autonomy orientations?

Materials and Methods

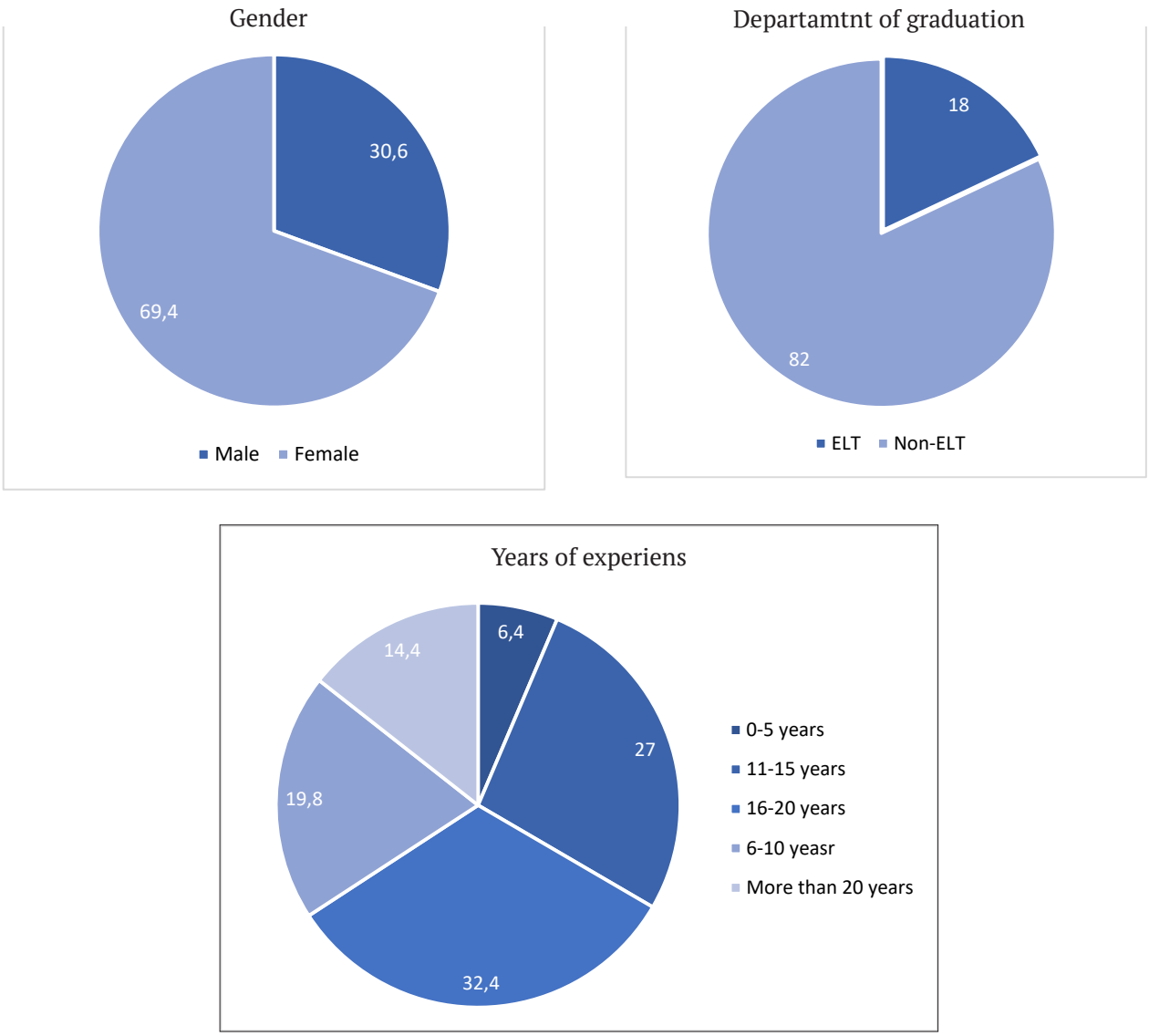
Participants and Setting

Being one of the biggest state universities in Turkey, Anadolu University School of Foreign Languages (AUSFL) was chosen as the setting of the study as it offers an intensive language program to university students before they start their education in their academic fields. The participants of the present study were 111 English language teachers at the School of Foreign Languages. Eleven of these participants also voluntarily contributed to the qualitative part of the study, which aimed to gather more detailed data on their autonomy orientations. Since teachers' gender, their years of experience, and majors were variables of the study, they were asked to reveal information regarding these three factors. Figure 1 presents information about the teachers' backgrounds.

As shown in Figure 1, 69.4 % of the participants (77 out of 111) were females while 30.6 % of them (34 out of 111) were males. In terms of the department of graduation, 82% of the participants were English Language Teaching (ELT) graduates and 18% of them were non-ELT graduates. When it comes to their years of experience, 6.4% of them had 0-5 years of experience, 19.8% had 6-10 years, 27% of them had 11-15 years, 32.4% of them had 16-20 years, and 14.4% of them had 20 years or more experience in the field. To summarize, the majority of the participants of the study consisted of very experienced females who were graduates of the ELT department.

Eleven participants also volunteered for the open-ended responses that aimed to reveal teachers' perceptions of autonomy orientations in general, and their own orientations in particular with possible explanations for

Figure 1
Teachers' gender, department of graduation, and years of experience



the reasons behind them. Volunteers from the three types of scores on the Autonomy Orientation Questionnaire were chosen; namely, among the teachers who got the highest score (3 autonomy supportive teachers), the lowest score (4 controlling teachers), and average scores (5 moderately autonomy supportive teachers) considering the mean score of all of the participants.

Assessments and Measures

The Questionnaire

The main instrument utilized in this study was a questionnaire that aimed to find the extent to which teachers are being controlling or autonomy supportive with their students. Being a type of "Motivators' Orientations Questionnaire", the "Problems in Schools (PIS)" Questionnaire by Deci, Schwartz, Sheinman, and Ryan (1981) was adopted for use in this study. Since the original scale was designed according to the primary and middle school levels, the PIS scale was adapted to the university-level setting with the help of an expert committee. For adaptation, the specific wording used for the school levels in the questionnaire were altered to the ones that would be used in a university setting and the names used for the example vignettes were changed into Turkish ones to help the participants relate to the vignettes more easily. The lexical items and the situations given either

in the vignettes or in their options were checked and modified to match a university setting. The PIS Questionnaire has eight situations, each including four options of response. The participants were asked to rate the degree of appropriateness of each of the four options (on a 7-point scale) for each situation. Those four options provided for each case represent four sub-scales: Highly Controlling (HC), Moderately Controlling (MC), Moderately Autonomous (MA), and Highly Autonomous (HA). Lastly, in order to be representative of cases teachers might experience in an intensive language teaching context, two more vignettes were added to the questionnaire, making a total of 40 questionnaire items to rate (Appendix A). The new version of the questionnaire was given to five ELT experts to ensure content validity. Overall score that one gets from this questionnaire, ranging from -180 for Highly Controlling to +180 to Highly Autonomous, gives information about the autonomy orientation of the participant. Zero is considered Neutral. Based on this scale, every positive score gained is considered autonomy supportive and every negative score is considered controlling. The scores up to +90 are considered moderately autonomy supportive and when the score is greater than +90, it is highly autonomy supportive. Likewise, the scores as low as -90 are considered moderately controlling, and highly controlling if it is lower than -90.

Evaluation Guestions

In order to gain a deeper understanding of teachers' opinions on the autonomy support of their colleagues they are working with in general and their own orientations, 11 of them with low, medium, and high orientations were invited to evaluate the findings of the quantitative part. The participants were asked to share their opinions on the results regarding the autonomy orientation scores as well as their overall opinions on autonomy support and the possible reasons behind these results. The teachers who participated in the qualitative part were asked to write detailed answers to six open-ended questions that were prepared based on the research questions with the guidance of expert opinion (Appendix B).

Data Collection Procedures

Before administering the questionnaire to the participants, a pilot study was conducted with a similar group of teachers who were working at another intensive langauge program. The possible problems in the scale items such as the appropriateness of the wording or the vignettes to the setting were investigated. After making the necessary modifications considering the opinions of the teachers in the pilot study, the questionnaire was administered to the participants towards the end of the fall semester of the 2017-2018 academic year. On the consent form that was presented online, the participants were informed about the anonymity and confidentiality and were asked to volunteer to participate by the researcher. Demographic information including gender, years of experience and their department of graduation was also added to the questionnaire. After analysing the questionnaire results according to their autonomy orientation scores, 11 participants were asked for their volunteer participation in the second part of the study.

Data Analysis

The percentages for the gender, years of experience, and department of graduation of the participants were calculated. Then, an overall autonomy score was calculated following the suggested formula by Deci, et al. (1981). Next, to test the normality of the distribution of the 111 participants' autonomy scores, a Kolmogorov-Smirnov (K-S) test was applied. Since this test is very sensitive to the sample size, the coefficient of skewness was also determined. Proving the normality of distribution meant that parametric statistical methods could be used in order to compare the autonomy orientation scores of the participants in accordance with their demographic information. A t-test was utilized for the relationship between gender and autonomy as well as the relationship between the department of graduation and autonomy scores of the participants. In addition, an ANOVA was used to examine the relationship between the participants' years of experience and their autonomy score. Lastly, the qualitative data were analysed identifying the codes and the themes as suggested in the content analysis scheme of Creswell (2012). This was done by two separate researchers in order to increase inter-rater reliability.

Results

Results of the Guantitative Data

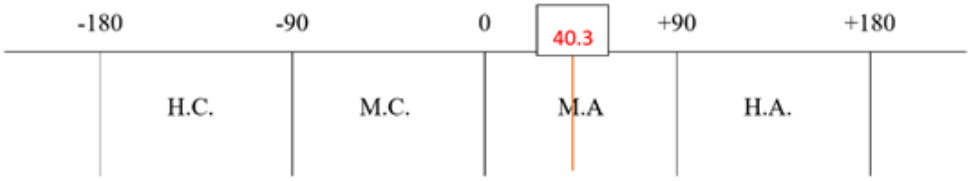
The distribution of the autonomy orientations of language teachers obtained from 111 participants is presented in Table 1 below.

Table 1
Descriptive statistics of the teachers' autonomy orientations

Statistics	Value
Mean	40.31
Standard Deviation	20.92
Kurtosis	-.964
Skewness	-.064
Variation	437.76

Firstly, it should be mentioned that the majority of the teachers had a positive score on the Autonomy Orientation Questionnaire, which meant that autonomy-supportive orientations dominated the controlling ones. The mean score of the 111 participants was 40.31, which falls in the "Moderately Autonomy Supportive" group. The highest score obtained was 82, which was very close to the limit of Highly Autonomy Supportive orientation. The lowest score obtained was -4, very close to 0, which was neutral. None of the participants could be considered completely controlling or completely autonomy supportive. An illustration representing the average placement of the participants in the autonomy orientation scale is presented in Figure 2 below.

Figure 2
The place of participants in the autonomy orientation scale



As the second step, a one-sample t-test was applied to test whether the mean score of all the participants was significantly different from the midpoint of the scale, "0". The results obtained are shown in Table 2.

Table 2
One-sample t-test results

N	Mean Score	S.D.	t	d.f.	Significance (p)
111	40.31	20.92	20.3	110	.000

As seen in Table 2, the one-sample t-test results were statistically significant at the .05 level ($p < .05$). In other words, the average of the scores obtained from the scale was found to be statistically significant to the reference score "0".

In order to find the teachers' autonomy orientations and their dependence on the teachers' gender, educational background, and experience, fa K-S test was applied and the result that was obtained for the distribution of scores was found to be significant ($p = .028 < .05$). This value indicates that distribution is not a normal one; thus, the coefficient of skewness was obtained and interpreted. The coefficient of skewness was found to be -.06. Since this value remains within the range of +/-1, it can be concluded that the distribution met normality assumptions. As the number of categories in the independent variable category was two, an independent samples t-test was applied to find out the effect of gender on teachers' autonomy orientations.

Table 3
Autonomy orientations and gender

Gender	N	Mean Score	S.D.	t	d.f.	Significance (p)
Male	34	34.18	20.17	-2.08	109	.040
Female	77	43.01	20.8			

As it can be seen in Table 3 above, the mean of the males' scores was 34.18 while it was 43.01 for the females. The difference between the mean of the scores was found statistically significant at the 0.05 level (p=.040<.05), concluding that the female EFL teachers had higher autonomy support scores compared to the males.

Having the differences between the means in the results of both t-test and variation analysis be significant is not usually sufficient because these analyses cannot provide a variation rate of the independent variable's power of explanation for the dependent variable. Therefore, the effect size was measured. In this study, the effect size was obtained through Cohen's d coefficient for the analyses conducted for t-tests in cases where the difference between means was significant. Cohen's coefficient to measure effect size was found to be 0.44. This value can be interpreted as the difference between the means of scores having a medium effect size in practice.

To identify if teachers' autonomy orientations varied depending on their teaching experience, they were categorized into five groups as having 0-5 years, 6-10 years, 11-15 years, 16-20 years, and more than 20 years of experience; then a one-way ANOVA was applied to test whether EFL teachers' autonomy orientations varied across these five categories. Firstly, the homogeneity of variations was reported with the Levene test. According to the obtained results, the Levene test was not found statistically significant at the .05 level (p=.277>.05). Since this condition means that the pre-requisition of variation homogeneity was met, the ANOVA table was interpreted for reporting descriptive statistics. Table 4 is presented below.

Table 4
Autonomy orientations and experience

Experience	N	Mean Score	S.D.
0-5 years	7	37.57	27.50
6-10 years	22	43.86	21.16
11-15 years	29	46.44	16.13
16-20 years	36	39.25	21.35
20+	17	28.58	21.16

The highest mean score of 46.4 belonged to teachers with 11 to 15 years of experience. This was followed by the teachers with 6 to 10 years of experience (M=43.9), 16 to 20 years (M=39.3) and 0 to 5 years (M= 37.6). The lowest mean score belonged to the teachers who had the most teaching experience, with a mean of 28.6. Although the most experienced teacher group had the lowest autonomy orientation score, these results were not found statistically significant. The related table is presented below.

Table 5
The variation analysis results of the autonomy orientations

	Sum of squares	d.f.	F	Significance (p)
Within Groups	3799.24	4	2.27	.066
Between Groups	44354.34	106		
Total	48153.58	110		

As seen in Table 5, since P=.066>.05, was found insignificant, it is possible to conclude that autonomy orientation scores did not vary according to the participants' years of experience.

In order to find out the effect of educational background on teachers' autonomy orientations, the participants were divided into the ones graduating from English Language Teaching (ELT) departments and the ones graduating from the other departments (non-ELT). An independent samples t-test was applied to find the answer for this sub-question. The obtained results are presented in Table 6.

Table 6
Autonomy orientations and departments of graduation

Graduation	N	Mean Score	S.D.	t	d.f.	Significance (p)
ELT	91	39.17	20.24	-1.22	109	.226
Non-ELT	20	45.45	20.65			

As a result, the mean scores of teachers who graduated from English Language Teaching departments was identified as 39.17 while the mean of scores non EFL graduates was 45.45. Although the non-ELT graduates had a higher autonomy orientation score compared to ELT graduates, this difference was not found to be statistically significant (p=.226>.05).

Teachers' Interpretations of their Orientations

The teachers who participated in the qualitative part were asked to interpret the overall results, the possible reasons behind these results, as well as their own individual scores compared with the others in the group.

The teachers interpreted a moderate autonomy-supportive orientation as having "outdated educational beliefs in Turkey" and following an "inflexible syllabus". They thought females were more autonomy supportive because of "the natural inherited differences between males and females". Participants believed that all teachers "adapt to the context of the institution in which they work", that is why, years of experience or the department of graduation didn't make a significant difference.As another explanation, the "personality of the teacher", which does not change with experience or graduation, was suggested by the participants. Some teachers also stated that "being more familiar with the latest technological and educational trends" could the motive for less experienced teachers having a more autonomy-supportive orientation.

When it comes to interpreting the participants' own personal autonomy orientation scores, the possible reasons listed were "personal self-development efforts", "inflexible syllabi", and the "problems with the Turkish educational belief system" for all teachers from all autonomy orientation scores. They also stated that since "supporting autonomy leads to better learning", EFL teachers should be more autonomy supportive. Finally, when the teachers were asked for their suggestions for other EFL teachers to become more autonomy supportive, most of them felt the need for "institutionally organized workshops and seminars" as well as "a more flexible syllabus".It should be highlighted that similar ideas were shared by the teachers from all levels of autonomy orientation, including very high scoring and very low scoring teachers. For example, one teacher with a moderate controlling score mentioned trying to find ways to foster learner autonomy by working on her personal development by reading recent articles, while another teacher whose score was quite above the average mentioned negative ideas such as the syllabus stopping him from being creative and autonomy supportive.

Discussion

The results of the study indicated that the EFL teachers working at AUSFL had a moderate autonomy-supportive orientation with a mean score of +40,3, which is actually close to being controlling.Thus, it is possible to state that even though language teachers are currently in the autonomy-supportive area, there is a risk of exhibiting a more controlling approach towards their teaching. In fact, there were a few participants who actually had scores lower than zero (-4 and -3). There might be several reasons for this result. The first one is teachers' belief that since the schools of foreign languages have to ensure systematicity and coordination, they have to follow a common syllabus with common materials and assessment for a crowded group of learners and teachers.This might create a feeling of not being able to act autonomously. Yet, this belief also discloses that teachers do not have clear ideas about what autonomy-supportive teaching is and how they can facilitate it for their learners, as also argued by Reeve and Jang (2006). The following quotation from a high-scoring teacher supports this idea of teachers' beliefs regarding standardized applications:

> – *While trying to fit into a well-detailed schedule and assign a number of tasks, we may be gaining more control over the students... (T1) (Score: +80)*

Teachers also blamed the lack of autonomy support on the Turkish education system for their own autonomy orientation. Moreover, they believed that autonomy was something for the students, rather than the teachers, as seen in the except below:

> *– In my opinion, "autonomy" is a term to which Turkish students are strangers. They are led or even interfered with by their families or teachers till they come to university so much that they sometimes get lost or do not know what to do when they are not led. (T6) (Score: +39)*

It is worth adding Maehr's (1976) claim here that supporting or not supporting the psychological needs of the learners, which will in return lead to intrinsic motivation, can be a cultural matter. Schools exist within a society and they are influenced by public policy as well as the cultural milieu. If the culture is pressuring people to be successful, it may become rather difficult for the teachers and the administrators to be able to maintain an autonomy-supportive orientation in the classroom. Supporting this view in his study, Peacock (2001) found significant differences in the teaching styles among teachers by ethnic origin, which points to culture being the source of those perceptions. In addition, Reeve, et al. (2014) argue that a society's collectivist or individualistic orientations predict its teachers' autonomy orientations. In their study, teachers who lived and worked in collectivist cultures self-reported adopting a more controlling style; stating that they believed it to be the right classroom practice according to their cultural norms. Parallel to this finding, Kozan, Ergin, and Varoğlu (2007) argued that Turkish organizations are known for their centralized decision making, strong leadership, and collectivist orientation, all of which can cause Turkish schools to adopt similar orientations. That might be the rationale behind teachers' beliefs that they cannot act autonomously even if they are working with university-level students who are expected to act independently and be more responsible for their own learning.

The findings revealed significant differences between male and female teachers' autonomy orientations, females having higher autonomy support scores. Although there aren't many studies focusing on the gender factor, Chudgar and Sankar's (2008) and Özkal and Demirkol's (2014) studies found female teachers to be more autonomy supportive compared to their male colleagues. This result can be been explained by females' having been found to be more emphatic and more forgiving compared to males (Broidy, et al., 2003; Hoffman, 1977). Females' communication skills and their ability to create more equally distributed social relations, while males tend to be more dominant, might also be presented as an explanation as to why females can be more autonomy supportive (Merchant, 2012). Furthermore, Strober and Tyack (1980) argued that by being mothers, women need to be patient, nurturing, and understanding, all of which cause them to be more supportive and "better teachers than men" (p.496). In their responses, Turkish teachers in this study also claimed that inborn and later gained characteristic differences between males and females, as well as their power relationships, may be the reason behind this significant result. The extract below explains this difference:

> *– The possible reason for this is the sexist stereotypes the society imposes. Or, the reasons could simply be rooted in some cliché gender-oriented generalizations. For example- with all the shame of saying this- and totally not a supportive of the cliché- men are bossier. (T1) (Score: +80)*

It appeared that while younger teachers appeared to be less controlling, the most experienced teacher group where teachers had more than 20 years of experience, had the most controlling orientation. This may be a result of the changing student profiles and the technological gap that is expanding between the older generation of teachers and the students. Since the new generation is thought to be digital natives and the older generation of teachers are identified as being more digital immigrants (McMahon, 2016), the senior teachers may not feel sufficiently equipped when addressing the new generations, which might in turn lead them towards holding a more controlling approach. Supporting this result, Nakata (2011) argues that having more years of experience, especially in an exam-oriented educational context might be seriously undermining the practices of autonomy. However, considering the fact that the autonomy orientation scores did not have any significance statistically according to the participants' years of experience, the differences observed in the sample cannot be used as a basis for making a deduction towards the population. According to the participants, adapting to the environment in which one works was the main reason for not finding significant differences according to experience variable. The quotation below from a moderately controlling teacher exemplifies this:

> – *When you start working in an institution that has specific goals, the effects of experience and graduation start to lose their importance. In order to keep the balance, the needs of the institutional goals, you feel the need to change and be more autonomous. (T2) (Score: -3)*

No significant differences were observed between EFL teachers' department of graduation and autonomy orientations. While ELT graduates might be expected to be more autonomy supportive due to having a more comprehensive pedagogic education, the lack of explicit training on learner autonomy, like other affective variables during pre-service teacher education, might be the reason of this insignificance. Contrary to the expectations, the non-ELT graduates were found to have higher autonomy orientations scores. As illustrated in the following quotation, personal professional efforts such as participating in a CELTA training program mattered more than their previous educational background:

> – *The main reason could be CELTA, during which I got more insight into allocating more time to students for activities they conduct on their own. Another reason is my belief that -taking their ages into account- as they are the learners, they should decide how to learn if not what to learn. This gives them the opportunity to choose the better and also more enjoyable way to learn.(T1) (Score:80)*

The teachers' evaluations revealed mostly similar opinions for the reasons behind their autonomy orientations. Thus, it can be argued that the teachers' opinions were not shaped according to their autonomy orientations. For example, language teachers with a wide variety of scores seemed to agree with each other in terms of the gender-related results, namely females being more autonomy supportive compared to males. Moreover, teachers with high, low, and average autonomy orientations mentioned the same reasons about the fact that years of experience and department of graduation did not make a difference on teachers' autonomy orientations. Personality factors and adapting to the work environment were their explanations. There was not a group of high scorers or low scorers claiming any different reasons. To sum up, teachers can have every different scores yet share similar ideas, all depending on their interpretations of the system they are living in; a teacher with a moderate controlling score can mention working on her personal development by reading recent articles and trying to find ways to foster learner autonomy, whereas another teacher with a score quite above the average can blame the syllabus for stopping him from being creative and autonomy supportive. Thus, starting by identifying teachers' beliefs and the rationale behind these beliefs should be the first step in encouraging teachers to act more autonomously and only then can they act as role models for their students in their journeys to become more autonomous learners.

Conclusion

The results of this study found that the majority of language teachers had autonomy-supportive orientations. There was no teacher who was completely controlling or completely autonomy supportive. Yet, the average scores were not far from being in the controlling zone. Gender was the only factor making a difference in teachers' orientations; female teachers were identified as more supportive. No significant differences were observed in terms of teachers' experience or their department of graduation and autonomy orientations. The teachers' explanations mainly blamed the lack of autonomy support on the Turkish education system for their own autonomy orientation scores. Their opinions also disclosed that they did not have clear ideas about what autonomy-supportive teaching was or how they could facilitate it for their learners. The teachers' responses also revealed that their autonomy orientations and their perceptions did not match. It seems, after all, that how teachers perceive the world around them is very important for the decisions and deductions they make, yet this does not seem to be related to how autonomy supportive they are. This can be the key in terms of creating a more autonomy-supportive environment. The teachers' negative perceptions causing them to think that their power is limited also affects the way they perceive their roles and the options available to them.

The results of this study have some significant implications for language teachers as well as administrators. In-service and pre-service education programs can also benefit from the findings. Firstly, it should be made explicit to all teachers that supporting learners' autonomy leads to better learning (Reeve & Jang, 2006). Thus, teachers should be encouraged to raise their own awareness on their individual choices for becoming autonomous. They need to be informed that creating and maintaining an autonomy-supportive climate in their classrooms is under their control and being autonomous is not a responsibility of the learners only. Explicit

training on what autonomy-supportive teaching is, how it can be encouraged, and what learners' and teachers' responsibilities are were suggested by almost all of the participants of the study. Showing real examples of autonomy-supportive actions by teachers as well as how the language they use and the decisions they make affect their students might be suggested as the most effective methods of training for the teachers. Sometimes just encouraging teachers to see events from different perspectives might be enough. As suggested by the teachers' verbal feedback. this might have a tremendous effect on students' development and teachers' awareness on this effect should be increased.

Being aware of the autonomy orientations of the teachers in their institutions might help administrators make decisions that will lead to a better learning environment in their institutions. For example, if there is more than one teacher instructing a class, a balanced combination of teachers can be assigned to a group of students; that is, rather than two controlling teachers teaching the same class, one autonomy supportive teacher together with one controlling teacher can be a better arrangement for creating a fair teaching environment and giving students opportunities to practice becoming autonomous language learners.

The insights gained from this study may also contribute to the design of and both pre- and in-service training programs by providing opportunities for the teachers to make choices and internalize the rationales behind events that are beyond the students' control. This way, rather than demanding a more flexible syllabus, teachers can realise their own power and start making their teaching more autonomy supportive by giving students opportunities to become autonomous language learners. Finally, if pre-service teachers can have a chance to experience autonomy support themselves, they can be more prone to supporting their students' autonomy when they start teaching in their own classes. This way, higher levels of autonomy support can be transferred to the next generations.

Limitations and Future Directions

The main limitation of this study is that the findings here cannot be generalized to all language schools, nor to all language teachers, because a more representative sample of Turkish teachers is needed. This study was conducted in one representing university and can simply be regarded as a case study; whether the language teachers in this study are autonomous or not may not be applied to all Turkish language teachers or language teachers in general. A more representative sample of Turkish teachers should be involved in order to be able to make generalizations. Examining the teachers' perceptions of the possible factors affecting their autonomy as well as their suggestions to address the issue might be considered good first steps for encouraging them to act more autonomously. What is more, teachers' perceptions, their levels of autonomy, the possible factors affecting their autonomy, and the corresponding solutions can be examined in future studies. Since autonomy support is a learnable concept, it may be very beneficial to study the autonomy orientations of the teachers working with young learners and setting up training sessions. Lastly, after organizing awareness-raising training on autonomy support, the post-effects of those training sessions can be investigated. All in all, one should not forget, no matter how well a training program is designed, its success depends on the teachers who apply what they learned from it (Aydın, 2017). Therefore, a training program concerning the motivation orientations of the EFL teachers working in schools of foreign languages seems to be necessary.

Conflict of interests

There are no conflicts to declare in the present study.

References

Aelterman, N., Vansteenkiste, M., Van den Berghe, L., De Meyer, J., & Haerens, L. (2014). Fostering a need-supportive teaching style: Intervention effects on physical education teachers' beliefs and teaching behaviors. *Journal of Sport and Exercise Psychology, 36*(6), 595-609. https://doi.org/10.1123/jsep.2013-0229

Arslan, M., & Akbarov, A. (2010). Türkiye'de yabanci dil öğretiminde motivasyon-yöntem sorunu ve çözüm önerileri [The matter of motivation-method and solution offers in foreign language teaching in Turkey]. *Selçuk Üniversitesi Edebiyat Fakültesi Dergisi*, *24*, 179-191.

Assor, A., Kaplan, H., Kanat-Maymon, Y., & Roth, G. (2005). Directly controlling teacher behaviors as predictors of poor motivation and engagement in girls and boys: The role of anger and anxiety. *Learning and Instruction, 15*(5), 397-413. https://doi.org/10.1016/j.learninstruc.2005.07.008

Aydın, B. (2017). *Türkiye'de hazırlık okullarında yaşanan sorunlara bir çözüm önerisi: Anadolu Üniversitesi örneği* [Solutions for the problems of a preparatory school in Turkey's Anadolu University instance]. Pegem Yayınları.

Aygün, Ö. (2017). *A scale of Turkish preparatory school university students' demotivational factors towards learning Englis* [Unpublished master's dissertation]. Anadolu University.

Baranovskaya, T., & Shaforostova, V. (2018). Learner Autonomy through Role Plays in English Language Teaching. *Journal of Language and Education*, *4*(4), 8-19. https://jle.hse.ru/article/view/7942

Black, A. E., & Deci, E. L. (2000). The effects of instructors' autonomy support and students'autonomous motivation on learning organic chemistry: A self-determination theory perspective. *Science Education, 84,* 740–756. https://doi.org/10.1002/1098-237X(200011)84:6<740::AID-SCE4>3.0.CO;2-3

British Council. (2015). *The state of English in higher education in Turkey: A baseline study*. Yorum Basın Yayın Sanayi Ltd.

Broidy,L.,Cauffman,E.,Espelage,D.L.,Mazerolle,P.,&Piquero,A.(2003).Sexdifferencesinempathyanditsrelation to juvenile offending. *Violence and Victims, 18*(5), 503–516. https://doi.org/10.1891/088667003780928143

Chirkov, V. I., & Ryan, R. M. (2001). Parent and teacher autonomy-support in Russian and US adolescents: Common effects on well-being and academic motivation. *Journal of Cross-cultural Psychology, 32*(5), 618-635. https://doi.org/10.1177/0022022101032005006

Chudgar, A., & Sankar, V. (2008). The relationship between teacher gender and student achievement: Evidence from five Indian states. *Compare, 38*(5), 627-642. https://doi.org/10.1080/03057920802351465

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

Csikszentmihalyi, M., & Nakamura, J. (1989). The dynamics of intrinsic motivation: A study of adolescents. In R. Ames, & C. Ames (Eds.), Research on motivation in education: Goals and cognitions (pp. 45–71). Academic Press.

Deci, E. L., & Ryan, R. M. (2000). The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4), 227-268. https://doi.org/10.1207/S15327965PLI1104_01

Deci, E. L., Betley, G., Kahle, J., Abrams, L., & Porac, J. (1981). When trying to win: Competition and intrinsic motivation. *Personality and social psychology bulletin, 7*(1),79-83. https://psycnet.apa.org/doi/10.1177/014616728171012

Deci, E. L., Schwartz, A. J., Sheinman, L., & Ryan, R. M. (1981). An instrument to assess adults' orientations toward control versus autonomy with children: Reflections on intrinsic motivation and perceived competence. *Journal of Educational Psychology, 73*(5), 642–650. https://psycnet.apa.org/doi/10.1037/0022-0663.73.5.642

Deci, E., & Ryan, R. M. (1985). *Intrinsic motivation and Self-Determination in human behavior*. Springer Science and Business Media.

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

Gardner, Robert C. (1985) *Social psychology and second language learning*. Edward Arnold.

Griffin, B. W. (2016). Perceived autonomy support, intrinsic motivation, and student ratings of instruction. *Studies in Educational Evaluation, 51,* 116-125. https://doi.org/10.1016/j.stueduc.2016.10.007

Güvenç, E. & Güvenç, H. (2014). İlköğretim matematik ile fen ve teknoloji öğretmenlerinin sınıf yönetim biçemleri ve özerklik desteği algıları [Classroom management styles and autonomy support perceptions of primary school mathematics and science and technology teachers]. *NWSA-Education Sciences, 9*(3), 311-322

Güvenç, H. (2011). Sınıf öğretmenlerinin özerklik destekleri ve mesleki özyeterlik algıları [Autonomy supports and professional self-efficacy perceptions of classroom teachers]. *Kuram ve Uygulamada Egitim Yönetimi Dergisi, 17*(1), 99-116.

Hargreaves, E. (2014) The practice of promoting primary pupils' autonomy: Examples of teacher feedback, *Educational Research*, *56*(3), 295-309. https://doi.org/10.1080/00131881.2014.934554

Hazaea, A. N., & Alzubi, A. A. (2018). Impact of mobile assisted languagelLearning on learner autonomy in EFL reading context. *Journal of Language and Education*, *4*(2), 48-58. https://doi.org/10.17323/2411-7390-2018-4-2-48-58

Hofferber, N., Eckes, A., & Wilde, M. (2014). Effects of autonomy supportive vs. Controlling teachers' behavior on students' achievements. *European Journal of Educational Research, 3(4),* 177-184. https://doi.org/10.12973/eu-jer.3.4.177

Hoffman, M. L. (1977). Sex differences in empathy and related behaviors. *Psychological Bulletin, 84*(4), 712-722 https://doi.org/10.1037/0033-2909.84.4.712

Hu, P., & Zhang, J. (2017). A pathway to learner autonomy: a self-determination theory perspective. *Asia Pacific Education Review*, *18*(1), 147-157. https://doi.org/10.1007/s12564-016-9468-z

Jang, H., Reeve, J., & Deci, E. L. (2010). Engaging students in learning activities: It's not autonomy support or structure, but autonomy support and structure. *Journal of Educational Psychology,*102*(3), 588-600*. https://doi.org/10.1037/a0019682

Karatas, H., Alci, B., Yurtseven, N., & Yuksel, H. G. (2015). Prediction of ELT students' academic (language) achievement: Language learning orientation and autonomous learning. *International Online Journal of Educational Sciences, 7*(1), *160-171*. http://dx.doi.org/10.15345/iojes.2015.01.014

Kozan, M.K., Ergin, C. & Varoglu, D. (2007). Third party intervention strategies of managers in subordinates' conflicts in Turkey. *International Journal of Conflict Management, 18*(2), 128-147. https://doi.org/10.1108/10444060710759345

Little, D. (2007) Language Learner Autonomy: Some Fundamental Considerations Revisited. *International Journal of Innovation in Language Learning and Teaching, 1*(1), 14-29. https://doi.org/10.2167/illt040.0

Maehr, M. L. (1976). Continuing motivation: An analysis of a seldom considered educational outcome. *Review of Educational Research, 46*(3), 443-462. https://doi.org/10.3102/00346543046003443

Maehr, M. L. (1984). Meaning and motivation: Toward a theory of personal investment. *Research on Motivation in Education, 1,* 115-144.

McMahon, M. J. (2016) The adoption of a virtual learning environment among "digital Immigrant" engineering lecturers: A case study. *Irish Journal of Academic Practice, 5*(1), Article 3. https://doi.org/10.21427/D78B04

Merchant, K. (2012). *How men and women differ: Gender differences in communication styles, influence tactics, and leadership styles*. CMC Senior Theses.

Nakata, Y. (2011). Teachers' readiness for promoting learner autonomy: A study of Japanese EFL high school teachers. *Teaching and Teacher Education, 27*(5), 900-910. https://doi.org/10.1016/j.tate.2011.03.001

Oishi, D., Diener, E., Suh, E. M., & Lucas, R. E. (1999). Value as a moderator in subjective well-being. *Journal of Personality and Social Psychology, 69,* 915-924. https://doi.org/10.1111/1467-6494.00051

Opdenakker, M. C., & Damme, J. V. (2007). Do school context, student composition, and school leadership affect school practice and outcomes in secondary education? *British Educational Research Journal, 33(2),* 179-206. https://doi.org/10.1080/01411920701208233

Özkal, N., & Demirkol, A. Y. (2014). Opinions of teachers on the necessity and execution of learners' Autonomy support. *Education Sciences, 9*(3), 293-310. https://doi.org/10.12739/NWSA.2014.9.3.1C0619

Peacock, M. (2001). Match or mismatch? Learning styles and teaching styles in EFL. *International Journal of Applied Linguistics, 11*(1), 1-20. https://doi.org/10.1111/1473-4192.00001

Reeve, J., Bolt, E., & Cai, Y. (1999). Autonomy-supportive teachers: How they teach and motivate students. *Journal of Educational Psychology, 91*, 537–548. https://doi.org/10.1037/0022-0663.91.3.537

Reeve, J. (2002). Self-determination theory applied to educational setting. In E. L. Deci, & R. M. Ryan (Eds.), *Handbook of self-determination research* (pp. 183–203). University of Rochester Press.

Reeve, J., Jang, H., Carrell, D., Jeon, S., & Barch, J. (2004). Enhancing students' engagement by increasing teachers' autonomy support. *Motivation and Emotion, 28(2),* 147-169. https://psycnet.apa.org/doi/10.1023/B:MOEM.0000032312.95499.6f

Reeve, J., & Jang, H. (2006). What teachers say and do to support students' autonomy during a learning activity. *Journal of Educational Psychology, 98*(1), 209-214. https://doi.org/10.1037/0022-0663.98.1.209

Richards, J. C., & Lockhart, C. (1994). *Reflective teaching in second language classrooms.* Cambridge University Press.

Strober, M. H., & Tyack, D. (1980). Why do women teach and men manage? A report on research on schools. *Signs: Journal of women in culture and society, 5*(3), 494-503. https://doi.org/10.1086/493734

Şakrak-Ekin, G., & Balçıkanlı, C. (2019). Does Autonomy Really Matter in Language Learning? *Journal of Language and Education*, 5*(4),* 98-111. https://doi.org/10.17323/jle.2019.8762

Wang, M. T., & Eccles, J. S. (2016). Retracted: Multilevel Predictors of Math Classroom Climate: A Comparison Study of Student and Teacher Perceptions. *Journalof Research on Adolescence, 26*(3), 617-634. https://doi.org/10.1111/jora.12153

Wu, X. (2003). Intrinsic motivation and young language learners: The impact of the classroom environment. *System, 31,* 501 – 517. https://doi.org/10.1016/j.system.2003.04.001

Yeşilyurt, S. (2008). A Self-Determination approach to teaching writing [Unpublished doctoral dissertation]. Atatürk University.

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology, 25*(1), 82-91. https://doi.org/10.1006/ceps.1999.1016

Appendix A

AUTONOMY ORIENTATIONS QUESTIONNAIRE

Dear Colleague,

In this questionnaire, there are ten situations that language instructors who work in a preparatory language school can come across, and each of these ten situations has four different options of responding to it. Please rate the degree of appropriateness of each of the four options (on a 7-point scale) for each of the ten situations. If you consider the option to be extremely appropriate respond to the option with the number "7" or if you consider the response highly inappropriate respond with the number "1". Thus, there are 40 ratings. There are no right or wrong ratings on these items. People's styles differ, and we are simply interested in what you consider appropriate given your own style. Please respond to each of the 40 items using the following scale:

very inappropriate 2- 3- 4 moderately appropriate 5-6-7 very appropriate

A. Cem is an average student who is at B level. During the past two weeks he has not been participating in the class activities. The work he does in the class is accurate but he has not been completing tasks. The most appropriate thing for Cem's teacher to do is to:

1. Impress upon him the importance of finishing his tasks since he needs to learn this material for his own good. (MC)
2. Let him know that he doesn't have to finish all of his work now and see if it is possible to help him work out the cause of the listlessness. (HA)
3. Make him stay for more practice after the class. (HC)
4. Let him see how he compares with the other students in terms of his assignments and encourage him to catch up with the others. (MA)

B. At a teachers' meeting it was mentioned that Selen has made more progress than expected since the last meeting. All teachers hope she continues to improve so that she does not have to repeat the preparatory year. They have been expecting this since her last midterm exam results. The teachers decide that the best thing to do is to:

5. Promise her a cinema ticket to her favorite film if she continues to improve. (HC)
6. Tell her that she's now doing as well as many of the other students in her class. (MA)
7. Tell her about the report, letting her know that they're aware of her increased independence in school. (HA)
8. Continue to emphasize that she has to work hard to get better grades. (MC)

C. Deniz loses his temper a lot and has a way of agitating other students. He doesn't respond well to what the teacher tells him to do and the teacher is concerned that he will disturb his class. The best thing for the teacher to do with him is to:

9. Emphasize how important it is for him to "control himself" in order to succeed in his academic and social life. (MC)
10. Ask the director to change his classroom. (HC)
11. Help him see how other students behave in these various situations and praise him for doing the same. (MA)
12. Realize that Deniz is probably not getting the attention he needs and start being more responsive to him. (HA)

D. One of Ms. Cansever's students is a very good player on the university football team, which has been winning most of its games. However, she is concerned because he has been missing most of his quizzes due to the games and has not been doing his homework. The best thing to do is to:

13. Ask him to talk about how he plans to handle the situation. (HA)
14. Tell him he ought to decide to give up football so he can catch up in his classes. (MC)
15. See if others on the team are in the same dilemma and suggest he does as much preparation as the others. (MA)
16. Make him miss tomorrow's game to study; football has been interfering too much with his school work. (HC)

E. The Beginner-1 level group has been having trouble catching up with the other groups all year. The best thing for Miss Vatan to do is to:

17. Organize competitions in English so that they will be motivated to do better. (MA)
18. Make them practice more and give them special rewards for improvements. (HC)
19. Have each student keep a notebook and emphasize how important it is. (MC)
20. Help the group devise ways of learning together (games, and so on). (HA)

F. In Mr. Coşkun's class there is a female student named Mehtap. She is quiet and usually alone. Despite the efforts of all her teachers, Mehtap has not been accepted by the other students. The best thing for Mr. Coşkun to do is to:

21. Push her into interactions and provide her with a lot of praise for any social initiative. (HC)
22. Talk to her and emphasize that she should make friends, so she'll be happier. (MC)
23. Invite her to talk about her relationships with the other students and encourage her to take small steps when she is ready. (HA)
24. Encourage her to observe how other students relate and to join in with them. (MA)

G. For the past few weeks, things have been disappearing from the teacher's desk and a student's money has been stolen. Today, Mehmet was seen by the teacher taking a pen from her desk. The best thing for the teacher to do is to:

25. Talk to him about the consequences of stealing and what it would mean in relation to the other students. (MA)
26. Talk to him about it, expressing your confidence in him and attempting to understand why he did it. (HA)
27. Report this situation to the school's manager; stealing is something that cannot be tolerated, and he has to learn that. (HC)
28. Emphasize that it was wrong and have him apologize and promise not to do it again. (MC)

H. Mr. Arthur's student Canan has been getting average grades, and he'd like to see her improve. The best thing to do is to:

29. Encourage her to talk about her grades and what they mean to her. (HA)
30. Go over her grades with her; point out where she stands in the class. (MA)
31. Stress that she should do better; she'll not be able to finish preparatory school with grades like these. (MC)
32. Offer her big rewards for every A and smaller ones for every B for her future grades. (HC)

I. There is a student in Miss Tarçın's class who insists on using his mobile phone during the lesson even though she has warned him many times not to do so. The best thing to do is to:

33. Talk to him about why they have such a rule in class and the consequences he will face if he keeps using his mobile phone during the lessons, including getting a lower grade. (MC)
34. Warn him one more time telling him this is his last chance before he gets a time out. (HC)
35. Talk to him and assure him that his attention is very necessary in class. (MA)
36. Convince him it is distracting her and hindering her teaching performance in class. Not for him but for the teacher he needs to stop, so that she can teach to her full potential. (HA)

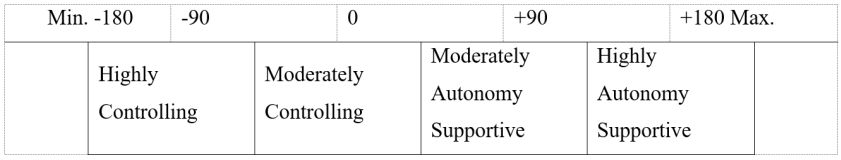
J. One of Mrs. Ander's students has a serious attendance problem and she knows that it is going to result in his failure that year. The best thing for Mrs. Ander to do is to:

37. Make him talk to a previous student who has failed due to the same reason (absenteeism) and make him see how regretful this previous student is. (HA)
38. Get his friends involved and try to convince him to come to class regularly. (MA)
39. Have a private talk with him in order to find out why he really doesn't come to class and try to help with whatever problem he is having. (MC)
40. He needs to take responsibility for his own behavior and face any kind of results that he causes. He is not a child anymore. (HC)

Appendix B

EVALUATION QUESTIONS

1. According to the "**Motivators' Orientations Questionnaire**" that was applied for the purposes of this study, the mean of the Autonomy Scores of the EFL teachers working at AUSFL was **40.3** which correlates to a **Moderately Autonomy Supportive** orientation. You can find a representation of the scores one could possibly get from this questionnaire below. How would you evaluate this situation? What do you think the possible reasons behind this result are?



2. When we take a look at the results in terms of **gender** differences, it was found that **female** teachers were slightly **more autonomy supportive** compared to their male colleagues. How would you evaluate this situation? What do you think the possible reasons behind this result are?
3. In the study, it was also investigated whether the Autonomy scores varied according to the teachers' **years of experience** or **their department of graduation** (ELT / non-ELT).It was found that neither of these variables made a significant difference. How would you evaluate this situation? What do you think the possible reasons behind this result are?
4. Your personal Autonomy Orientation score out of this questionnaire was "X". According to the figure presented on the previous page, how would you evaluate this situation? What do you think the possible reasons behind this result are?
5. Do you think EFL teachers should be more Autonomy Supportive? Why?
6. In order for the EFL teachers to become more Autonomy Supportive, what do you think should be done personally and institutionally? Why?

National Research University Higher School of Economics
Journal of Language & Education Volume 7, Issue 1, 2021

Deveci, T., & Ayish, N. (2021). Qualitative Adjectives in EFL Students' Reflective Writing Essays. *Journal of Language and Education, 7*(1), 64-77. https://doi.org/10.17323/jle.2021.10979

Qualitative Adjectives in EFL Students' Reflective Writing Essays

Tanju Deveci, Nader Ayish
Khalifa University

Correspondence concerning this article should be addressed to Tanju Deveci, English Department, Khalifa University, Al Saada St - Zone 1 - Abu Dhabi, UAE. E-mail: tanjudeveci@yahoo.com

Qualitative adjectives are often used in expressive writing, including reflective writing. They express and (de)intensify feelings and emotions, thereby expressing stance. This study investigates the adjective profiles of 60 first-year EFL students' reflective essays and compares male and female university students' utilization of qualitative adjectives and those used as attitudinal stance markers. Data were collected from a reflective writing task after students participated in a seminar on effective listening. Analyses were conducted considering the General Service List (GSL), the Academic Word List (AWL), and words that do not appear in either of the preceding lists. The results indicated that qualitative adjectives accounted for 6% of the words in these reflective writing essays, and the male students used a greater number of adjectives than the female students. This difference, which was at a statistically significant level, likely stems from male students' greater use of adjectives from the Academic Word Lists. The results also showed that 47.5% of the adjectives used in these essays were attitudinal. There was no statistically significant difference between the frequencies with which the male and the female students used these attitudinal adjectives. The results are discussed and recommendations are made to increase students' effective use of adjectives in reflective writing.

Keywords: adjectives, attitudinal markers, freshman students, gender, qualitative adjectives, reflective writing

Introduction

Adjectives are used to describe the qualities or attributes of nouns. In this way, they point to the distinctions made between nouns. Such is the place of adjectives in language that a sentence without adjectives provides information, but it does not say much (Murray, 2014). Adjectives are also used to express attitude and stance. As such, they are utilized quite frequently in expressive writing, which encourages individuals to use verbal expressions on their thoughts and feelings about lived experiences (Beaumont, 2018). Expressive writing comes in many forms, one of which is reflective writing. In addition to certain core skills such as description, critical analysis, self-awareness, and evaluation, reflective writing requires clear and concise articulation of thought processes, which necessitates the use of good written English (Handler, Handler, & Gill, 2011).

Although various aspects of effective language use in reflective writing have been noted (e.g., organization, syntax, sentence length, appropriate lexis, and personal language), adjectives—a key component of expressive writing—have received inadequate attention. This is despite a few studies that have highlighted the essential role adjectives play in expressive writing (Adams, 2013; Lee, Fitria, & Ginting, 2019). While reflective thinking and writing are taught and practiced in different disciplines of higher education, especially in professional degree programs such as teacher education (Hagevik, Aydeniz, & Glennon, 2012; Hatton & Smith, 1995; Moon, 2006; Zulfikar & Mujiburrahman, 2018) medical and health professional education (Bernard, et al., 2012; Crowe & O'Malley, 2006; Nguyen, Fernandez, Karsenti, & Charlin, 2014; Schön, 1987) and engineering (Kilgore, Sattler, & Turns, 2013; Reidsema, 2009; Turns, Newstetter, Allen, & Mistree, 1997; Wallin & Adawi, 2018), little research has examined how students use language to write reflectively (Hatton & Smith, 1995; Russell, 2005). One reason for this lack of research is due to the nature of reflective writing. Students, for example, often perceive reflective writing as an unfamiliar genre and ancillary to their core studies. Questions as to how to teach and assess reflective writing simply add to the challenge of conducting such research (Dyment & O'Connell, 2011; Kember et al., 2008; Plack, et al., 2005; Sumsion & Fleet, 1996).

Unless trained well, using language effectively can be a formidable challenge for students, especially those pursuing education in a foreign language. The lack of research in this area is what led us to want to better understand how our students write reflectively. This is particularly true relative to the use of adjectives, because so little is known about the linguistic patterns of how students use adjectives to express themselves in reflective writing. So too can examining how females and males use adjectives in reflective writing. In this paper, therefore, we describe our Emirati freshman students' reflective writing behavior at the end of a seminar within the context of an undergraduate English and Communication course.

Theoretical Background

Adjectives

An adjective is

> a word used with a noun to denote some quality, attribute, or fact, which we connect in thought with that for which the noun stands, without making as distinct *assertion* that the quality or attribute belongs to what we are speaking about (Mason, 1858, p. 13).

Mason (1858, pp. 32–36) identified three types of adjectives:

a) qualitative adjectives denoting some quality or attribute;
b) quantitative adjectives denoting how much or how many of a particular thing we have in our thoughts. These include the indefinite article, the cardinal numeral adjectives, words such as *all, few, more, most, both,* etc., and;
c) demonstrative/determinative adjectives pointing out which thing(s) we are speaking of out of the class of things denoted by a common noun. These include the definite article, the adjective pronouns, pronominal adjectives, and the ordinal numbers.

Another classification is made in Collins Cobuild English Grammar (as cited in Stillar, 1988), according to which there are five major categories: a) qualitative, b) emphasizing, c) classifying, d) color, and e) postdeterminers. We chose the first two of these five types of adjectives based on the aims of our study. In addition, they are the ones one would expect to see more often in reflective essays (Çapar, 2014; Ferris, 2009). Qualitative and emphasizing adjectives, for example, express feelings, attitudes, and stance and serve as building blocks of expressive and reflective writing. Postdeterminers, on the other hand, are outside the scope of our study because they function as grammatical structures rather than lexical items. Classifying adjectives are also outside the study's scope because they serve as attributive adjectives and classify their associated nouns. As for color adjectives, while they function to simply describe the color of something, we paid special attention to any use of color adjectives used to denote feelings and attitudes.

The majority of adjectives are comprised of qualitative adjectives. This is one reason why instructional materials targeting learners of English tend to focus on qualitative adjectives in their explanation of what an adjective is. This appears to be the case for lower-level learners in particular. The other two types of adjectives, 'quantifiers' or demonstratives, are often treated as distinct grammatical structures with little, if any, mention of their attributes as adjectives. Many learners of English, as well as native speakers, are likely unfamiliar with the classification of such grammatical structures as adjectives. This study, therefore, only focuses on qualitative adjectives in order to account for this unfamiliarity. Among the various meanings qualitative adjectives have is the one of position/attitude, as in the examples 'a *key* point' and 'the *beautiful* message' (Stillar, 1988). Such examples highlight the attributes qualitative adjectives assign to things, people, concepts, etc. from some point of view. From this perspective, they play a key role in expressive language, which is under the scope of the current research.

Reflective Thinking and Reflective Writing

Reflective thinking allows individuals an opportunity to internalize and reconstruct newfound knowledge and make sense of lived experiences (Boud, Keogh, & Walker, 1985; Burns & Bulman, 2000; Mezirow, 1998; Ryan, 2011). While definitions of reflective thinking vary, a common theme among many is that thinking reflectively is essential for one's personal growth and development (Lew & Schmidt, 2011; Rodgers, 2002). According to Boyd and Fales (1983), for example, reflection is the "Process of internally examining and exploring an issue of concern, triggered by an experience, which creates and clarifies meaning in terms of self and which results in a

changed conceptual perspective" (p. 99). This definition best captures our understanding of reflective thinking as it applies to our research context.

Reflective writing stems from reflective thinking and it is often used as a pedagogical strategy to clarify thinking and to give more time and space for deeper and more meaningful reflection about something of significance (Tsingos-Lucas et. al, 2017; Nunn & Brandt, 2016). In this sense, reflective writing reinforces learning by encouraging students to thoughtfully think through key concepts and ideas related to a course. Rusche and Jason (2011) reiterate this by saying reflection in the form of writing engages students in an examination of their own writing which results in their learning to "value the process of learning, inquiry, and critical self-reflection while acquiring and constructing self-knowledge" (p. 1). Reflective writing has also been shown to help students contextualize the new information they are learning, refrain from meaningless memorization of factual information, and "see learning as a never-ending journey that does not stop outside of the classroom or after they have graduated" (Cisero, 2006, p. 231).

Moon (2004) notes that reflective writing is the external expression of the mental process of reflection though such expression is unlikely to be the same as what takes place in one's head. The cycle in which reflection takes place is comprised of three stages, which have come to be known as the reflective cycle (Deveci & Nunn, 2018). The first stage includes the description of an experience in relation to what happened in context as well as the writer's feelings aroused by the experience. The second phase includes the writer's evaluation and analysis of the experience. The final stage includes the writer's reflection on what he/she has learned from the experience and how he/she plans to improve future experience. Considering the fact that reflective writing allows the writer to explain his/her rationale and thought processes to others (Brown, 2016), it engages the writer in interpersonal communication. This is supported by Gibbs' observation (as cited in Rodgers, 2002, p. 845) that "reflection needs to happen in community, in interaction with others." This points to the role writing as a linguistic skill plays in effective communication, which is described in detail in the following section.

Language Use in Reflective Writing

Reflective writing often shares a number of features with traditional essay writing. Examples include a well-organized structure, the correct use of tenses, and lexical variety. Reflective writing may also require students to make references to relevant content from lectures so that their instructor can gain insight into their thinking and learning processes as well as how they interpret the information they have been presented. Reflective writing also adds the elements of personal experience, first person, and present and future tense (Eckstein & Ferris, 2018; MacLellan, 2004). The language used in reflective writing, therefore, tends to be mostly personal and subjective. One reason for this is because reflective writing tasks are often personal in nature (e.g., "Reflect on an experience you recently had with your team"). While students are expected to write in the first person and incorporate concrete and abstract language into their reflections, many find this type of genre writing challenging and uncomfortable compared to more familiar and traditional academic writing tasks (Burton & King, 2004; King, 2001; Lyubormirsky, Sousa, & Dickerhoof, 2006). Part of the reason for this stems from the experience some students have had with academic essays that tells them that the use of the personal pronoun "I" should be avoided. Another reason is because students may worry about how such tasks are going to be assessed (Wong et al., 1995).

Tone is another complicating factor students must contend with when writing reflectively. Although tone is dependent on the topic and purpose of a particular style of writing, reflective writing, unlike traditional academic essays (which tend to be serious and formal) may be ironic and informal or something in between. This flexibility in tone adds to the challenge of writing reflectively (Ferris, 2009; Lu & Ai, 2015). So too does the use of 'show, don't tell' as an effective writing strategy where the student is expected to use sensory details to illustrate something (e.g., "He was sweating profusely as he stood on stage before the large audience") rather than simply state something directly (e.g., "He was very nervous on stage"). These reflective writing characteristics, in particular, encourage the use of adjectives to bring writing to life.

Writing in general and reflective writing in particular appear to be more challenging for those who write in English as a foreign language (EFL), which is the context in which the current study was conducted. These writers have to put more cognitive effort into their writing (van Weijen, 2014), find it more difficult to write due to factors like limited lexis, syntax, and organization (Al Murshidi, 2014; Englander, 2014; Handler et al., 2011), and suffer from a lack of stylistic competence (Buckingham, 2008). In fact, Deveci (2015) and Deveci (2017)

found that Emirati freshman students' vocabulary usage in reflective writing essays was heavily based on the list of the first 2000 most common words only (i.e., General Service List (West, 1953)), which may not be adequate for a writer to fully express themselves. Along with these difficulties, students' inability to think critically (Thabet, 2008), think reflectively (Ayish, 2020), and write reflectively prior to coming to university creates additional problems, all of which require tailor-made support (Moussa-Inaty, 2015).

The Use of Adjectives in Reflective Writing

According to OxbridgeEssays (n.d., parag. 22)[1], reflective writers should utilize adjectives to describe their experiences. Therefore, it is best to tap into a "wide range of adjectives" and use specific rather than "vague adjectives such as '*okay'* or '*nice'* to express oneself." For example, "My *little* brother wants to be a bigshot lawyer when he grows up." Writing in this way helps the reader better visualize what is being described. Similarly, Jefferys (2018) points to the importance of using concrete and abstract forms of language in reflective essays. He notes that the former, in particular, helps the writer bring life to his/her writing and draw the reader into an essay through the use of sensory or descriptive details. Using adjectives with concrete language enhances the impact of the description (e.g., "her *high* voice" and "a *metallic* smell"). "This reflection of the details in the essay is part of what brings the author's voice to the reader. The point of this type of essay is to bring that reader as close to the experience as possible," argues Jefferys (2018, parag. 4). Birney (2012, p. 203) also found that strong writers of reflective essays used adjectives judiciously when they "were 'appraising'" or reasoning in writing "about a particular topic." Other research, too, revealed that the use of feeling and thinking verbs with adjectives (e.g., "a *positive* impact") and reasoning adverbs with adjectives (e.g. "extremely *challenging*") was among the linguistic features strong reflective writers were able to utilize (Gibson et al., 2017).

It is also important to note expressions of stance (i.e., the expression of personal feelings, attitudes, value judgements, or assessments (Biber et al., 1999)) in reflective essays. These expressions are at the heart of a writer's ability to express feelings, emotions, and lived experiences (Beaumont, 2018). Among the various grammatical devices writers can use to express stance are adjectives, which can appear in adjectival predicates (e.g., "To me, it's just *amazing* that …") or the writer's lexical choice (e.g. "They are very *nice*, cats are."). Such use of adjectives, then, can be labelled as an *attitudinal stance marker* in that an adjective "can reasonably be interpreted as either a positive or negative attribute to a proposition" since they can be, for example, "'*good'* or '*bad'*, '*important'* or '*unimportant'*, '*certain'* or '*uncertain'*" (Forsell, 2011, p. 27). According to the Scottish Examination Board (1992 as cited in Klenowski, 2002), writers ought to express a clear and appropriate stance in reflective essays. Similarly, Hardy and Friginal (2016) note that expressions of stance are used to express stance in other types of university-level writing including creative writing and response papers.

Despite the important role adjectives play in reflective essays, to our knowledge, there is no research conducted to identify the actual adjective coverage in a well-written effective reflective essay. However, we analyzed the sample essay we provide our students with when teaching reflective writing skills. We found that the adjective coverage of the essay was 8%[2]. We also analyzed the adjective coverage of reflective essays written by five high-achieving students in our previous classes. The average coverage of adjectives in their essays was 7.5%. We therefore concluded that a successful reflective essay should have adjective coverage of approximately 8%.

Language Use and Gender Differences

The personal and contextual nature of reflective writing makes it difficult to compare and quantify differences between the genders (Magno, 2009). Nonetheless, some research has shed light on the use of language by males and females in general and on adjectives in particular. The findings appear to be inconsistent. For example, research by Alkhrisheh, Aziez, and Alkhrisheh (2019) showed that male and female university students used language in a similar fashion. Likewise, Rubin and Greene (1992) found that male and female university students' writing is far more similar than different. On the other hand, other research revealed that female students, in comparison to their male counterparts, had a greater tendency to use adjectives indicating personal feelings (Ishikawa, 2015) and attitudes (Barczewksa & Andreasen, 2018; Wenjing, 2012).

[1] Oxbridge Essays (2020, January 20). *A complete guide to reflective essay.* https://www.oxbridgeessays.com/blog/complete-guide-to-writing-a-reflective-essay/

[2] The sample essay had been written by the course coordinator a few years before the current study took place. Several course instructors had also given feedback on it.

According to Lakoff (1975, p. 51), some adjectives are gender neutral and may be used by males or females (e.g., *great*, *terrific*, *cool*, and *neat*), but other adjectives are mostly used by females (e.g., *adorable*, *charming*, *sweet*, *lovely*, and *divine*). For example, both males and females may say, "What a *terrific* idea," but only females are likely to say, "What a *divine* idea" (Lakoff, 1975, p. 52). Ishikawa (2015, p. 597), on the other hand found that male students tended to use "adjectives and verbs related to social economic activities (*social*, *democratic*, *lose, eat, appear, avoid*)" more often than female students. While Lakoff (1975) highlighted linguistic differences between males and females, it is not very clear if it is gender that drives language use (Baker, 2014; Eckert,1989; Newman, Groom, Handelman, & Pennebaker, 2008; Tannen, 1990), factors such as the age of the speaker, listener, or social status of the interlocutors (Coates, 2016; Ishikawa, 2011), or a combination of elements. Taken together, previous literature on the use of adjectives points to the need to better understand how males and females perceive and use language (and especially adjectives) to communicate and write reflectively.

Research Questions and Hypotheses

In this research, we describe the performance of the students without instructional intervention. Towards this end, we sought answers to the following questions:

1 a) What is the adjective profile in the students' reflective essays?
b) Is there a difference between the adjective profiles in the female and male students' reflective essays?

2 a) What adjectives are used as attitudinal stance markers in the students' reflective essays?
b) Is there a difference between the female and male students' use of these adjectives?

Based on these questions and the relevant literature reported above, we formulated the following hypotheses:

1. Approximately 8% of the students' reflective essays will be comprised of qualitative adjectives.
2. A greater number of the qualitative adjectives in the student essays will be based on the General Service List (GSL).
3. An important number of the qualitative adjectives will be attitudinal markers.
4. The female and male students will use qualitative adjectives with similar frequencies.
5. The female students will use attitudinal adjectives more often than their male counterparts.

Materials and Methods

Participants

Sixty Emirati first-year students (30 male and 30 female) in a freshman communication course at an Abu Dhabi-based university participated in the study. Their ages ranged between 18 and 21, with a mean age of 19. Most students (87%) attended government schools where Arabic is the medium of instruction and where English is taught as a foreign language starting in grade 1. While students are required to have IELTS scores of 6 overall to enter our university, they do not necessarily have a 6 in writing.

Data Collection Tool

Data were collected using a reflective writing task assigned at the end of a seminar on effective listening, which was the first of a set of seminars on a variety of communication-related topics including interpersonal communication, intrapersonal communication, and intercultural communication. The task required the students to respond to the following prompt:

Which of the poor listening practices do you think is the most harmful and why? Describe a situation in which you experienced this practice. Analyze and evaluate this experience in relation to the concepts presented in the readings, and state how you would modify any of your future listening behaviors to become a better listener.

While the content of the seminar text has a particular focus on poor listening practices, it also includes the characteristics of effective listeners and effective listening practices. During the seminar itself, ample classroom discussions were held on students' effective and ineffective listening experiences. Therefore, by the time

students were asked to respond to the prompt, their awareness had already been raised significantly. This assignment was chosen because reflective writing is an important aspect of the current course as well as other courses students take as part of their program of study. Identifying students' behaviors of reflective writing (without instructional intervention) could also serve as a diagnostic tool and help inform course instruction. Teaching ideas generated from this assignment are included in the conclusion of the paper.

While the students were not taught anything about adjectives, because the study aimed to describe their writing behavior at the beginning of the communication course, the rubric shared with the students included a general component of language use in addition to other aspects of writing such as content, organization, and task completion. The language component, which weighed less than the other components, included three main aspects:

a) style and expression: developed and engaging, with a consistent level of formality appropriate to the task.
b) grammar and vocabulary: sophisticated and appropriate use of a broad repertoire
c) clarity and accuracy: evidence of thorough, careful proofreading, correct spelling and punctuation.

Prior to completing the task, students were given instruction as to why reflective writing within the context of their academic studies is important. In addition, they received details about the reflective cycle as defined by Deveci and Nunn (2018). They also received instruction on different ways of organizing reflective essays, how to develop content that is supported by the relevant literature, and the need to include their personal experiences. They were asked to address their responses to their instructors, and they were reminded about the previously mentioned rubric that would be used to evaluate their writing.

Analysis

Of the three types of adjectives identified by Mason (1858), this study focused on qualitative adjectives. Therefore, only the qualitative adjectives in the student reflective essays were analyzed. Adjective profiles were examined based on the General Service List (GSL) (West, 1953), which is composed of the 2000 most common words only, the Academic Word List (AWL) (Coxhead, 2000), which contains 570 word families that appear frequently in academic texts, and the Off-list, which are words that do not appear in any of the preceding lists. In doing this, free online software was used (http://www4.caes.hku.hk/vocabulary/profile.htm).

Through a norming session, we (the two researchers) separately identified the adjectives in the students' reflective writing papers. The interrater agreement reached was 90%. We then compared the results, discussing the relatively low number of differences, until we reached an agreement. We followed a similar pattern in identifying the adjectives used as an attitudinal stance marker in the students' reflective writing. A third independent researcher was involved at this stage. We excluded the items on which we could not agree.

Descriptive statistics such as frequencies and percentages were used to describe the data sets. In comparing the data sets according to gender, log-likelihood (LL) similarity was used. Following Rayson, Berridge, and Francis (2004), a LL value of 3.84 or higher was considered significant at the level of $p < 0.05$.

Results

The first research question aimed to describe the adjective profile of the student reflective writing essays and asked whether the female and the male students' adjective profiles differed from each other. Results are presented in Table 1.

Table 1

Adjective profile of reflective writing essays

All			Male			Female				
Word count	**Adj count**	**% of adj**	**Word count**	**Adj count**	**% of adj**	**Word count**	**Adj count**	**% of adj**	**LL**	**p**
20,588	1,234	6	8,310	593	7.1	12,278	641	5.2	+29.84	0.000

According to Table 1, 6% of the student essays were made up of adjectives. Table 1 also shows that the male students had a greater tendency to use adjectives than the female students (7.1% vs. 5.2%). The computed value of log-likelihood (LL) was 29.84>3.84, which indicates a significant difference between the two data sets at a statistically significant level (p=0.000).

The adjectives in the texts were also analyzed according to word lists. Results are summarized in Table 2.

Table 2
Adjective profiles according to word lists

	All		Male		Female			
	#	%	#	%	#	%	LL	p
GSL	798	64.7	360	60.7	438	68.3	-2.77	0.078
AWL	269	21.8	157	26.5	112	17.5	+11.47	0.002
Off-list	167	13.5	76	12.8	91	14.2	-0.43	0.000
Total	*1,234*	*100*	*593*	*100*	*641*	*100*		

As can be seen in Table 2, students had a greater tendency towards the use of adjectives from the GSL (64.7%). This was followed by the AWL (21.8%) and the Off-list (13.5%). It is also shown that the two groups of students generally had a similar tendency towards using adjectives despite the slight differences between percentages. The only significant difference between the data sets was in regards to the AWL words; the male students tended to use a greater number of adjectives in this category than did the female students (26.5% vs. 17.5%). The LL value was +11.47>3.84 with a p value of 0.002, which indicates a statistically significant difference.

The second research question aimed to identify the adjectives used as an attitudinal stance marker in the students' reflective essays and whether or not the male and female students' usage of them differed. Data related to this question are summarized in Table 3 below.

Table 3
Attitudinal adjectives in relation to total number of adjectives

All		M			F				
#	% of total adjectives	#	%	# of total adjectives	#	%	# of total adjectives	LL	p
586	47.5	258	43.5	593	328	51.2	641	-3.82	0.052

According to Table 3, the total number of attitudinal adjectives used by the students was 586, which accounts for 47.5% of the total number of adjectives (1,234). The table also shows that the female students tended to use a greater number of attitudinal adjectives than the male students (51.2% vs, 43.5%) in relation to the total number of adjectives in each set of reflective writing essays (593 and 641 respectively). However, the computed LL value of -3.82 was slightly below the benchmark of 3.84, indicating a lack of statistical difference.

We also looked at the frequency with which each attitudinal adjective was used and compared these according to gender. The results can be seen in Table 4.

As shown in the table, the top ten most frequently used adjectives included *important*, *effective*, *bad*, *harmful*, *good*, *better*, *efficient*, *poor*, *negative*, and *hard*. And only two categories of adjectives compiled according to frequencies revealed differences at statistically significant levels, and these were to the advantage of the female students. One of the categories was *good,* which occurred 33 times (10%) in the female data set, but only 8 times in the male data set. The other category included those with single occurrence. Of the 72 adjectives in this category, a significant number (52) were present in the female data set. Examples included *thoughtful, bitter, careless, irritated, wonderful, perfect, problematic, key, minor,* and *mindful*. For this category, the log-likelihood value was 5.49>3.84. On the other hand, the adjectives *positive* and *big* were each used twice by the female students, but ten times by the male students (LL=+7.91>3.84). Similarly, the adjective *efficient* was used only by the male students.

Table 4
Attitudinal adjectives according to gender

#[3]	Adjectives	Male		Female		LL	p
		#	%	#	%		
69	important	29	11.2	40	12.2	-0.11	0.738
65	effective	32	12.4	33	10	+0.71	0.4
58	bad	27	10.5	31	9.5	+0.15	0.699
46	harmful	20	7.8	26	7.9	-0.01	0.94
41	good	8	3.1	33	10	-10.95	0.001
20	better	8	3.1	12	3.7	-0.13	0.716
16	efficient	16	6.2	-	-	-	-
14	poor	6	2.3	8	2.4	-0.01	0.93
13	negative	7	2.7	6	1.8	+0.5	0.478
11	hard	8	3.1	3	0.9	+3.72	0.054
9	wrong, easy, difficult	9	3.5	18	5.5	-1.29	0.000
8	successful	2	0.8	6	1.8	-1.25	0.001
7	huge	4	1.6	3	0.9	+0.48	0.000
6	positive, big	10	3.9	2	0.6	+7.91	0.008
5	right, essential	7	2.8	3	0.9	+2.75	0.003
4	helpful, serious, mad, main, complex	6	2.3	14	4.8	-1.66	0.001
3	mindless, easier, boring, relevant, embarrassing, strong, significant, necessary, uncomfortable, angry, fundamental	17	6.6	16	4.9	+0.74	0.000
2	sad, critical, painful, annoyed, comfortable, worse, fake, great, worst, disorganized, excellent, sufficient, basic, dangerous, harder, small, bored, simple, best, possible, ignorant, frustrated	20	7.8	24	7.3	+0.04	0.000
1	thoughtful, bitter, careless, false, useful, rare, irritated, responsible, distracting, earnest, hostile, irrelevant, careful, wonderful, respectful, perfect, weak, monotone, disoriented, competent, disappointed, excited, ethical, valuable, problematic, new, preconceived, suitable, addictive, active, fine, stimulating, mean, high, acceptable, disrespectful, distorted, suffocating, key, innocent, nonjudgmental, meaningful, unable, amazing, worried, informative, attentive, special, tiny, tough, dominant, stressed, annoying, average, central, smart, shocked, old, judgmental, normal, happy, outstanding, attracting, minor, appropriate, interesting, total, energetic, mindful, beneficial, misunderstood, enormous	22	8.5	50	15.2	-5.49	0.002
	Total	**258**	**44**	**328**	**56**		

Discussion

One of the aims of the current research was to describe the adjective profile of the students' reflective writing essays. Results showed that 6% of the student essays were comprised of adjectives. This result negates our first hypothesis stating approximately 8% of the students' reflective essays would be comprised of qualitative adjectives. The ratio of adjectives detected in the study is considered limited based on the arguments made in the literature that the personal nature of reflective writing often requires the writer to make references to feelings (Ullmann, 2019), which necessitates their utilization of adjectives in addition to other language elements. Effective reflective writers have also been noted to use "a wide range of adjectives" (OxbrigeEssays, n.d.) including concrete and abstract forms to help bring life to writing and engage the reader through the use of sensory and descriptive details (Jefferys, 2018). These observations, coupled with findings from past research indicating that adjectives in reflective essays frequently occur with nouns and adverbs, point to the relatively common use of adjectives in the genre investigated in this study. Still, the lack of previous research focusing on

[3] Delineates each word separately

corpus analysis of reflective essays prevents us from fully supporting our expectation with greater empirical data. Nevertheless, we consider the relevant literature and the preliminary data reported in the current study to be a modest contribution to the field.

Challenges students face in writing reflective essays, including language use, have been noted in other contexts, too. Students taught to describe events without deeply engaging their own thought processes, for example, encounter formidable obstacles when asked to reflect on these events from a personal stance (Mann, Gordon, & MacLeod, 2007). The same is true for students who have taken countless standardized tests requiring them to regurgitate factual information, which has an impact on their skills in "asking questions, problem-solving, and learning to think new ways" (Mauk & Metz, 2013, p. 448). Similarly, Thabet (2008) found that public schools in the UAE fell short of fostering critical thinking as a main objective for education and that teachers often opted for memorization as the main teaching strategy. It is likely that such an approach impacts students' reflection skills in general and reflective writing skills in particular.

From the perspective of language use, Handler et al. (2011) observed that many students—especially non-native speakers of English—face significant difficulties in using written English correctly, and exhibit poor grammar, spelling, punctuation, and sentence and paragraph cohesion. As non-native English speakers with little to no previous exposure to reflective writing, the students in our study likely experienced similar challenges. This is supported by recent research revealing that many Emirati students that attended high schools where Arabic is the medium of instruction encountered difficulty transitioning to a university where English is the medium instruction (EMI) (Ayish, 2020). Other research also showed that Emirati students at an EMI university had difficulty writing reflectively and required language support as well as guidance for reflection (Moussa-Inaty, 2015).

Evidence for students' somewhat limited skills in using adjectives in reflective writing essays may also be found in their greater dependency on the GSL. Results from the current study showed that the majority of adjectives (64.7%) in student essays belonged to the GSL, which confirmed our second hypothesis (i.e., "A greater number of the adjectives in the student essays will be based on GSL"). A more sophisticated and advanced level of language use could have encouraged students to tap into a wider variety of lexis from the AWL. A closer look at how student grades are determined also showed that language use is weighted less than the other components of the rubric used to assess the essays (i.e., content, organization, and task completion). In addition, it is worth noting that the language component focused on students' skills in using a broad repertoire of engaging, sophisticated, and appropriate vocabulary together with other aspects of grammatical and mechanical aspects of writing. Emirati first-year university students' reliance on the GSL in reflective writing essays was revealed in our previous research conducted at the same institution (Deveci, 2015). A similar study we conducted also showed that first-year students' self-rating of their ability to use the AWL words was low (Deveci, 2017). These two studies, in fact, helped us formulate the second hypothesis. Other research found that vocabulary was another challenging aspect of writing for Emirati international students (Al Murshidi, 2014). Although the focus of these past studies was not on adjectives per se, the findings point to the attention required for students' development of lexical dexterity within an academic environment.

The current research also sought to identify the adjectives used as an attitudinal stance marker. Of the 1,234 adjectives in the student reflective essays, a significant number (586), accounting for 47.5%, were attitudinal adjectives. Given the nature of the task, this was an expected result, which confirmed our third hypothesis that an important number of the adjectives in the student essays would be attitudinal markers. Attitudinal stance markers are indeed an important characteristic of reflective writing, and adjectives play a key role in expressing feelings, emotions, thoughts, and lived experiences (Beaumont, 2018). Other similar types of university-level writing such as response papers and creative writing have also been shown to include more language used to express stance and mental processes (Hardy & Friginal, 2016). In other contexts, stance is also considered a significant factor in determining the success of reflective writing. For instance, the Scottish Examination Board (1992 as cited in Klenowski, 2002) stipulate that successful reflective writing clearly shows the writer's distinctive voice through the use of a clear and appropriate stance. The writer is also required to make his/her writing vivid and original, exploiting the possibilities of language resourcefully and subtly. The significant proportion of the attitudinal adjectives utilized by the students in the current study can be considered a natural attempt to adhere to these requirements.

We also compared the students' adjective use according to gender. The data revealed that adjectives, in general, were used more by the male students than the female students, which was confirmed by statistical results (LL=+29.84>3.84), which negates our fourth hypothesis (i.e., "The female and male students will use qualitative adjectives with similar frequencies"). This also contradicts the findings of previous research by Alkhrisheh, Aziez, and Alkhrisheh (2019) who found *no* significant differences between male and female university students' use of language. Our finding also differs from earlier research findings indicating that the writing of male and female university students is far more similar than different (Rubin & Greene, 1992).

As per the students' use of attitudinal adjectives, however, the female students used them with higher frequencies than the male students (51.2% vs. 43.5%), confirming our fifth hypothesis despite the lack of statistically significant levels of differences in general, except for the adjective good, which was used extensively by the female students compared to the male students (LL=10.95>3.84). Ishikawa (2015) also found that female students had a greater tendency to use adjectives denoting personal feelings (e.g., *disgusting*) in their academic essays. Similarly, Wenjing (2012) found that strong attitudinal adjectives such as *precious, adorable*, and *charming* were used more often by women than by men. In their study of adjectives, Barczewska and Andreasen (2018), too, found that women tended to use more of the eight adjectives (i.e., g*ood, bad, big, small, ugly, pretty, different, important*), which could be categorized as attitudinal markers. Theirs was an academic spoken discourse, but the finding on the higher frequency of attitudinal adjectives by women concurs with our finding. Taken together, these findings support Newman et al.'s (2008) observation that women are more likely to use words that express personal attitude than men. These findings also contradict earlier observations that women's and men's use of language is more similar than different (Alkhrisheh et al., 2019; Rubin & Greene, 1992), at least from the perspective of adjective use.

We also observed that the female students' use of attitudinal adjectives was more varied, with a significant number of adjectives being used only once. This may have been a factor in encouraging the female students' use of a greater number of attitudinal adjectives. This result may also support Lakoff's (1975) observation that males and females often use different adjectives. Still though, the results of the current study do not seem to provide strong evidence for differences between male and female language use. We suspect that the nature of reflective writing made it difficult for us to compare and quantify differences between genders, as was previously suggested by Magno (2009).

Pedagogical Implications

The findings from this research point to important pedagogical implications. First, they offer insight into how students in our English courses, without direct instruction, incorporate the use of adjectives while writing reflectively. The results showed that qualitative adjectives accounted for only 6% of the students' reflective writing essays. As we noted earlier, an absence of adjectives provides information, but not much else (Murray, 2014). Indeed, the infrequent utilization of these lexical items in reflective writing can limit learners' effective expression of reasoning, emotions, and attitudes. Thus, for foreign language learners studying in an EMI institution, increasing student awareness of the benefits of using adjectives to enhance their writing will likely go a long way in improving their ability to more effectively express themselves. In particular, students can benefit from explicit instruction into how to incorporate adjectives into their writing. Towards this end, Paraskevas's (2006) "Grammar Apprenticeship" approach can be adapted. In this approach, students can be guided to analyze the variety of ways adjectives are utilized in reflective writing papers composed by advanced users of English. This will help students better understand the choices authors make and develop "a feel for the beauty of [adjectives], for [their] power and strength and grace" (p. 65). They may be asked to reflect on how they currently use adjectives and how they may adopt and adapt the strategies other authors use. They are then asked to revise their reflective writing accordingly. With this increased understanding of the role of adjectives in expressive writing, they can make better informed decisions about how to express their engagement with critical thoughts, feelings, and attitudes. Overall, Paraskevas's approach supports the theory that the acquisition of effective writing is facilitated by learners' engagement in extensive reading of texts composed by competent authors (Krashen, 1984).

Paraphrasing in the form of editing can also be a useful activity. To this end, learners can be given a reflective writing piece with a limited number of qualitative adjectives. They can be asked to rephrase sentences using appropriate items from a list based on more sophisticated adjectives (e.g., the AWL). Considering the results of

this study, female students, in particular, may benefit from having their attention drawn to adjectives in the AWL. Learners may also be given the freedom to use adjectives of their own choice, but they should be challenged to justify their choices. This could play an important role in engaging them towards greater reflection on language use in their reflective writing.

To raise their awareness of adjectives, learners can first be asked to identify all the qualitative adjectives in a well-drafted reflective writing essay and then to sort them out into categories according to meaning and usage. Their attention, in particular, can be drawn to attitudinal adjectives. To allow for interaction between students, groups/pairs of students may be asked to identify certain types of adjectives, after which an information exchange session is held.

The results of the current study regarding male and female students' use of adjectives are also important. While no significant differences were identified, it is important to develop a deeper understanding of how males and females approach the use of language generally and adjectives in particular in order to confirm or refute what others have found (Baker, 2014; Ishikawa, 2015; Lakoff, 1975; Tannen, 1990). One way to do this is to have students reflect on why they used particular adjectives in their writing. This can be done through brief reflections after they have completed a particular assignment. This would also serve as an ongoing reminder to students that it is important to remain cognizant of their use of adjectives as attitudinal stance markers.

Conclusion

In this paper, we attempted to describe how qualitative adjectives are used in reflective writing by first-year university students who are speakers of English as a foreign language. The nature of reflective writing requires authors to make references to their feelings, which necessitates the use of a significant number of adjectives. However, the students' use of adjectives in this study was relatively limited, pointing to the need for writing instructors to foster the development of students' reflective writing skills through purposefully designed pedagogical interventions.

We paid particular attention to the ways in which the male and the female students utilized adjectives in writing. The comparison between the data sets showed that the former had a greater tendency to use adjectives. However, the latter used adjectives as attitudinal markers more often. These findings point to gender differences in language use in reflective essays.

Considering the small scale of this study, the findings cannot be generalized to wider contexts. Further research is required for us to better understand the role adjectives play in students' pursuit of expressing themselves in reflective writing tasks. To this end, competent authors' writing habits can be investigated and compared to those of native and non-native students. This may reveal ways in which reflective writing instruction for EFL students can be better planned and implemented.

Conflict of interests

The authors declare that they have no conflict of interest.

References

Adams, K. (2013). *Expressive writing*: *Foundations of practice*. Rowman & Littlefield Education.

Al Murshidi, G. (2014). Emirati and Saudi students' writing challenges at U.S. universities. *English Language Teaching, 7*(6), 87-95. https://doi.org/10.5539/elt.v7n6p87

Alkhrisheh, H., Aziez, F., & Alkhrisheh, T. (2019). A study on gender and language differences in English and Arabic written texts. *Research and Innovation in Language Learning, 2*(2), 120-138. https://doi.org/10.33603/rill.v2i2.2028

Ayish, N. (2020). Attitudes toward using English as a medium of instruction among Emirati male and female freshman engineering students. In E. Mede, K. Dikilitas & D. Atay (Eds.), *Pedagogic and instructional*

perspectives in language education: The context of higher education (pp. 195-223). Peter Lang Publishers. http://dx.doi.org/10.3726/b16464

Baker, P. (2014). *Using corpora to analyze gender*. Bloomsbury

Barczewska, S., & Andreasen, A. (2018). Good or marvelous? Pretty, cute or lovely? Male and female adjective use in MICASE. *Suvremena Lingvistika, 44*(86), 194-213. http://dx.doi.org/10.22210/suvlin.2018.089.02

Beaumont, S. L. (2019). The art of words: Expressive writing as reflective practice in art therapy. *Canadian Art Therapy Association Journal, 31*(2), 55-60. https://doi.org/10.1080/08322473.2018.1527610

Bernard, A.W., Gorgas, D., Greenberger, S., Jacques, A., & Khandelwal, S. (2012). The use of reflection in emergency medicine education. *Academic Emergency Medicine, 19*(8), 978–982. https://doi.org/10.1111/j.1553-2712.2012.01407.x

Biber, D., Johanson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Pearson.

Birney, R. (2012). *Reflective writing: Quantitative assessment and identification of linguistic features* [Unpublished doctoral dissertation]. Waterford Institute of Technology.

Boud, D., Keogh, R., & Walker, D. (1985). *Reflection: Turning experience into learning*. Routledge, Abingdon, Oxon.

Boyd, E. M., & Fales, A. W. (1983). Reflective learning: key to learning from experience. *Journal of Humanistic Psychology, 23*(2), 99–117. http://dx.doi.org/10.1177/0022167883232011

Brown, S. (2016). *Effects of intrapersonal communication on reverse culture shock* [Research Report]. http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1236&context=comssp

Buckingham. L. (2008). Development of English academic writing competence by Turkish scholars. *International Journal of Doctoral Studies, 3*, 1-18. https://doi.org/10.28945/47

Burns, S., & Bulman, C. (2000). *Reflective practice in nursing*. Blackwell Science.

Burton, C. M., & King, L. A. (2004). The health benefits of writing about intensely positive experiences. *Journal of Research in Personality, 38*, 150-163. https://doi.org/10.1016/S0092-6566(03)00058-8

Çapar, M. (2014). Reflective writing in an EFL writing course. *International Journal of Language Academy, 2*(4), 470–486. https://doi.org/10.18033/ijla.118

Cisero, C. A. (2006). Does reflective journal writing improve course performance. *College Teaching, 54*(2), 231-236. https://doi.org/10.3200/CTCH.54.2.231-236

Coates, J. (2016). *Women, men and language* (3rd ed.). Routledge Linguistics Classics.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213–238. https://doi.org/10.2307/3587951

Crowe, M.T., & O'Malley, J. (2006). Teaching critical reflection skills for advanced mental health nursing practice: A deconstructive-reconstructive approach. *Journal of Advanced Nursing, 56*, 79–87. https://doi.org/10.1111/j.1365-2648.2006.03982.x.

Deveci, T. (2015). Communication students' use of lexis in a writing examination. *The Asian ESP Journal*, *17*(1), 39-63.

Deveci, T. (2017). Internet technology as an aid to traditional methods in the development of freshman students' use of academic words. *Journal of Teaching English for Specific and Academic Purposes, 5*(1), 55-76. https://doi.org/10.22190/JTESAP1701055D

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

Dyment, J. E., & O'Connell, T. S. (2011). Assessing the quality of reflection in student journals: A review of the research. *Teaching in Higher Education, 16*(1), 81-97. http://dx.doi.org/10.1080/13562517.2010.507308

Eckert, P. (1989). The whole woman: Sex and gender differences in variation. *Language Variation and Change, 1*(3), 245–268.

Eckstein, G., & Ferris, D. (2018). Comparing L1 and L2 Texts and Writers in First-Year Composition. *TESOL Quarterly, 52*(1), 137-162. http://dx.doi.org/10.1002/tesq.376

Englander, K. (2014). *Writing publishing science research papers in English: A global perspective*. Springer.

Ferris, D. R. (2009). *Teaching college writing to diverse student populations*. University of Michigan Press. https://doi.org/10.3998/mpub.263445

Forsell, S. (2011). *Promotion the European Union: Attitudinal stance adjectives in three genres of EU communication to public* [Unpublished Master's thesis]. University of Helsinki.

Gibson, A., Aitken, A., Sandor, A., Shum, S. B., Tsingos-Lucas, C., & Knight, S. (2017). Reflective writing analytics for actionable feedback. *Proceedings of the 7th International Conference on Learning Analytics and Knowledge* (pp. 153–162). ACM. http://dx.doi.org/10.1145/3027385.3027436

Hagevik, R., Aydeniz, M., & Glennon, R., C. (2012). Using action research in middle level teacher education to evaluate and deepen reflective practice. *Teaching and Teacher Education, 28*, 675–684. https://doi.org/10.1016/j.tate.2012.02.006

Handler, C., Handler, C., & Gill, D. (2011). *English and reflective writing skills in medicine: A guide for medical students and doctors*. Radcliffe.

Hardy, J. A., & Friginal, E. (2016). Genre variation in student writing: A multi-dimensional analysis. *Journal of English for Academic Purposes, 22*, 119-131. http://dx.doi.org/10.1016/j.jeap.2016.03.002

Hatton N., & Smith, D. (1995). Reflection in teacher education: Towards definition and implementation. *Teaching & Teacher Education, 11*(1), 33–49. http://dx.doi.org/10.1016/0742- 051X(94)00012-U

Ishikawa, Y. (2011). A corpus-based research on thanks in British dialogue. *International Proceedings of Economics Development & Research, 26*, 384-389.

Ishikawa, Y. (2015). Gender differences in vocabulary use in essay writing by university students. *Procedia - Social and Behavioral Sciences*, *192*, 593–600. http://dx.doi.org/10.1016/j.sbspro.2015.06.078

Jefferys, A. (2018, June 25). *Language to use for writing a reflective essay*. https://classroom.synonym.com/language-use-writing-reflective-essay-3292.html

Kember, D., McKay, J., Sinclair, K., & Wong, F. K. Y. (2008). A four-category scheme for coding and assessing the level of reflection in written work. *Assessment & Evaluation in Higher Education, 33*(4), 369-379. http://dx.doi.org/10.1080/02602930701293355

Kilgore, D., Sattler, B., & Turns, J. (2013). From fragmentation to continuity: Engineering students' narratives about the benefits of developing a professional portfolio. *Studies in Higher Education, 38*(6), 807-826. https://doi.org/10.1080/03075079.2011.610501

King, L.A. (2001). The health benefits of writing about life goals. *Personality and Social Psychology Bulletin, 27*, 298-807. https://doi.org/10.1177%2F0146167201277003

Klenowski, V. (2002). *Developing portfolios for learning and assessment: Processes and principles*. RoutledgeFalmer.

Krashen, S. (1984*). Writing: Research, theory and application*. Pergamon Institute of English. https://doi.org/10.1017/s0272263100006197

Lakoff, R. (1975). *Language and woman's place*. Harper and Row.

Lee, B., Fitria, A., & Ginting, H. (2019). Analysing literacy and other psychological tendencies using linguistic profile in English expressive writing: Are students able but unwilling to write? *Indonesian Journal of Applied Linguistics, 9*(1), 27-37. https://doi.org/10.17509/ijal.v9i1.12698

Lew, M. D. N. & Schmidt, H. G. (2011). Self-reflection and academic performance: Is there a relationship? *Advances in Health Sciences Education, 16*(4), 529–545. https://doi.org/10.1007/s10459-011-9298-z

Lu, X., & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing, 29*, 16-27. http://dx.doi.org/10.1016/j.jslw.2015.06.003

Lyubormirsky, S. Sousa, L., & Dickerhoof, R. (2006). The cost and benefits of writing, thinking about life's triumphs and defeats. *Journal of Personality and Social Psychology, 90*(4), 692-708.

MacLellan, E. (2004). How reflective is the academic essay? *Studies in Higher Education, 29*(1), 75-89. http://dx.doi.org/10.1080/1234567032000164886

Magno, C. (2009). Self-regulation and approaches to learning in English composition writing. *TESOL Journal, 1*(1), 1-16.

Mann, K., Gordon, J., & MacLeod, A. (2007). Reflection and reflective practice in health professions education: A systematic review. *Advances in Health Sciences Education, 14*, 595–621. https://doi.org/10.1007/s10459-007-9090-2

Mason, C. P. (1858). *English grammar: Including the principles of grammatical analysis*. Walton & Maberly.

Mauk, J., & Metx, J. (2013). *Inventing arguments*. Wadsworth.

Mezirow, J. (1998). On critical reflection. *Adult Education Quarterly, 48*(3), 185-198.

Moon, J. (2004). *A handbook of reflective and experiential learning: Theory and practice*. Routledge Falmer.

Moon, J. (2006). *Learning journals: A handbook for reflective practice and professional development*. Routledge.

Moussa-Inaty, J. (2015). Reflective writing through the use of guiding questions. *International Journal of Teaching and Learning in Higher Education, 27*(1), 104-113.

Murray, K. (2014). *Adjectives and adverbs*. PowerKiDS Press.

Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes, 45*(3), 211-236. https://doi.org/10.1080/01638530802073712

Nguyen, Q.D., Fernandez, N., Karsenti, T., & Charlin, B. (2014). What is reflection? A conceptual analysis of major definitions and a proposal of a five-component model. *Medical Education, 48*, 1176–1189. https://doi.org/10.1111/medu.12583

Nunn, R., & Brandt, C. (2016). A phenomenological approach to teaching reflective writing. *English Scholarship Beyond Borders, 2*(1), 130-142.

Paraskevas, C. (2006). Apprenticeship. *The English Journal, 95*(5), 65-70. https://doi.org/10.2307/30046591

Plack, M. M., Driscoll, M., Blissett, S., McKenna, R., & Plack, T. P. (2005). A method for assessing reflective journal writing. *Journal of Allied Health, 34*(4), 199-205.

Rayson, P., Berridge, D., & Francis, B. (2004). Extending the Cochran rule for the comparison of word frequencies between corpora. In G. Purnelle, C. Fairon, & A. Dister (Eds.). *Le poids des mots: Proceedings of the 7th International Conference on Statistical analysis of textual data (JADT 2004)* (pp. 926-936). Louvain-la-Neuve.

Reidsema, C. (2009). Assessing reflective writing: Analysis of reflective writing in an engineering design course. *Journal of Academic Language & Learning, 3*(3), 117-129.

Rodgers, C. (2002). Defining reflection: Another look at John Dewey and reflective thinking. *Teachers College Record, 104*(4), 842–866. https://doi.org/10.1111/1467-9620.00181

Rubin, D. L., & Greene, J. (1992). Gender-typical style in written language. *Research in the Teaching of English, 26*(1), 7-40.

Rusche, S., & Jason, K. (2011). You have to absorb yourself in it: Using inquiry and reflection to promote student learning and self-knowledge. *Teaching Sociology, 39*, 338-353. https://doi.org/10.1177%2F0092055X11418685

Russell, T. (2005). Can reflective practice be taught? *Reflective Practice, 6*(2), 199–204. https://doi.org/10.1080/14623940500105833

Ryan, M. (2011). Improving reflective writing in higher education: A social semiotic perspective. *Teaching in Higher Education, 16*(1), 99–111. http://dx.doi.org/10.1080/13562517.2010.507311

Schön, D.A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. Jossey-Bass.

Stillar, G. F. (1988). *Analyzing everyday texts: Discourse, rhetoric, and social perspectives*. SAGE.

Sumsion, J. & Fleet, A. (1996). Reflection: Can we assess it? Should we assess it? *Assessment & Evaluation in Higher Education, 21*(2), 121–130. http://dx.doi.org/10.1080/0260293960210202

Tannen, D. (1990). *You just don't understand: Women and men in conversation*. Ballantine Books.

Thabet, R. (2008). *Do public schools in UAE foster critical thinking as one of the main objectives of education?* xUnpublished Master's thesisъ. British University in Dubai.

Turns, J., Newstetter, W., Allen, J. K., & Mistree, F. (1997). *Learning essays and the reflective learner: Supporting reflection in engineering design education*. In Proceedings of the ASEE Annual Conference (pp. 2.274.1 – 2.274.14). Milwaukee, WI. https://doi.org/10.18260/1-2--6665

Tsingos-Lucas, C., Bosnic-Anticevich, S., Schneider, C. R., & Smith, L. (2017). Using reflective writing as a predictor of academic success in different assessment formats. *American Journal of Pharmaceutical Education*, *81*(1), Article 8. https://doi.org/10.5688/ajpe8118

Ullmann, T. D. (2019). Automated Analysis of reflection in writing: Validating machine learning approaches. *International Journal of Artificial Intelligence in Education, 29*, 217–257. https://doi.org/10.1007/s40593-019-00174-2

van Weijen, D. (2014, September 29). *How to overcome common obstacles to publishing in English*. https://www.elsevier.com/authors-update/story/publishing-tips/how-to-overcome-common-obstacles-to-publishing-in-english

Wallin, P., & Adawi, T. (2018). The reflective diary as a method for the formative assessment of self-regulated learning. *European Journal of Engineering Education, 43*(2), 507–521. https://doi.org/10.1080/03043797.2017.1290585

Wenjing, X. (2012). Study on gender differences in English. *International Conference on Education Technology and Management Engineering*, (16-17), 218-221.

West, M. (1953). *A general service list of English words*. Longman.

Wong, F. K. Y., Kember, D., Chung, L. Y. F., & Yan, L. (1995). Assessing the level of student reflection from reflective journals. *Journal of Advanced Nursing, 22*(1), 48-57. http://dx.doi.org/10.1046/j.1365-2648.1995.22010048.x

Zulfikar, T., & Mujiburrahman, A. (2018). Understanding own teaching: Becoming reflective teachers through reflective journals. *Reflective Practice, 19*(1), 1–13. https://doi.org/10.1080/14623943.2017.1295933

National Research University Higher School of Economics
Journal of Language & Education Volume 7, Issue 1, 2021

Gavriilidou, Z., & Mitits, L. (2021). The Socio-linguistic Profiles, Identities, and Educational Needs of Greek Heritage Language Speakers in Chicago. *Journal of Language and Education, 7*(1), 78-95. https://doi.org/10.17323/jle.2021.11959

The Socio-linguistic Profiles, Identities, and Educational Needs of Greek Heritage Language Speakers in Chicago

Zoe Gavriilidou, Lydia Mitits
Democritus University of Thrace

Correspondence concerning this article should be addressed to Lydia Mitits, Department of Greek Philology, Democritus University of Thrace, Komotini, Greece. E-mail: lydiamitits@gmail.com

The present study aims to further the research on heritage language speakers (HLSs) by providing the socio-linguistic profiles and identities of an uninvestigated community of heritage speakers, namely the Greeks of Chicago, thus offering data for a less-studied HL, Greek. The participants were fifty-four (N=54) first, second, and third-generation Greek HLSs. The socio-linguistic data were collected through an online survey, while identification with Greek culture as well as ethnic attachment and practice of Greek traditions were investigated through the content analysis of data from the Greek Heritage Language Corpus. The results of the study are discussed with respect to how they can improve our knowledge of the educational needs of Greek HL learners. This research-based knowledge can be employed for addressing the academic needs of HL learners through educational programs. The authors propose an agenda for a more linguistically and culturally responsive education program for HL learners, in general, and Greek HL learners in diasporic communities, in particular.

***Keywords*:** bilingualism, language profile, heritage language, heritage language speaker, identity, heritage language education

Introduction

Classrooms around the world are increasingly characterized by the presence of heritage speakers. Heritage languages (HLs) are spoken by simultaneous or sequential early bilinguals, commonly referred to as heritage speakers (HSs), whose one shared characteristic is a restricted knowledge of a home language mainly because of insufficient linguistic input. As a result, HSs can understand the language of the home and probably speak it to some degree but are more proficient in the dominant language of their host country (Polinsky, 2011). Previous research has shown that HLs are often related with identities and that HSs' identities are negotiated constantly, shaped by multiple variables, and are dynamic and socially constructed (Berard, 2005; Block, 2007; Valdés, 2001).

As far as Greek HLSs (GHLSs) are concerned, even though there are a considerable number of them in the U.S., Australia, and Canada, they are still in uncharted territory. Only recent research (Aravossitas 2016; Gardner-Chloros, 2009; Gavriilidou & Mitits, 2019; Karatsareas, 2018, 2019) has documented the linguistic competency and sociolinguistic characteristics of Greek or Cypriot-Greek HSs, so there is an urgent need to collect more data from GHLSs around the world in order to delineate their characteristics, investigate their identities, and address specific educational requirements stemming from their needs. Otherwise, as research on other HLs has shown, heritage Greek is faced with a loss of language contact and a pending language death with every subsequent generation of HLSs.

Thus, the purpose of the present paper is to extend previous research on HLs by offering data for a less-studied HL, Greek, and a HSs' community that has not been investigated yet - GHLSs of Chicago, in order to cast light on predictable (but variable) sociolinguistic profiles of GHSs. Based on previous research highlighting various parameters closely related to HSs' socioliolonguistic profiles, our specific aim is to investigate their age of arrival and possible contact with Greece, language preferences in various contexts, the formal instruction they might have received, exposure to Greek, self-assessed proficiency, attitudes toward Greek, and identity characteristics in order to discuss how these may connect with and what they imply about their educational needs when learning Greek as an HL.

Literature Review

The Community of Greeks of Chicago

Since 1820, the U.S. has been one of the principal destinations for Greek immigrants (Abbot, 1909; Kopan, 1989). Migratory flows reached a peak from 1900 to 1910 and from 1911 to 1920. The 1990 Census reported the number of people claiming at least one ancestor as Greek at 1,110,373, while 321,144 people older than five years of age spoke Greek at home. Most of the immigrants were from Laconia, notably, from the city of Sparta (a province of the Peloponnesus in southern Greece). From the 1890s, Greeks began arriving from other parts of the country, principally from Arcadia, another province in the Peloponnesus. They settled in major urban areas, including the industrial cities of the Northeast and Midwest (Kopan, 1989; Salutos, 1964). Initially, immigrants were mainly uneducated, low-skilled males who came to the U.S. in search of work. After the Immigration Act of 1965, new immigration waves consisted of better educated males and females in equal numbers as well as family groups (Kopan, 1989; Moskos, 1989).

One of the highest concentrations of Greek heritage speakers is in the Chicago metropolitan area. The Chicago's Greek Town is the oldest, largest, and one of the more lively Greek settlements in the U.S. (Kopan, 1989). Actually, by 1990, the U.S. census counted more than 70,000 people in metropolitan Chicago claiming Greek ancestry, approximately one-third in the city and two-thirds in the suburbs. The 2000 census counted 93,140 people of Greek ancestry in the metropolitan region. Community estimates, however, ranged from 90,000 to 125,000 people (http://www.encyclopedia.chicagohistory.org/pages/548.html).

Initially, the first generation of Greek immigrants in Chicago worked on railroads and in mines, but later they entered the service industries and after accumulating their own capital started their own businesses, mainly restaurants, confectionary shops, floral shops, and dry cleaning, and gained control of the wholesale commission market. The majority of the second generation of immigrants are college graduates (Kopan, 1989).

Heritage Speakers' Characteristics

In the last few years, HL research has focused on HL education, HL attrition, or structural characteristics of HLs, while less attention has, to date, been paid to the sociolinguistic profiles of HSs. In what follows, we foreground the role of sociolonguistic characteristics of HSs by offering a literature review of the main variables that affect HSs' sociolinguistic profiles, such as the age of arrival that determines the generation to which a HS belongs, the motivation of HSs to speak an HL, the quantity, quality, and situation of HL use, language attitudes, and so on. The fundamental factors listed above are included in the present study because they are both regularly and strongly implicated in HL maintenance (Kasstan et al 2018).

Age of Arrival and Contact with the Heritage Country

Previous research has stressed the correlation between proficiency in an HL, the age of immigration to the host country, and the identity of HSs, which is why these variables were included in the present study. The more years HSs have spent in the host country, the less proficient they become in their HL, since they may choose to integrate in the host country's culture and thus adopt a monocultural identity. Veltman (1988) claimed that all ethnic groups follow the two-generation model of adaptation. In other words, a shift to the majority language followed by a significant loss of the mother tongue begins with second-generation immigrants. Fishman (1991) observed a complete loss of the HL in most families within three generations. Valdés (1999) found that fourth-generation immigrants in the U.S. generally become monolingual English speakers. Tamošiūnaitė (2008) maintained that the turning point in the shift from the HL to the majority language is an eight-to-twelve-year stay in the immigration country and that the correlation between the age of arrival and linguistic proficiency corresponds with the correlation between the generation and linguistic proficiency.

According to relevant studies, first-generation HSs are more proficient than all other generations, demonstrate a monocultural identity, and often reject the language and culture of the dominant society. Second-generation HSs are characterized by simultaneous or sequential early bilingualism, depending on the time of the initial exposure to both languages. In both groups, however, the HL is the weaker one due to the limited exposure and input as well as the shift in the functional needs (the host country language is the formal language of education and the dominant language of the society). Being under the influence of two cultural systems, second-

generation HSs often undergo the demanding process of choosing an identity, which may result in the rejection of the HL and culture, more rarely in the rejection of the majority language and culture, or finally in a bilingual-bicultural identity (Lambert, 1975). In third-generation HSs, language attrition is predominant (Carreira & Kagan, 2011) and quite often accompanied by a monocultural identity. Furthermore, previous research has shown that that younger learners are more at risk to lose proficiency in their HL (Carreira & Kagan, 2011).

On the other hand, there is no previous research investigating the direct effect of contact with the HSs' home country. It may include travel to visit relatives, study abroad programs, volunteering, or internships, and could be classified as what is usually called 'heritage tourism'. Actually, visiting a country where the HL is spoken has many benefits. In addition to improving their language skills, HSs who travel to countries where their HL is spoken gain self-confidence, deepen their cultural knowledge, and strengthen their heritage identity. It seems that travel to heritage countries, with or without formal study, is one of the best ways to boost language skills in an HL. But this hypothesis needs to be supported with empirical data.

Language Preference and Use in Different Situational Contexts

The study of language preference and use of an HL in different situational contexts allows for predictions about the linguistic ability of the different generations of HSs and that was the main reason for incorporating this variable in the present research. It is evident that the majority of second and third-generation HSs in the U.S. are, if not English dominant, English preferent. Some of them may continue to function in two languages, especially for communicating with family or community members of the first generation. Valdés (1999) noticed that HL use is restricted to everyday discourse; it is narrower in range and is a low variety, appropriate mainly for casual, informal everyday interactions. Other researchers (Chevalier, 2004; Shin, 2013) pointed out that the restricted use of an HL in informal settings in home and local communities is often the result of monolingual educational policies, which also lead to monocultural identities. These restricted situational contexts in which HLs are used, compared to the use in home countries, lead to a number of idiosyncratic changes and attrition in all linguistic levels of the HL. This attrition results in HSs transferring their HL "in a 'mutilated' form to the next generation" (Valdés, 1999: 10). Functionally, though, HSs may be able to carry out conversations on everyday topics with ease and a relevant fluency, and may even be able to understand spoken language that includes the use of humor.

Formal Instruction in an HL

The correlation between the years of formal instruction and proficiency in an HL has been duly acknowledged in previous research and it was decided to further investigate this variable in the present study as more years of instruction may ensure a higher level of proficiency (Kagan, 2005; Kagan & Dillon, 2008). Some studies have focused on the effect of specific instruction programs (Potowski et al., 2009) on HSs' proficiency in an HL. What is important to note here is that instruction in an HL aims not only to enhance HSs' linguistic ability but also to achieve community advancement, since it addresses broader academic and identity challenges and may strengthen cultural ties between HSs and their home and community. Furthermore, it cultivates a positive attitude toward the HL as it can counteract the hegemony of the majority language (Martinez, 2016).

Various types of programs for HL teaching have emerged in recent years: school based-programs in primary and secondary schools, higher-education programs in private and public universities, and finally community-based programs which offer after-school, Saturday-school, or summer school instruction that is hosted in museums, cultural, religious, or community centers (Beaudrie, 2016).

Actually, Greek Orthodox churches in the U.S. and more specifically in Chicago started establishing parochial schools in the 1900s, some of which are full-day schools with a bilingual English and Greek curriculum, while others are afternoon and Saturday schools with only a Greek-language curriculum. According to data reported in Michopoulos (2008), the total enrollment of Greek ancestry pupils was 6,000 pupils in day schools and 30,000 pupils in afternoon or Saturday Schools. While the vast majority of Greek children attend Chicago public schools (except for those enrolled in Greek day schools), practically all Greek children attend afternoon (following public school attendance) and Saturday schools, where they learn the rudiments of the Greek Orthodox faith along with Greek language and culture (Kunkelman, 1990). In these schools, teachers are usually volunteers, with no previous teacher training, teaching methods are often outdated, and traditional and pedagogic material is often obsolete; as a result, these schools are not always attractive to young Greek HSs. However, since the U.S. educational system does not prioritize bilingualism in different

heritage languages, specifically in Greek, such programs fill this void and play a crucial role in Greek language and culture maintenance (Nikolidakis, 2005). It should be noted that there are also numerous cultural organizations and unions promoting Greek cultural heritage and traditions. Thus, such programs in combination with the education received by HSs in mainstream schools help them acquire a unique bicultural identity and a sense of belonging to two cultures, and also assign equal status to the two cultures and languages.

Exposure to and the Use of an HL

Exposure is a predictor of possible bilingual or HL development, but we may also add maintenance, as reported in previous studies (De Houwer, 2007; Dixon et al. 2012), which is why it was deemed necessary to include this parameter in the present study. The length and manner of HSs' exposure to an HL may vary according to socio-economic parameters and this variation can have a great impact on mastering it. Research has shown that HSs do not have much exposure to their HL outside the home (Carreira & Kagan, 2011) and that parental involvement can be crucial for HL and culture maintenance. For instance, Park and Sarkar (2007) found that one way that parents helped their children maintain Korean as an HL in Montreal was through encouragement of the internet use in Korean. Prevoo et al. (2014) reported that reading to children in the HL may boost children's' vocabulary in that language and expose them to a broader range of concepts and perspectives. Furthermore, it is the parents' wish to expose their children to the family HL that motivates them to enroll their children in community-based programs for HL instruction (Liu, 2013). It should be noted that activities such as reading literature in the HL, navigating the internet, watching TV, reading newspapers, etc., may help HSs maintain contact with their HL and culture.

Language Proficiency in an HL

Proficiency in an HL can be defined narrowly or broadly. It can refer to reading, writing, speaking, and listening skills in an HL or to basic linguistic abilities such as vocabulary or grammar (Zyzik, 2016). Some scholars claim that proficiency in an HL does not necessarily mean knowledge of the culture carried by the language and vice-versa (Hoffman, 1991). Others believe that "the further one progresses in bilingual ability, the more important the bicultural element becomes, since higher proficiency increases the expectancy rate of sensitivity towards the cultural implications of language use" (Beardsmore, 1986: 24).

Different research protocols have been used to assess the language proficiency of HSs such as self-report assessments (Kagan & Friedman, 2003), vocabulary knowledge (Polinsky, 2008a), speech rate (Polinsky, 2008b), cloze-tests (Montrul & Foote 2014), picture naming (Montrul et al., 2014), etc. (for the proficiency assessment of HSs, see Ilieva & Clark, 2016). Regardless of the instrument used to assess proficiency, the majority of studies have concluded that HSs demonstrate strong oral skills, while other skills lag behind (Ilieva & Clark, 2016). Valdés (1999: 11) observed that HSs "may seem quite superior in some respects and quite limited in others", e.g. HSs' reading and writing skills may be restricted or even nonexistent (Valdés, 2000). Similarly, Carreira and Kagan (2011) found that most HL learners who had acquired English after acquiring their HL could understand and speak the HL better than they could read and write it. Zyzik (2016) noted that HSs' language proficiency is limited to basic-level cognition which includes listening and speaking, and high frequency lexical items and frequent grammatical expressions commonly used in everyday interactions. In the same vein, Swender et al. (2014) in their study of Russian and Spanish HSs found that their sample demonstrated an inability to express abstract thinking and reasoning and lacked extended discourse and precise vocabulary.

Attitudes towards the HL

Research (Cho et al., 1997; Park & Sarkar, 2007) has shown that positive attitudes towards HLs are interrelated with the desire to learn an HL, successful HL learning, and/or different patterns in HL maintenance (Polinsky, 2018). It was, thus, deemed necessary to examine this variable. Research on socio-cultural aspects of HL learning has emphasized that attitudes toward an HL are not static but emerge from interactions and social practice (Zyzik, 2016). Furthermore, HSs tend to express a growing interest in maintaining their HL, leading to an increased enrollment in HL instruction classes (Lee & Kim, 2008; Yang, 2003). Carreira and Kagan (2011: 48) investigated HL learners' attitudes towards their HL and found that most of the students had "largely positive feelings and experiences" with their HL. Similar results were reported by Polinsky (2011), but also by Karatsareas (2018) who found a positive attitude among the members of the Greek-Cypriot Community in London towards their HL.

Heritage Speakers' Identities

Since language has a symbolic value, is a carrier of cultural capital, and determines one's conception of the world and one's relation to otherness and position in society, it is in an intimate relation with one's identity and is actually a strong indicator of a person's or group's identity (Lee, 2002). HSs connect with at least two linguistic communities and cultures and are in constant negotiation within a multicultural environment (Crawshaw et al., 2001). Whether they develop a dual linguistic and cultural identity or not depends on (a) how able they are to adopt hybrid-situated, non-conflicted identities in the frame of the dominant and heritage socio-cultural discourses (He, 2006) and on (b) how much the educational and linguistic policies adopted aim at strengthening dual identities or just simply promote monolingualism, thus weakening HL learner programs.

Thus, based on the above review of the literature and the need for investigations into the group of GHLSs, the following research questions were posed:

RQ1. What is the socio-linguistic profile of GHLSs living in Chicago?
RQ2. What are the GHLSs' defining elements of their Greek identity?
RQ3. What are the educational needs of GHLSs based on RQ1 and RQ2?

Materials and Methods

Participants

The present study investigated fifty-four (N=54) GHLSs. The participants were included in the study by convenience sampling, after contacting the Greek community in Chicago. Informed consent was obtained from all participants. There were twenty-three (N=23, 42.6%) male and thirty-one (N=31, 57.4%) female participants whose ages ranged from <12 to 55+ (see Table 1), with all education levels from primary education to postgraduate degree holders represented (see Table 2). All procedures performed in the study involving human participants were in accordance with the ethical standards of the institutional ethics committee of Democritus University of Thrace (60589/2111/31-8-2018).

Table 1
Age range of the participants

<12	12-17	18-22	23-28	29-40	41-55	55+	Total
8	10	2	2	9	20	3	54
14.8%	18.5%	3.7%	3.7%	16.7%	37%	5.6%	100%

Table 2
Education level of the participants

Primary	secondary	university	post-graduate	Total
11	12	17	14	54
20.4%	22.2%	31.5%	25.9%	100%

Instrumentation

The data collection instrument used to answer RQ1 was the 29-item *Greek Heritage Language Background Questionnaire* (GHLBQ), adapted to and validated in Greek, English, and Russian within the frame of the project *Greek Heritage Language Varieties*. Its design was based on a survey that investigated the reading skills of HL learners (Jensen & Llosa, 2007) and parts from other surveys available from the *National Heritage Language Resource Center* (Carreira, 2009; Gignoux, 2009; Lyutykh, 2012; Montrul, 2012; Torres, 2012). The design and validation procedure of GHLBQ followed the adaptation protocol described in Gavriilidou and Mitits (2016).

The GHLBQ contains 29 questions in total. More specifically: (a) three questions (Q 1-3) on demographic information about the participants (age, gender, and the level of education), (b) six questions (Q 4-9) on biographical information related to the contact with the languages they use, (c) five questions (Q 10-14) on language use and preference, (d) one question (Q 15) on heritage language instruction, (e) seven questions (Q 16-22) on previous exposure to written/spoken heritage and dominant languages, (f) four questions (Q 23-26) on self-rated language proficiency, and (g) three questions (Q 27-29) on their attitude towards the languages they spoke.

The Cronbach's alpha coefficient for the 46 five-point Likert scale items grouped in questions 23-26 for assessing self-rated language proficiency was .969, suggesting an excellent degree of internal consistency. The survey was administered online to GHLSs in the spring of 2019.

To investigate RQ2, the authors relied on data from the Greek Heritage Language Corpus (GHLC), which is a speech corpus comprising 90 hours of recordings of speech from Greek HSs living in Moscow, St. Petersburg, and Chicago. More specifically, the study investigated the 85,000 words found in the Chicago sub-corpus.[1]

Finally, answers to RQ3 were based on the sociolinguistic profile of the sample as well as on supplementary data from the Chicago sub-corpus of the GHLC.

Data Analysis

To investigate RQ1, SPSS v.24 Multiple Dichotomy Analysis was used for the multiple response sets and descriptive statistics were used in order to compute the frequencies of the responses containing nominal data from the GHLBQ. Content analysis was adopted for the analysis of the data obtained from the GHLC in order to answer RQ2 and RQ3.

Results

The results of the quantitative analysis for RQ1 are presented in the following section, while content analyses of the GHLC data for RQ2 and RQ3 are included in the discussion section.

Age of Arrival and Contact with Greece (Qs 4-9)

It was found that the majority of the participants (N=43, 79.6%) were born in the U.S., while 11 of them (20.4%) reported being born in Greece. Those participants who were not born in the host country arrived there at an early age <14, with the exception of two cases who arrived between the ages of 14 and 18. All of them had attended school in Greece before emigrating. However, the school years they completed varied (see Table 3). There were also cases of those who were born in the U.S. but then moved to Greece for a period of time before returning.

Table 3
Years of school in Greece

1-2 years	3-5 years	6-8 years	9-12 years	12+ years
4	3	1	2	1
36.4%	27.2%	9.1%	18.2%	9.1%

As for their visits to Greece, only 3 out of 54 have never traveled there, while a large number of the respondents have visited Greece three to five times (n=26, 48.1%) and one-third (n= 17, 31.5%) go there at least once a year. Lastly, when it comes to the language that they learned to read in first, the participants report an almost even distribution among English, Greek, and both languages (n=16, 29.6%; n= 19, 35.2%; n=19, 35.2% respectively).

[1] The GHLC is available at http://synmorphose.gr/index.php/el/projects-gr/ghlc-gr-menu-gr/ghlc-transcriptions-sample

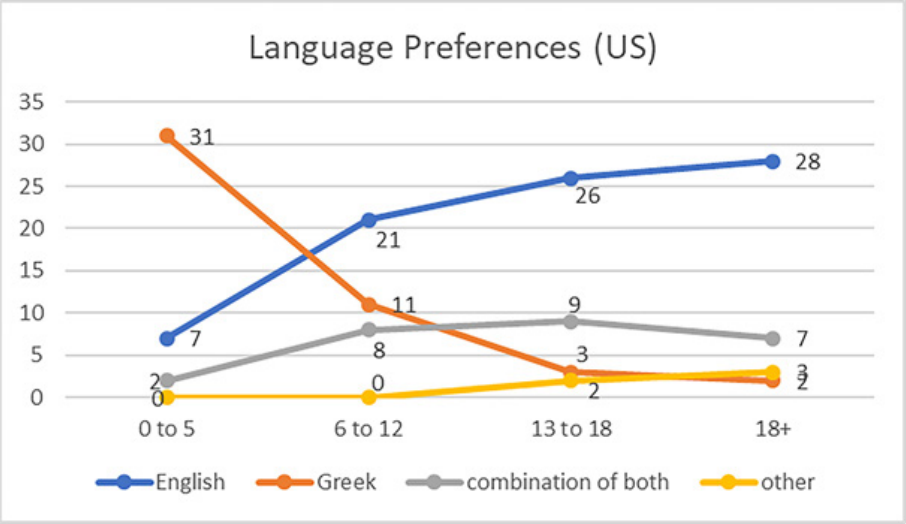
Language Preference and Use in Different Situational Contexts (Qs 10-14)

Questions 10-14 investigated language use and preferences in everyday communication. More precisely, these questions focus on determining which language(s) the participants prefer to use and with whom, as well as whether those preferences changed at different periods of their lives. The results showed (a) a decline in the use of Greek as HSs age and (b) an extended use of Greek only at home or with family.

When asked what language they spoke most of the time, only two (3.7%) opted for Greek while the largest number (29, 53.7%) chose English only, followed by both English and Greek (N=23, 42.6%). However, 70.3% of the participants reported using both English and Greek, and 16.7% only Greek when speaking to their family members as opposed to merely 13% who use English at home. When asked which language(s) they spoke with their friends, the participants reported the host country language first (n=39, 72.2%), the combination of dominant and HL second (n=13, 24.1%), and Greek (N=2, 3.7%) third.

The study produced very interesting findings with respect to language use preference changes over time (see Figure 1).

Figure 1
Language preferences



There is a dramatic decline in the speakers' use of Greek as they age. While 74.1% (n=40) of respondents said they used Greek predominantly before the age of 5, only 3.7% (n=2) aged 18+ reported the same. At the same time, the number of those who use both languages on a daily basis seemed to increase by age before it drops after the age of 18.

Formal Instruction in Greek as an HL (Q 15)

The type and amount of formal instruction refers mainly to community-based instruction with the majority (N=39, 72.2%) of the participants receiving more than four years of instruction in Greek as an HL (see Table 4).

Table 4
Formal instruction in Gree

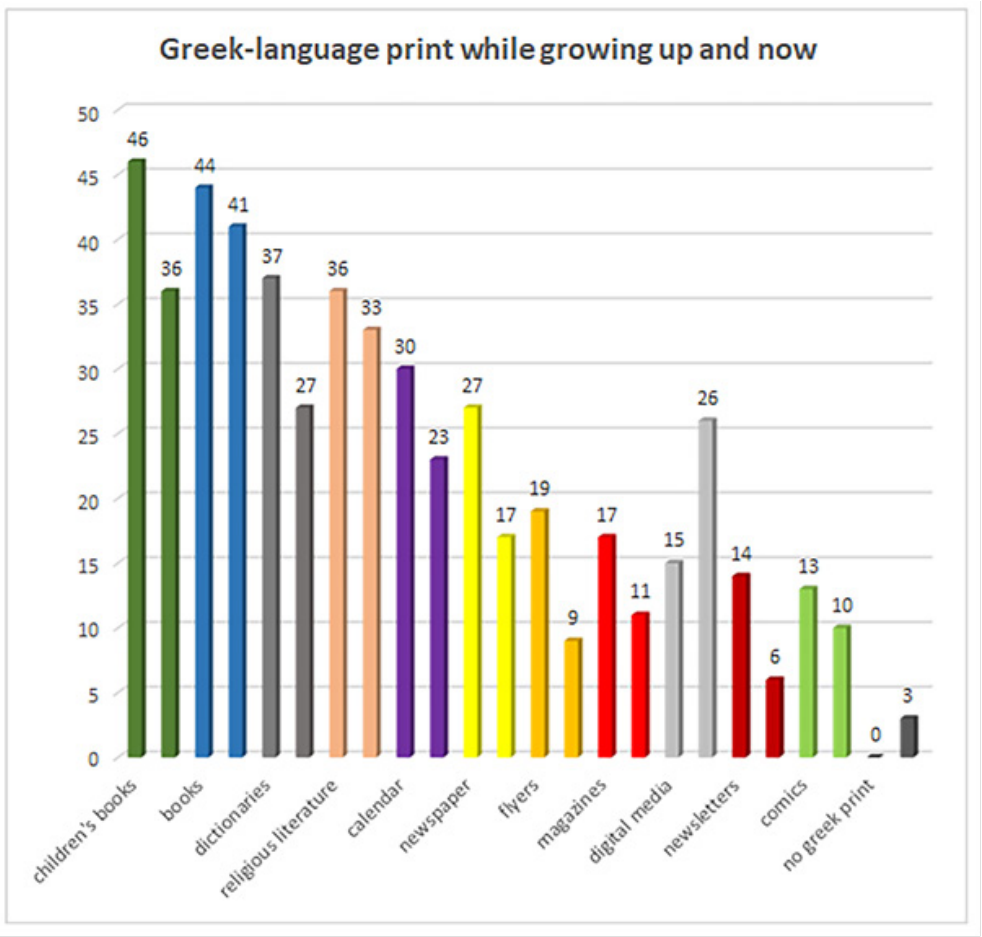
Have you studied Greek at a community/church school?					
Never	**1 year or less**	**2-3 years**	**3-4 years**	**More than 4 years**	**Total**
7	3	2	3	39	54
13%	5.6%	3.7%	5.6%	72.2%	100%

Exposure to and Use of Greek as a Heritage Language (Qs 16-22)

Next, the study investigated the exposure to Greek that the participants have had throughout their lives, mainly to the written language, which is generally underrepresented in the upbringing of HSs but also to certain tasks and activities that require the comprehension and use of spoken Greek.

Question 16, *What type of Greek print did you have as you were growing up*, revealed that the GHLSs grew up with a variety of Greek print media at home, with children's books (N=46, 85.2%), books (N=44, 81.5%), dictionaries (N=37, 68.5%), religious literature (N=36, 66.7%), and calendars (N=30, 55.6%) the most widely represented. They were followed by newspapers (N=27, 50%) and to a lesser percentage by flyers, magazines, digital media, and comics (between 35.2% and 24.1%). However, the above numbers drop when compared to what Greek-language print they have at their home at present (Q17), except for digital media, the use of which has risen from 27.8% to 48.1 % (see Figure 2).

Figure 2
Greek-language print



As parental involvement plays an important role in HL development, the participants were asked to select what type of print they were read to as children by their family members. Their responses unveiled that those were mostly children's books (N=46, 85.2%), literature (N=29, 53.7%), religious literature (N=17, 31.5%), and comics (N=10, 18,5%).

Accessing the internet in Greek is another opportunity for the GHLSs to get additional language input. Although a general increase in the exposure to digital media in Greek was documented by the previous question, GHLSs do not appear to take advantage of this particular tool since only 10 of them (18.5%) use it frequently, whereas 35 (64.8%) and nine (16.7%) do so rarely or never respectively. In general, the time

our participants spend reading non-school related material in Greek varies between not at all (n=5, 9.3%) and more than two hours (n=6, 11.1%) with the majority of them reading for less than 15 minutes (n=23, 42,6%).

Table 5 shows, in descending order, which activities involving contact with Greek the participants have been engaged in recently. Their exposure to and use of Greek comes mostly from their oral communication in informal contexts, their contact with the popular cultural elements such as music and TV, and their participation in Greek cultural/religious celebrations and events.

Table 5
Activities done in Greek

In the past six months, which of the following have you done in Greek?	N	Percent
Spoken on the phone	51	15.2%
Listened to music	49	14.6%
Attended an event	45	13.4%
Watched TV	34	10.1%
Listened to the radio	30	8.9%
Written an email/letter	27	8.0%
Visited a website	27	8.0%
Used Facebook	19	5.7%
Read a book/short story	17	5.1%
Watched a movie	15	4.5%
Tweeted/chatted online/used Instagram	14	4.2%
Read a newspaper	8	2.4%
Total	336	100.0%

Self-Assessment of Language Skills (Qs 23-26)

Most respondents rated their writing as the least developed of their skills in Greek (low 24.1%, intermediate 38.9%, advanced 22.2%, and native-like 14.8%). It was followed by reading (low 11.1%, intermediate 44.4%, advanced 27.8%, and native-like 16,7%). Speaking was self-rated higher with only 3.7% of the participants reporting low skill, 46.3% intermediate, 31.5% advanced, and 18.5% native-like. They were more confident about their listening skill, with 33.3% rating it as native-like, 29.6% each as advanced or intermediate, and 7.4% as low. See Figure 3 for the frequencies of the responses.

When asked to self-rate their English language skills, the respondents provided answers for the four skills that depicted an equal distribution of responses as the vast majority feel they have mastered them. Writing (1.9% intermediate, 7.4% advanced, and 90.7% native-like) and reading (1.9% intermediate, 5.6% advanced, and 92.6% native-like) were rated lower than listening and speaking (9.3% advanced and 90.7% native-like). See Figure 4 for the frequencies of the responses.

The acquisition of the language skills was further investigated by asking the respondents to rate how difficult it is for them to read numerous genres. The genres that most respondents find difficult to read are academic/technical papers, poetry, novels, theatrical plays, non-fiction, and textbooks, while flyers, dictionaries, letters, and emails were rated as being easy to understand by most of them.

Next, when speaking Greek, the respondents find it relatively easy to very easy to accomplish most of the tasks they were asked about, such as telling a joke, a fairy tale, a story, using polite language, or being rude. Talking about current events or debating an idea show a moderate level of difficulty, while only giving a formal presentation was rated as challenging. Finally, the GHLSs are more confident in their listening skills and judge most of the tasks (eavesdropping, understanding humor, songs, TV, and movies) as relatively easy to very easy to accomplish, except for formal talks and news reports, which have a more equal distribution of answers from difficult to very easy.

Figure 3
Greek language skills

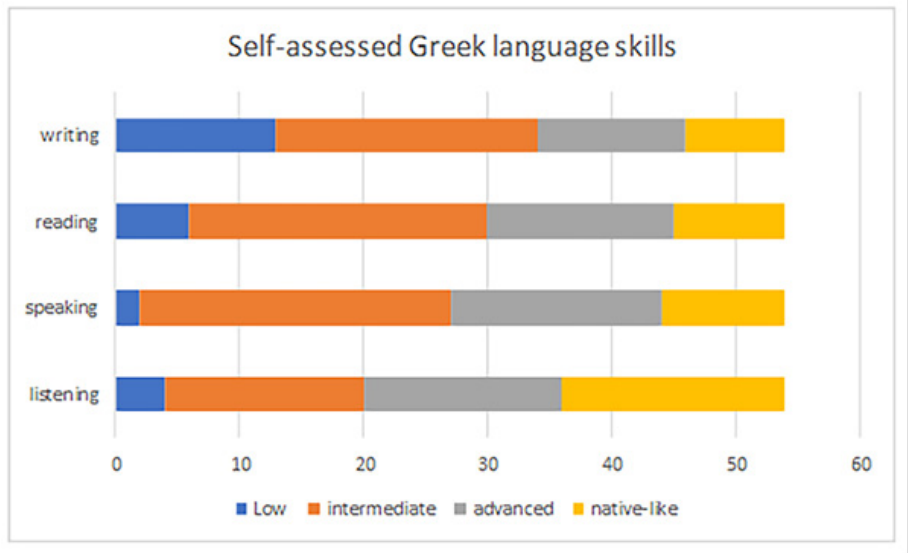
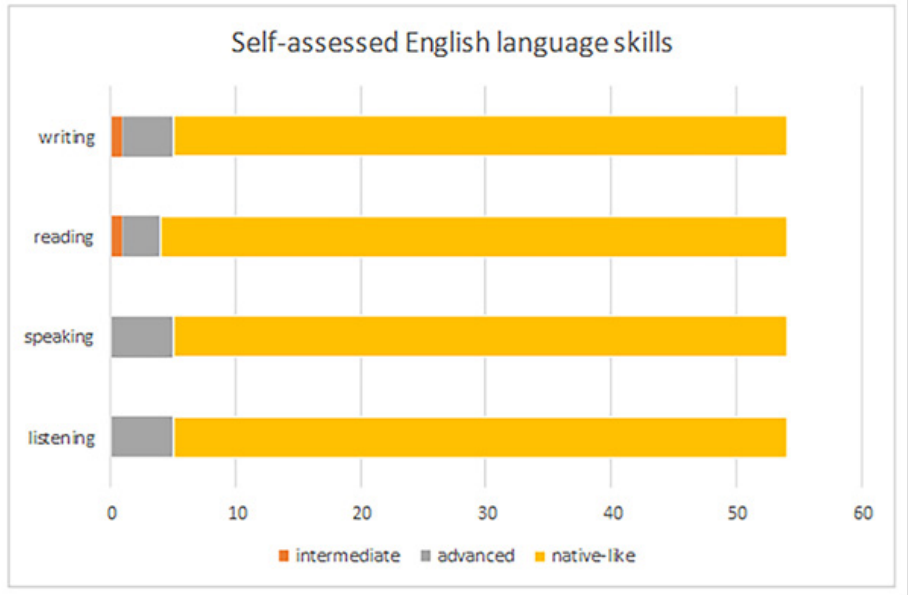


Figure 4
English language skills



Attitudes towards Greek (Qs 27-29)

An investigation of GHLSs' attitudes towards their HL was done with the last set of questions. When asked what language(s) they preferred to speak, the majority preferred their dominant language (English) whereas about one-third of the respondents favored speaking a mixture of both (see Table 6).

Table 6
Language preference

What do you prefer to speak?				
Greek	**English**	**a mix of both**	**no preference**	**Total**
3	31	17	3	54
5.6%	57.4%	32.5%	31.5%	100%

It was found that the overwhelming majority of participants reported having their family's support in maintaining Greek (98.1%). Similarly, all of them would want to teach their children Greek (100%). Although the scale tips towards English as a preferred language (particularly among the 2nd and overwhelmingly among the third generation), the results showed a very positive attitude toward Greek.

Discussion

The purpose of the present study was to investigate the sociolinguistic profiles of GHSs (RQ1) and to discuss how these may connect with their identities (RQ2) and what they imply about their educational needs when learning Greek as an HL (RQ3).

RQ1: What is the Socio-Linguistic Profile of GHLSs Living in Chicago?

The results of the first set of questions (Questions 4-9) of the GHLBQ allowed for the initial profiling of our participants concerning the HS generation they belonged to and the type of bilingualism they had developed. Thus, the first generation were successive bilinguals in Greek and English, the second generation were either sequential or simultaneous (English dominant) bilinguals, while the third generation were simultaneous, English-dominant bilinguals.

By comparing the participants' responses with their language backgrounds (Questions 10-14), we found a trend with respect to which generation of HSs they belonged to. The first generation prefer and maintain Greek as the language they still feel more proficient in. The second generation (sequential or simultaneous English-dominant bilinguals), who have been given opportunities to stay in contact with their HL continuously and systematically, prefer to speak a combination of both languages. The third generation (simultaneous English-dominant bilinguals), whose contact with Greek is mainly through formal education, prefer English. Karatsareas (2019) reported a similar linguistic situation with Cypriot-Greek heritage speakers in London. Also, data presented in this paper confirm previous findings (Cho et al., 2004; Montrul, 2008, 2016) showing that there is a shift in dominance from the HL to the majority language in the transition from the first to the second generation and a possible loss of the HL by the third generation. Our study also documents this dramatic decline. It occurs particularly after the age of 18 as a result of interrupted formal schooling in Greek but also because of language preferences only in restricted contexts (family meetings, conversations with friends, socializing, etc.). As our results do not confirm an extended use of Greek at home or with family, this can help explain the shift in the language preference from Greek to English over time and from generation to generation. The finding of restricted use of Greek at home may be attributed to the fact that parents and children are often pressured to speak only English at home, a practice that "not only contributes to heritage language loss, but also is likely to lead to a less enriched language environment with fewer opportunities for interactions about interesting topics incorporating sophisticated vocabulary, ideas, and concepts, given that many immigrant parents have limited English proficiency" (Lee & Oxelson 2006: 462-463).

The results of Question 15 about formal instruction in Greek indicated that GHSs who have never attended day or afternoon/Saturday schools belong to the first-generation GHLSs who had studied in Greece before moving to the United States. From our results, it can be concluded that GHLSs systematically receive formal education in their HL. However, when interviewed, GHSs commented that when they were younger, they were forced by their parents to attend such schools, and as a result they lacked motivation to learn Greek, while later they regretted doing so.

When it comes to the quantity and variety of written Greek they have been exposed to since their early years (Questions 16-22), it was found that it depends on parental involvement, whose importance has been documented in related studies (Li, 2006). In other words, it is the close family that provides both spoken and written contact with the HL and the quality of linguistic input is a determinant of how proficient a HS may become. It was also found that accessing the internet in Greek is quite restricted. This finding confirms Carreira and Kagan (2011), who claimed that most HSs are not electronically literate in their HL as they use the internet in English rather than in their HL. Another interesting finding was that the participants did not have reading experiences in Greek outside school. This finding highlights the fact that HSs' knowledge of Greek is not supported by extracurricular reading that could improve their academic vocabulary and writing skills in Greek.

This finding is in line with Carreira and Kagan (2011), who found that most HL learners who had acquired English after acquiring their HL did not have much exposure to their HL outside the home.

GHSs' self-rated proficiency was also investigated (Questions 23-26). It can be argued that self-assessed proficiency as an approach to measuring the level of language acquisition has its drawbacks, in that it requires linguistic and metalinguistic awareness on the part of HSs. Moreover, speakers tend to assess their knowledge as lower or higher, showing an inverse correlation between a person's self-assessment and their actual proficiency (see Polinsky, 2011). However, studies have shown that HSs' self-assessment correlates very well with independent measures of proficiency (Montrul et al. 2010). Furthermore, asking the participants to self-rate their proficiency in English helped us to further validate our method and confirm the claim that HSs are more proficient in the language of the host country, particularly those belonging to the second and third generations. Thus, in light of the previous findings, we hypothesized that the sample would report a higher self-rated proficiency in listening and speaking and lower in reading and writing, which was confirmed. This finding is in line with Valdés (2000) and Carreira and Kagan (2011), who also found that HSs could understand and speak the HL better than they could read and write it. Actually, the reading and writing abilities of HSs depend on the amount of formal instruction that HSs have received in their HL. However, a different pattern was found when they were asked to self-rate their proficiency in English, where they did not differentiate much between the four skills in English. This finding supports the generally expressed views that HSs are by definition dominant in the host language (Montrul, 2016; Polinsky, 2011).

To further investigate the acquisition of language skills in HG, the respondents were asked to rate how difficult it is for them to read in various genres. It was found that academic/technical papers, poetry, novels, theatrical plays, non-fiction, and textbooks were rather difficult for them. This finding agrees with previous research (Leeman & Serafini, 2016), which showed that HSs are not acquainted with the registers associated with academic or professional settings. Sánchez-Muñoz (2016) attributed these drawbacks in academic language to different social contexts.

It is obvious that the development of language skills observed in our study follow the order of the L1 acquisition process of a child naturally learning the language of the home. Namely, the participants rate their linguistic abilities regressively starting from listening, speaking, reading, to finally writing. Moreover, the phenomenon of the dominant language overtaking the HL (Campbell & Rosenthal, 2000; Polinsky & Kagan, 2007) is documented by very high self-rating skills in English as opposed to Greek. Similarly, oral production (in particular tasks in English) poses no problem, while more limited contact with spoken modalities of Greek lead to lower self-rated speaking ability.

Questions 27-29 investigated GHSs' attitudes towards Greek, which proved to be highly positive. Studies have shown that many parents view HL maintenance as a way to preserve their cultural identity and wish for their children to maintain it (Cho et al., 1997; Guardado, 2010; Hashimoto & Lee, 2011; Lao, 2004; Li-Yuan & Larke, 2008; Wong Fillmore, 2000). Others have suggested that parental attitudes are strong indicators of children's success with HL maintenance (Garcia, 2003; Lao, 2004; Wong Fillmore, 2000). Further findings point to positive parental attitudes being directly associated with more successful maintenance (Hashimoto & Lee, 2011). Our results also agree with Polinsky (2011) and Karatsareas (2018) who report a similar positive stance among their participants towards their HL. Carreira and Kagan (2011: 48) found that most of their students had "largely positive feelings and experiences" with their HL. In our study, Greek is viewed as a symbol of GHSs' ethnolinguistic identity, culture, and history that has to be preserved and there is a subsequent strong desire, especially from the first and second generation, to maintain it.

RQ2: What are the GHLSs' Defining Elements of their Greek Identity?

Data from the GHLC showed that the two most salient indicators of the Greek identity are the Greek language and Greek orthodox religion. First-generation HSs demonstrate high proficiency in Greek and claim they are Greek because they were born and raised in Greece. They relate to Greek more than to American culture, are active members of the Greek community of Chicago, and go to local Greek-orthodox churches every week (there are 26 of them in the Chicago area). They also actively participate in events organized by local Greek communities and follow all Greek customs and traditions (e.g. Easter, Christmas, etc.). They prefer Greek food, wish to preserve the Greek identity of their children by sending them to community-based schools, feel

nostalgic about what they left behind when they immigrated, and have a sense of belonging to and love of Greece. As a result, they visit relatives in Greece frequently and some of them say that they hope to return to Greece when they retire. For most of them, cultural heritage is considered as a form of support in their life in the host country.

On the other hand, there is a variation in second and third-generation HSs. The more their parents have engaged them in cultural and social activities associated with the Greek community and culture, the higher sense of appreciation for Greek language and culture the children have. In such cases, second-generation HSs combine Greek and American culture and adopt a bicultural Greek-American identity as can be seen from the extracts from interview 1000 from the GHLC in (1a) and (b) below:

> (1a) *When I was younger, our mom was alive and we used to go to church during the Holy Week. Now that my mother passed, we don't go very often. My father is old and sometimes he can't go, but when he is well, we go. On Good Wednesday we usually go to church to take the unction. On Good Friday all Greek churches get out the Epitaph and go around the church […] On Good Saturday we used to go to Church for the Resurrection and then we used to go home, and my dad and mom prepared avgolemono[2]. This year my dad didn't prepare avgolemono, he was sick, but in the past, he also used to cook avgolemono and we used to come home at two or three in the morning to eat it.[3]*
>
> (GHLC, interview 1000, l. 10-43)[4]
>
> (b) *Here in America, we have a custom called Halloween. Small children dress up as witches, go to neighbouring houses and ask for sweets and people give them sweets. We also used to go when we were kids.*
>
> (GHLC, interview 1000, l. 111-118)

Second-generation HSs (e.g. interviewee 1000) maintain a balance by adapting to American culture without ignoring their Greek background. However, in cases where second-generation parents did not invest in the preservation of Greek culture, sometimes due to their workload or because one of the two parents is not Greek, third-generation children feel mainly American, distance themselves from Greek customs, traditions, and/or cuisine, and very often demonstrate a low proficiency in Greek.

> *- Is your wife Greek?*
> *- No, she isn't.*
> *- Does she speak Greek?*
> *- She understands a little bit, but she doesn't' speak.*
> *- Do you speak Greek with your children?*
> *- A little, but they don't go to Greek school. I bought a computer for them to learn Greek. They go to American school the whole day and to after-school athletic activities and more classes, you know, it's not easy. And I work a lot.*
>
> (GHLC, interview 1001, l. 48-59)
>
> *I haven't eaten magiritsa[5] since I was a kid. Nobody eats magiritsa nowadays, only old men and women. My mom used to cook it.*
>
> (GHLC, interview 1001, l. 76-77)

It is important to highlight here that, since language proficiency in the HL is low for third-generation HSs, the Greek orthodox religion seems to be the defining element of Greek identity for this generation. This finding supports Lee's (2002: 120) claim that "the loss of one marker does not automatically entail the loss of cultural identity and secondly, the loss of one or more identity markers tends to reinforce the remaining identity markers."

RQ3: Implications for Greek Heritage Language Education Based on RQ1 and RQ2

[2] Avgolemono (or magiritsa) is a traditional soup eaten after the Resurrection Ceremony in Greek Easter.
[3] The original interview extract from the GHLC is in Greek and is translated by the authors in English for the purpose of the present study.
[4] This citation indicates the number of the interview and the number or the lines of the specific interview found in the Greek Heritage Language Corpus (GHLC), from which we extracted the data for this paper.
[5] See note 1 above.

Addressing the educational needs of GHLSs, particularly the second and third generations, is of the utmost importance for the future of Greek as an HL. The future of any HL is at stake unless a systematic approach to the design of educational material is taken. However, a first step in deciding the content and form of more linguistically and culturally responsive educational resources is to identify the socio-linguistic needs of HL learners. Their motivation to learn/maintain their HL as a part of their identity must also be addressed, since designing an HL program without considering their individual proficiencies and needs would compromise the success of the program (Beaudrie, 2016).

The thorough consideration of the language proficiency levels and instructional needs of GHSs, as they appeared in their sociolinguistic profiles presented in this study, suggest that in order to motivate students, the GHL program should include interesting, inspiring, and flexible activities that prioritize systematic contact with Greece. Moreover, utilizing the COVID-19 pandemic experience with remote learning methods gained by all parties involved (students, parents, and educators), e.g. video conferencing with peers and/or teachers from Greece can help prevent the sudden drop in exposure to the HL. The content of such a program ought to depict which tasks HL learners are capable of completing independently and which (formal talks, etc.) could be achieved only with targeted assistance/scaffolding, and should also include teaching of the less developed language skills of reading and writing by using Open Educational Resources (OER) (e.g. online tasks, textbooks, images, quizzes, videos, podcasts, etc.) that address the text types (genres) that Greek HL learners have not been sufficiently exposed to. Content should also enhance cultural and linguistic diversity, foster sustainability through reuse, and address the needs of policymakers and educators (Bijlsma et al., 2016) with OER.

Furthermore, through the program, students need to build HL learning strategies for becoming effective HL leaners while parents need to be trained to effectively motivate their children, particularly the third generation, to stay in touch with Greek culture by relating it to their particular age groups and interests and not relying solely on the traditional elements brought down by previous generations. Finally, since teachers often feel frustrated and inadequate when teaching HL learners, because they essentially use L2 or L1 teaching methods that are not appropriate for this specific target group, they should be offered systematic, additional training in HL teaching and the use of OERs in order to develop the skills and knowledge needed for teaching this population.

Conclusion

The present study has furthered the research on HSs by offering data for a less-studied HL and providing the socio-linguistic profiles and identities of an uninvestigated community of HSs: Greeks living in Chicago. By comparing the results of studies on various HLs with the ones presented here, the authors hope to have revealed some underlying dimensions important to the fate of heritage languages and shed light on questions such as how socio-linguistic profiles and identity issues can help tailor educational programs for HSs. This research-based knowledge can be employed in addressing the academic needs of GHLSs through educational programs. One drawback of the study is that identification with Greek culture, ethnic attachment, and the practice of Greek traditions were investigated only through a structured interview, as such questions were not included in the GHLBQ. Assessing the above-mentioned characteristics using a survey, followed by an interview, would help verify respondents' answers. Finally, further research should investigate the relationship between cultural identity and HL maintenance.

Conflict of interest

The authors declare that they have no conflicts of interest.

Acknowledgements

This study is part of the project Greek Heritage Language Varieties (HEGREEK MIS 5006199). It was held in the frame of the National Strategic Reference Frame (Ε.Σ.Π.Α) and was co-funded by resources of European Social Fund and national resources.

References

Abbot, G. (1909). A study of the Greeks in Chicago. *American Journal of Sociology*, *15*, 379-393.

Aravossitas, T. (2016). *The hidden schools: Mapping Greek heritage language education in Canada* [Unpublished doctoral dissertation]. University of Toronto.

Beardsmore, H. B. (1986). *Bilingualism: Basic principles* (vol. 1). Multilingual matters.

Beaudrie, S. M. (2016). *Building a Heritage Language program*. In M. Fairclough, & S. Beaudrie *Heritage language teaching* (pp. 80-98). Georgetown University Press.

Berard, T. J. (2005). Evaluative categories of action and identity in non-evaluative human studies research: Examples from ethnomethodology. *Qualitative Sociology Review*, *1*(1), 1-25.

Bijlsma, M., Creelman, A., & Titlestad, G. (2016). *Policy recommendations enhancement of cultural and linguistic diversity through OER* [Project Report]. http://langoer.eun.org/c/document_library/get_file?uuid=ef295694-ce7b-477e-b43d-8e921ddb9df4&groupId=395028

Block, D. (2007). *Second language identities*. Continuum.

Campbell, R., & Rosenthal, J. (2000). Heritage languages. In J.W. Rosenthal (Ed.), *Handbook of undergraduate second language education* (pp. 165-184). Lawrence Erlbaum Associates.

Carreira, M. (2009). *Student Questionnaire*. National Heritage Language Resource Center. http://nhlrc.ucla.edu/nhlrc/research#researchproficiency.

Carreira, M. (2016). A general framework and supporting strategies for teaching mixed classes. *Advances in Spanish as a heritage language*, *49*, 159-176. https://doi.org/10.1075/sibil.49.09car

Carreira, M., & Kagan, O. (2011). The results of the national heritage language survey: Implications for teaching, curriculum design, and professional development. *Foreign Language Annals, 44*(1), 40-64. https://doi.org/10.1111/j.1944-9720.2010.01118.x

Chafe, W. L. (1980). The deployment of consciousness in the construction of narrative. In W.L. Chafe (Ed.), *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Praeger.

Chevalier, J. (2004). Heritage language literacy: Theory and practice. *Heritage Language Journal, 2*(1), 1–19.

Cho, G, Cho, K-S., & Tse, L. (1997). Why ethnic minorities want to develop their Heritage language: The case of Korean-Americans. *Language, Culture and Curriculum*, *10*(2), 106-112.

Cho, G., Shin, F., & Krashen, S. (2004). What do we know about heritage languages? What do we need to know about them? *Multicultural Education*, *11*(4), 23-26.

Crawshaw, R., Callen, B., & Tusting, K. (2001). Attesting the self: Narration and identity change during periods of residence abroad. *Language and Intercultural Communication*, *1*(2), 101-119. https://doi.org/10.1080/14708470108668067

De Houwer, A. (2007). Parental language input patterns and children's bilingual use. *Applied Psycholinguistics*, *28*(3), 411-424. https://doi.org/10.1017/S0142716407070221

Dixon, L. Q., Zhao, J., Quiroz, B. G., & Shin, J. Y. (2012). Home and community factors influencing bilingual children's ethnic language vocabulary development. *International Journal of Bilingualism*, *16*(4), 541-565. https://doi.org/10.1177%2F1367006911429527

Fishman, J. A. (1991). *Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages* (vol. 76). Multilingual matters.

García, M. (2003). 2. Recent research on language maintenance. *Annual Review of Applied Linguistics, 23*, 22-43. https://doi.org/10.1017/S0267190503000175

Gardner-Chloros, P. (2009). *Code-switching*. Cambridge University Press.

Gavriilidou, Z., & Mitits, L. (2016). Adaptation of the Strategy Inventory for Language Learning (SILL) for students aged 12-15 into Greek: Developing an adaptation protocol. In M. Mattheoudakis, & K. Nicolaidis (Eds.), *Selected papers from the 21st International Symposium on Theoretical and Applied linguistics* (pp. 588-601). Aristotle University of Thessaloniki.

Gavriilidou, Z., & Mitits, L. (2019). Profiling Greek heritage language speakers in the USA and Russia. *European Journal of Language Studies*, *6*(1), 28-42.

Gignoux, A. (2009). *Survey of demographic and self-identification information for heritage learners of Mexican descent*. http://nhlrc.ucla.edu/data/questionnaires.asp

Guardado, M. (2010). Heritage language development: Preserving a mythic past or envisioning the future of Canadian identity? *Journal of Language, Identity & Education, 9*(5), 329-346. https://doi.org/10.1080/15348458.2010.

Hashimoto, K., & Lee, J. (2011). Heritage-language literacy practices: A case study of three Japanese American families. *Bilingual Research Journal, 34*(2), 161-184. https://doi.org/ 10.1111/j.1749-818X.2007.00022.x

Hoffman, C. (1991). *An introduction to bilingualism*. Longman.
Ilieva, G. N., & Clark-Gareca, B. (2016). Heritage language learner assessment. In M. Fairclough & S.M. Beaudrie (Eds.), *Innovative strategies for heritage language teaching: A practical guide for the classroom* (pp. 214-237). Georgetown University Press.
Jensen, L., & Llosa, L. (2007). Heritage language reading in the university: A survey of students' experiences, strategies, and preferences. *Heritage Language Journal, 5*(1), 98-116.
Kagan, O. (2005). In support of a proficiency-based definition of heritage language learners: The case of Russian. *International journal of bilingual education and bilingualism*, *8*(2-3), 213-221.
Kagan, O., & Dillon, K. (2008). Issues in heritage language learning in the United States. In N. Van Deusen-Scholl, & S. May (Eds.), *Second and foreign language education. Encyclopedia of language and education* (3rd ed.). Springer. https://doi.org/10.1007/978-3-319-02246-8_3
Kagan, O., & Friedman, D. (2003). Using the OPI to place heritage speakers of Russian. *Foreign Language Annals*, *36*(4), 536-545. https://doi.org/10.1111/j.1944-9720.2003.tb02143.x
Karatsareas, P. (2018). Attitudes towards Cypriot Greek and standard modern Greek in London's Greek Cypriot community. *International Journal of Bilingualism*, *22*(4), 412-428. https://doi.org/10.1177%2F1367006918762158
Karatsareas, P. (2019). Cypriot Greek as a heritage and community language in London: (Socio)linguistic aspects of a non-standard variety in a diasporic context. In I. Manolessou, N. Pantelidis, & G. Karla (Eds.), *Volume in honour of Basea-Bezantakou* (pp. 145-170). Kardamitsa
Kasstan, J., Auer, A., & Salmons, J. (2018). Heritage-language speakers: Theoretical and empirical challengers on sociolinguistic attitudes and prestige. *International Journal of Bilingualism, 22*(4), 387-94. https://doi.org/10.1177%2F1367006918762151
Kopan, A. (1989). The Greeks of Chicago: The survival of an ethnic group through education. *Journal of the Hellenic Diaspora*, *16*(1-4), 47-59.
Kunkelman, G. (1990). *The religion of ethnicity: Belief and belonging in a Greek-American community.* Garland.
Lambert, W. (1975). Culture and language as factors in learning and education. In A. Wolfgang (Ed.), *Education of immigrant students*. Ontario Institute for Studies in Education. https://eric.ed.gov/?id=ED096820
Lao, C. (2004). Parents' attitudes toward Chinese-English bilingual education and Chinese-language use. *Bilingual Research Journal, 28*(1), 99-121. https://doi.org/10.1080/15235882.2004.10162614
Lee, J. S. (2002). The Korean language in America: The role of cultural identity in heritage language learning. *Language culture and curriculum*, *15*(2), 117-133. https://doi.org/10.1080/07908310208666638
Lee, J. S., & Kim, H. Y. (2008). Heritage language learners' attitudes, motivations and instructional needs: The case of postsecondary Korean language learners. In K. Kondo-Brown, & J. D. Brown (Eds.), *Teaching Chinese, Japanese, and Korean heritage languages students: Curriculum needs, materials and assessment* (pp. 159–85). Lawrence Erlbaum.
Lee, J. S., & Oxelson, E. (2006). "It's not my job": K–12 teacher attitudes toward students' heritage language maintenance. *Bilingual Research Journal*, *30*(2), 453-477. https://doi.org/10.1080/15235882.2006.10162885
Leeman, J., & Serafini, E. J. (2016). Sociolinguistics for heritage language educators and students. In M. Fairclough & S.M. Beaudrie (Eds.), *Innovative strategies for heritage language teaching: A practical guide for the classroom* (pp. 56-79). Georgetown University Press.
Li, G. (2006). Biliteracy and trilingual practices in the home context: Case studies of Chinese-Canadian children. *Journal of Early Childhood Literacy*, *6*(3), 355-381. https://doi.org/10.1177%2F1468798406069797
Liu, N. (2013). The linguistic vitality of Chinese in the United States. *Heritage Language Journal*, *10*(3), 1-11.
Liao, L. Y., & Larke, P. J. (2008). The voices of thirteen Chinese and Taiwanese parents sharing views about their children attending Chinese heritage schools. *US-China Education Review, 5*(12), 1-8.
Lyutykh, E. (2012). *Survey for Parents in Russian and English*. National Heritage Language Resource Center. http://nhlrc.ucla.edu/data/questionnaires.asp
Martínez, G. (2016). Goals and beyond in heritage language education. In M. Fairclough, & S.M. Beaudrie (Eds.), *Innovative strategies for heritage language teaching: A practical guide for the classroom* (pp. 39-55). Georgetown University Press.
Michopoulos, A. (2008). The Greek-American Education in the dawn of the 21rst Century. In M. Damanakis (Ed.), *Globalization and Greek diaspora* (pp. 31-45). Rethymnon University of Crete.
Montrul, S. (2008). *Incomplete acquisition in bilingualism. Re-examining the age factor.* John Benjamins.
Montrul, S. (2012). *Bilingual background questionnaire for Spanish/English speakers*. National Heritage Language Resource Center. http://www.nhlrc.ucla.edu/data/questionnaires.asp
Montrul, S. (2016). *The acquisition of heritage languages*. Cambridge University Press.

Montrul, S., & Foote, R. (2014). Age of acquisition interactions in bilingual lexical access: A study of the weaker language of L2 learners and heritage speakers. *International Journal of bilingualism*, *18*(3), 274-303. https://doi.org/10.1177%2F1367006912443431

Montrul, S., Bhatt, R., & Benmamoun, A. (2010). Morphological errors in Hindi, Spanish and Arabic heritage speakers. *Studies in Second Language Acquisition*, *33*(2), 155–161. https://doi.org/10.1017/S0272263110000719

Montrul, S., Davidson, J., De La Fuente, I., & Foote, R. (2014). Early language experience facilitates the processing of gender agreement in Spanish heritage speakers. *Bilingualism*, *17*(1), 118. https://doi.org/10.1017/S1366728913000114

Moskos, C. (1989). *Greek Americans: Struggle and Success* (2nd ed.). Transaction.

Nikolidakis, N. (2005). Greek Education in the United States. *Études helléniques/Hellenic Studies*, *13*(2), 79-95.

Park, S. M., & Sarkar, M. (2007). Parents' attitudes toward heritage language maintenance for their children and their efforts to help their children maintain the heritage language: A case study of Korean-Canadian immigrants. *Language, culture and curriculum*, *20*(3), 223-235. https://doi.org/10.2167/lcc337.0

Polinsky, M. (2008a). Gender under incomplete acquisition: Heritage speakers' knowledge of noun categorization. *Heritage language journal*, *6*(1). https://doi.org/10.46538/hlj.6.1.3

Polinsky, M. (2008b). Heritage language narratives. *Heritage language education: A new field emerging* (pp. *149* – 164). Routledge.

Polinsky M. (2011). Annotated bibliography of research in heritage languages. In *Oxford Bibliographies, Linguistics* (pp.1-26). Oxford University Press.

Polinsky, M. (2018). *Heritage languages and their speakers* (vol. 159). Cambridge University Press.

Polinsky, M., & Kagan, O. (2007). Heritage languages: In the 'wild' and in the classroom. *Language and linguistics compass*, *1*(5), 368-395. https://doi.org/10.1111/j.1749-818X.2007.00022.x

Potowski, K., Jegerski, J., & Morgan-Short, K. (2009). The effects of instruction on linguistic development in Spanish heritage language speakers. *Language learning*, *59*(3), 537-579. https://doi.org/10.1111/j.1467-9922.2009.00517.x

Prevoo, M. J., Malda, M., Mesman, J., Emmen, R. A., Yeniad, N., van Ijzendoorn, M. H., & Linting, M. (2014). Predicting ethnic minority children's vocabulary from socioeconomic status, maternal language and home reading input: Different pathways for host and ethnic language. *Journal of child language*, *41*(5), 963-984. https://doi.org/10.1017/s0305000913000299

Salutos, T. (1964). *The Greeks in the United States*. Harvard University Press.

Sánchez-Muñoz, A. (2016). Heritage language healing? Learners' attitudes and damage control in a heritage language classroom. *Advances in Spanish as a heritage language*, *49*, 205-217. https://doi.org/10.1075/sibil.49.11san

Shin, S. J. (2013). Transforming culture and identity: Transnational adoptive families and heritage language learning. *Language, Culture and Curriculum*, *26*(2), 161-178.

Swender, E., Martin, C. L., Rivera-Martinez, M., & Kagan, O. E. (2014). Exploring oral proficiency profiles of heritage speakers of Russian and Spanish. *Foreign Language Annals*, *47*(3), 423-446. https://doi.org/10.1080/07908318.2013.809095

Tamošiūnaitė, A. (2008). The Lithuanian language in the United States: Shift or maintenance. *Lituanus*, *54*(3), 60-78.

Torres, J. (2012). *Language background questionnaire for heritage speakers of Spanish*. National Heritage Language Resource Center. http://nhlrc.ucla.edu/data/questionnaires.asp.

Valdés, G. (1999). Heritage language students: Profiles and possibilities. In J. K. Peyton, D. A. Ranard, & S. McGinnis (Eds.), *Heritage languages in America: Preserving a national resource* (pp. 37-80). Center for Applied Linguistics

Valdés, G. (2000).Introduction. In *AATSP Professional Development Series Handbook for Teachers K–16* (vol. 1: *Spanish for Native Speakers,* pp. 1–20). Harcourt.

Valdés, G. (2001). Heritage language students: Profiles and possibilities. In J. K. Peyton, D. A. Ranard, & S. McGinnis (Eds.), *Heritage languages in America. Preserving a national resource* (pp. 37–80). Center for Applied Linguistics.

Valdés, G. (2003). *Expanding definitions of giftedness: Young interpreters of immigrant background*. Lawrence Erlbaum.

Veltman, C. (1988). *The future of the Spanish language in the United States*. Hispanic policy development project. https://cdn.ymaws.com/www.aatsp.org/resource/resmgr/advocacy/The_future_of_spanish.pdf

He, A. W. (2006). Toward an identity theory of the development of Chinese as a heritage language. *Heritage Language Journal*, *4*(1), 1-28.

Wong-Fillmore, L. W. (2000). Loss of family languages: Should educators be concerned? *Theory into Practice, 39*(4), 203-210.
Yang, J. S. R. (2003). Motivational orientation of Korean learners and ethnic identity development of heritage learners. *The Korean language in America*, *8*, 295-314. https://doi.org/10.1111/j.1944-9720.2003.tb01931.x
Zyzik, E. (2016). Toward a prototype model of the heritage language learner. In M. Fairclough and S. Beaudrie (Eds.), *Innovative approaches to heritage language teaching* (pp. 19-38). Georgetown University Press.

National Research University Higher School of Economics
Journal of Language & Education Volume 7, Issue 1, 2021

Heydarnejad, T., Fatemi, A. H., & Ghonsooly, B. (2021). The Relationship between Critical Thinking, Self-regulation, and Teaching Style Preferences among EFL Teachers: A Path Analysis Approach. *Journal of Language and Education, 7*(1), 96-108. https://doi.org/10.17323/jle.2021.11103

The Relationship between Critical Thinking, Self-regulation, and Teaching Style Preferences among EFL Teachers: A Path Analysis Approach

Tahereh Heydarnejad[1], Azar Hosseini Fatemi[2], Behzad Ghonsooly[2]

[1] University of Gonabad

[2] Ferdowsi University of Mashhad

Correspondence concerning this article should be addressed to Tahereh Heydarnejad, University of Gonabad, Department of English Language, Faculty of Literature and Humanities, University of Gonabad,Khorasan Razavi Province, Gonabad, Iran. Email: t.heydarnejad88@yahoo.com

Teachers are world makers. They can change the thinking, attitudes, and lives of their students. Thus, it is essential to study the factors that foster teachers' competency. Critical thinking, self-regulation, and teaching style are some of the factors influencing the effectiveness of teachers. In line with this argument, the present study delved into the possible impact of critical thinking abilities and self-regulatory strategies of English as a foreign language (EFL) teachers on their teaching style preferences. The possible influence of critical thinking abilities on teacher self-regulation was also studied. To do this, Watson-Glaser's Critical Thinking Appraisal (Form 1), the Teacher Self-Regulation Scale (TSRS), and Grasha's Teaching Style Inventory (TSI) were administered to 320 EFL teachers who were teaching at different private language institutions in Iran. A path analysis was utilized to ponder their causal relationships. The findings indicated that teachers with higher critical thinking abilities and self-regulatory skills tend to implement learner- centered styles (namely, Facilitator and Delegator) while teachers with lower critical thinking abilities and self-regulatory skills tend to do the opposite. Moreover, the significant effects of critical thinking on teacher self-regulation were determined. The implications of this study may open up new perspectives into successful pedagogy for policymakers, curriculum designers, and teachers.

Keywords: critical thinking, self-regulation, teaching styles, EFL teachers, path analysis

Introduction

The era of technology and information uncovers new challenges every day, which calls for expanding novel approaches to achieve efficient teaching and instruction. These changing conditions push educational capabilities to cultivate Critical Thinking (CT) as a higher-order thinking skill. That is, CT ensures better-thinking capacities in EFL teachers and enables them to evaluate and analyze (Zhang, Yuan, & He, 2020) before drawing conclusions or making a judgment. As a result, with better decisions, the future is more secure and successful.

Dewey (1933), who is referred to as the father of modern critical thinking, considered CT to be an active lens that helps to make inferences based on the existing evidence. In other words, CT puts forward a royal road to thinking (Ghanizadeh & Heydarnejad, 2015). That means CT, along with sophisticated thinking skills, takes individuals from the surface layer of meaning into the deeper one involving inferences (Renner, 1996), which is necessary for insightful education. CT also offers a standard of intellectual excellence that is essential for people's academic and social lives (Scriven & Paul, 2004). Therefore, by helping teachers to cultivate CT, the educational system and society are sowing the seeds of a generation with better thinking capabilities. Hence, it is part of the priorities of a proper instruction to provide efficient room for the students to think deeply and share their thoughts with their peers. This goal is not reachable unless the teachers are critical thinkers themselves.

CT can also be defined as a deliberate and self-regulatory judgment about one's actions and thoughts (Giancarlo & Facione, 2001). Self-regulation reflects the utilization of self-assessment and self-awareness techniques

(Zimmerman & Risemberg, 1997) with a profound influence on professional success (Davis & Gray, 2007). Teachers play a pivotal role in effective instruction; to that end, they should monitor and evaluate their ideas, behaviors, and other self- regulatory factors related to educational contexts (Dembo, 2001; Peel, 2020). Self-regulation provides a golden chance for teachers to define objectives, act decisively, and modify their teaching accordingly (Karimnia & Mohammadi, 2019; Yesim, Sungur, & Uzuntiryaki, 2009). Key to this discussion, Paris and Winograd (1999) claimed that self-regulation enables teachers to better adjust to the needs of the students as well as evaluate teaching and learning strategies deeply. Tying together CT and self-regulation, society can take a step toward insightful and practical instruction.

Another crucial aspect of successful instruction is teachers' adopted teaching styles, which represent teachers' attitudes, approaches, and practices in teaching (Cooper, 2001; Tavakoli & Karimnia, 2017). Moreover, Jarvis (2004) stipulated that teaching style specifies teachers' perceptions and attitudes toward all the parts related to teaching and learning. In other words, teaching styles are general principles as well as educational and management strategies for classroom instruction. As individuals learn differently, there are different teaching styles as well.

Successful teachers are those who are aware of students' different learning styles and attempt to implement suitable teaching styles to motivate students and increase their engagement in learning. To achieve these objectives, teachers should be equipped with knowledge and skills to evaluate and modify themselves, and consequently design instruction in the best way. Facilitating knowledge and skills, especially with higher-order thinking skills (Zhang, Yuan, & He, 2020), is not an easy task and sheds light on some avenues for further research. Considering the crucial role of teaching style, this study sought to explore the possible influence of CT and self-regulation on successful styles of teaching.

Background

In the following sections, the relevant literature on critical thinking, self-regulation, and teaching style is briefly reviewed.

Critical Thinking

About two thousand years ago, the concept of critical thinking was put forward by Socrates, who emphasized reasoning, questioning, analyzing, and evaluating as the basis of every judgment. Reviewing the existing literature on CT reveals no agreement upon a unified definition (Fasko, 2003; Fisher, 2001; Halanon, 1995; Minnich, 1990). In Ennis's words (1996), CT is an ongoing evaluation of ideas to make inferences based on the available evidence. Similarly, Ivie (2001) describes critical thinking in terms of reflective practices, which connects early assumptions to justified conclusions in a logical way. An indisputable definition of CT is proposed by The American Philosophical Association, which characterized CT as deliberate and self-regulatory decision making via verification, inquiry, appraisal, and inference (Giancarlo & Facione, 2001).

Although there are no agreed-upon definitions of CT, a considerable body of literature highlights the significant role of CT in all aspects of life, especially in achieving academic objectives (Facione & Facione, 1996). As Dewey (1933) noted, the core aim of education is learning to think; to reach this goal, students need to learn how to use CT strategies in the academic domain (Keeley, Holland, & Watson, 2005). Key to this discussion, Elder and Paul (2003) emphasized the crucial role of CT in converting learners into active questioners. CT is not an inborn ability (Ghanizadeh & Heydarnejad, 2015); thus, it should be created and shaped by teachers as the drivers of education vehicles who are responsible for providing and practicing deep comprehension (Heydarnejad, Ebrahimi, & Najjari, 2018).

The capability to think critically and reflectively is related to a teacher's intellect and experience (Dewey, 1933). In this regard, Jenkins (1998) stressed that CT develops teacher competencies and turns a novice teacher into an autonomous one. According to Reynolds (1992), competent teachers analyze their activities using higher-order thinking skills (CT and reflective thinking). In a similar way, Mayfield (2001) stressed that a critically reflective teacher raises students' awareness of critical thinking by engaging them in critical thinking activities. Thus, the evidence is overwhelming that teachers need to expand their knowledge and skills in CT.

Taking these standpoints in mind, critical thinking skills and their positive effects on educational achievement in EFL settings (among many others) provide a fertile field in which to undertake educational research. For instance, in a recent study by Bobkina and Stefanova (2016), the benefits of developing a model of teaching critical thinking skills in the EFL classroom is investigated. They concluded that critical thinking skills enable learners to create their own reflective learning styles. In another study, Kavanoz and Akbaş (2017) inferred that EFL teachers are capable of practicing critical thinking in their classrooms if they develop an adequate knowledge of critical thinking. Similarly, Zhang et al. (2020) referred to EFL teachers' narrow understanding of CT as well as the gap between EFL teachers' attitudes toward CT and their actual class activities as the reasons for not successfully implementing CT in EFL classes.

Self-regulation

Self-regulation simply outlines self-supporting thoughts, sensations, and behaviors that are premeditated and adjusted to achieve personal objectives (Zimmerman, 2000). In the learning context, self-regulation refers to the active involvement of students in the goal setting, self- monitoring, and self-evaluation of their learning (Pintrich, 2000). In the teaching domain, self- regulation reflects teachers' strategies to regulate and manage their activities (Yesim et al., 2009) and emotions (Heydarnejad, Hosseini Fatemi, & Ghonsooli, 2017b). All in all, theories on student self-regulation were generalized to theories and principles on teacher self-regulation.

Self-regulation and self-control seem to have similar meanings, but they are different. While self- regulation has to do with impeding strong impulses, self-regulation lessens the frequency and strength of strong impulses by controlling stress-load and recovery (Shanker, 2016). To get more specific, self-regulation manages self-control. The notion of teacher self-regulation can also be discussed from personal and social viewpoints (Delfino, Dettori, & Persico, 2010). From a particular viewpoint, teachers armed with self-regularity strategies evaluate and monitor their progress; they adjust their actions to achieve the desired outcome. From the social perspective, self-regulation helps teachers adapt themselves to students' needs and the fast speed of technology. Self-regulation covers a wide range of behaviors from minute-to-minute choices to the larger, more significant decisions that can have a substantial impact on whether people meet their goals.

Key to this discussion, the association between Iranian EFL teachers' self-regulation and self- resiliency was explored in a study by Partovi and Tafazoli (2016). They found that Iranian EFL teachers who adopted more self-regulatory strategies were more resilient. In a similar manner, Monshi Toussi and Ghanizadeh (2012) found a significant relationship between teachers' self- regulation and internal locus of control. They also concluded that EFL teacher self-efficacy had no significant impact on the relationship between self-regulation and locus of control. In another recent study, a positive relationship between teachers' reflective practices and their self-regulation was discovered (Pazhoman & Sarkhosh, 2019).

Overall, leafing through the empirical studies conducted on self-regulation revealed that the majority of studies in this realm were related to the learners (Alvi & Gillies, 2020; Dignath & Büttner, 2018; Peel, 2020). Those that drew attention to teacher self-regulation focused on the self-regulatory practices of teachers as being learners or considered teachers as the responsible ones to learn self-regulatory strategies for the sake of transferring them to their own learners (Dembo, 2001; Peel, 2019; Randi, 2004; Yesim et al., 2009). Thus, teacher self-regulation is still in its infancy and requires a more insightful analysis.

Teaching Style

Individuals are different due to their distinctive style of thinking, preferences, and ways of doing things that influence their behavior (Heydarnejad, Hosseini Fatemi, & Ghonsooli, 2017a). Teachers are not an exception; their attitudes toward teaching and the strategies that they employ are their distinctive teaching style (Grasha, 2002). In other words, teaching style is teachers' personal preferences, knowledge, and beliefs that influence how they present information to their learners and receive feedback (Kaplan & Kies, 1995). In fact, teachers, like students, could employ many teaching and management strategies for classroom instruction, which are due to different personal attitudes, inspirations, ways of thinking, and social background (Vaughn & Baker, 2001). As Grasha (1996) put it, these styles are like colors on an artist's palette, which converge into various clusters and make up the distinctive ways teachers design instructional settings.

Studies on teaching styles emerged in the 1970s (Biggs, 2001). An overwhelming number of classifications can be found in the literature that portrays different teaching styles. The following are some of these classifications:

Direct style-Indirect style (didactic-student-centered) by Flanders (1970), Formal style-Informal style by Bennett, Jordan, Long, and Wade (1976), Open style- Traditional style by Solomon and Kendall (1979), and Intellectual Excitement style-Interpersonal Rapport style by Lowman (1995). In the current research, Grasha's model (1994, p.143) was applied due to its integrity and comprehensiveness. This model identifies styles of teaching as:

1. Expert: A teacher with expert style strives to maintain status as an expert among learners by presenting detailed knowledge.
2. Formal Authority: A teacher who teaches by formal authority style supervises learners critically and generally concentrates on the content. This style of teaching does not emphasize the student's participation, needs, and affective factors.
3. Personal Model: A teacher who implements personal model style expects the learners to copy his/her attitudes and approaches as the central model in the class.
4. Facilitator: A facilitator teacher focuses on self-learning and self-discovery in the class. Independent and responsible learners are the product of facilitator style of teaching.
5. Delegator style: A delegator teacher enhances autonomy and confidence in the learners. He/ She emphasizes group work and collaboration. Collaboration and peer to peer learning are the dominant goals in this style.

Expert, Formal Authority, and Personal Model styles are considered to be teacher-centered styles; whereas, Facilitator and Delegator styles correspond to student-centered styles. As Grasha (1994) put it, this classification originated from the traditional binary between teacher-centered and student-centered styles of teaching. The teacher-centered approach originated from the behaviorist theory, which claimed that changes in behavior are due to external stimuli (Skinner, 1974). Each teacher may adopt all these styles but to a varying extent. To ensure effective instruction, the teacher, as the center of the education, should challenge themselves to tailor their tutoring to different learners and in various educational contexts. Furthermore, educational and cultural backgrounds as well as teaching and learning experiences are influential in teaching style adaptation (Faruji, 2012; Nouraey & Karimnia, 2016; Tavakoli & Karimnia, 2017).

A recent study dealt with the effects of teachers' gender, teaching experience, and brain dominance on their teaching style in the EFL context of Iran (Karimnia & Mohammadi, 2019). Their findings implied that a teacher's gender makes a difference in the formality of their teaching style. Regarding the impact of teaching experience on teaching style, a significant difference between teachers with different teaching experience was concluded. However, no significant finding was reported with regard to the effects of brain dominance on teaching style.

The above-mentioned studies highlight the contributing effect of CT, self-regulation, and teaching style on teachers' practices and instructional behaviors; however, to the researchers' best knowledge no study has ever considered these variables in a single study. Moreover, few researchers have ever attended to some of these goals in an Iranian context. Henceforth, the current study is a response to the call for better understanding the association of EFL teachers' critical thinking with self- regulation and teaching style. Therefore, the following research questions are explored in the current investigation:

1. Does EFL teachers' critical thinking have any impact on their self-regulation?
2. Does EFL teachers' critical thinking have any impact on their teaching style preferences?
3. Does EFL teacher self-regulation have any impact on their teaching style preferences? Based on the research questions, the following null hypotheses could be formulated:
   HO1. EFL teachers' critical thinking has no impact on their self-regulation.
   HO2. EFL teachers' critical thinking has no impact on their teaching style preferences.
   HO3. EFL teacher self-regulation has no impact on their teaching style preferences.

Materials and Methods

In this section, the methodological steps that were taken in conducting this study are introduced. First, it introduces the participants and a detailed account of the instrument used for the research purposes. Then, the procedure of the current study, which includes some information about data collection and data analysis, is presented.

Participants

The participants of the present study were 320 EFL teachers teaching at intermediate to upper- intermediate levels at private language institutions in different cities of Iran. They were selected based on convenience sampling. Also, to ensure generalizability, it was attempted to include participants from both genders with varying groups of age and teaching experience. There were 123 male and 185 female teachers (twelve teachers did not specify their gender), 22 to 48 years old, and with 1-26 years of teaching experience. Most of them held an MA degree or a BA degree in different branches of English studies: English literature (98), English teaching (123), English translation (48), and 25 teachers were Ph.D. candidates in English teaching. Moreover, those teachers who had majored in other fields of study had the necessary qualifications to teach English.

Instruments

A battery of three questionnaires was utilized in this study as follows:

- Watson–Glaser Critical Thinking Appraisal - Form A
- Teacher Self-Regulation Scale (TSRS)
- Grasha's Teaching Style Inventory (TSI)

Watson–Glaser Critical Thinking Appraisal-Form A

To juxtapose CT, the Watson-Glaser Critical Thinking Appraisal - Form A (1980) was employed. It includes five sections with 80 items, measuring the five aspects of CT based on Watson and Glaser's definition (1980): inference, recognizing of assumptions, making deductions, interpretation, and evaluation. Along with the face, content, construct, and criterion validity of this inventory, which were satisfactory, the test-retest reliability was reported (r=.81) by Watson and Glaser (2002). In the current research, the reliability of Watson–Glaser Critical Thinking Appraisal (Form A) estimated via Cronbach's alpha was .871.

Teacher Self-Regulation Scale (TSRS)

The Teacher Self-Regulation Scale (TSRS) developed and validated by Yesim et al. (2009), was utilized to measure teacher self-regulation. The formation of this scale is according to the self- regulation model proposed by Zimmerman (2000), along with semi-structured interviews with pre- service and in-service teachers. It comprises 40 items and nine subscales in the following sequence: goal setting, intrinsic interest, performance goal orientation, mastery goal orientation, self- instruction, emotional control, self-evaluation, self-reaction, and help-seeking. This inventory is on a six-point Likert-type scale, ranging from 'strongly disagree' to 'strongly agree'. Moreover, on this scale, there is one item as a filter, added to the 40 items, and it was not assessed in the following analyses. In the current study, the estimated reliability of the TSRS was .883, which indicated acceptable reliability.

Grasha's Teaching Style Inventory (TSI)

For measuring teaching style preferences, Grasha's Teaching Style Inventory (1996) was employed. It contained 40 items and was answered on a seven-point Likert-type scale. This inventory consists of five components as follows: expert, formal authority, personal model, facilitator, and delegator teaching style. The present study reported acceptable reliability for all sub-scales of the TSI ranged from .869 to .896.

Procedure

This study was carried out in different Iranian private language institutions, over two semesters from January to December 2019. The EFL teachers received the Watson–Glaser Critical Thinking Appraisal (Form A), Teacher Self-regulation Scale (TSRS), and Grasha's Teaching Style Inventory (TSI) in English versions because all of them were experienced English users. The questionnaires were administered either in a paper-and-pencil format or through a web-based platform.

Data Analysis

The reliability of the questionnaires was checked by the use of Cronbach's alpha formula. The Kolmogorov-Smirnov Test was also used to test the normality distribution of the scores. Then, for the purpose of describing

the collected data, descriptive statistics were applied. Finally, due to the normality of the data, Lisrel 8.8 was employed to analyze the data.

Results

The descriptive statistics of EFL teachers' teaching styles, self-regulation, and critical thinking are presented in the following table.

Table1
Descriptive statistics

Inventory	N	Mean	Std. Deviation	Minimum	Maximum
expert	320	5.049	1.032	2.000	7
formal.authority	320	4.935	1.055	1.875	7
personal.model	320	5.010	1.022	2.000	7
facilitator	320	5.464	1.035	2.125	7
delegator	320	5.342	1.017	2.125	7
TS	320	5.160	.935	2.150	6.900
TSR	320	4.729	.809	2.225	5.900
CT	320	57.406	16.687	0	80.000

As the results reported in Table 1 indicated, among teaching style subscales, Facilitator (M=5.464, SD=1.035) and Delegator (M=5.342, SD=1.017) got the highest mean scores. The other subscales got lower mean scores as follows: Expert (M=5.049, SD=1.032), Personal Model (M=5.010, SD=1.022), and Formal Authority (M=4.935, SD=1.055). Table 1 also displays the following mean scores for teacher self-regulation and critical thinking respectively: M=4.729, M=57.406.

To employ statistical methods and make logical inferences based on the researcher's hypotheses, the most significant step is to choose an appropriate statistical method for the research study. To this aim, data distribution must be taken into consideration. To do so, the Kolmogorov-Smirnov Test was employed to check normality distributions of the variables.

Table 2
The results of K–S Test

Inventory	Statistic	Sig
expert	.158	.200
formal.authority	.204	.095
personal.model	.191	.148
facilitator	.231	.052
delegator	.189	.156
TS	.215	.063
TSR CT	.202 .228	.102 .055

According to Table 2, the sig value for all the scales and their subcomponents was higher than .05, which suggests a normal distribution of the data. Thus, it can be implied that parametric methods could be employed for testing the related research hypotheses.

To this end, the LISREL 8.80 statistical package was used to test the structural relations in the proposed model. The fit indices examined to evaluate the model fit are as follows: the chi-square magnitude, the root mean square error of approximation (RMSEA), the comparative fit index (CFI), and the normed fit index (NFI). The acceptable criteria for fit indices are presented in Table 3.

Table 3
Fit indices

Model	1	2
Chi-square x2	421.67	3756.5
Chi-square/df ratio	2.83	2.75
df	149	1366
RMSEA	.076	.074
CFI	.94	.95
NFI	.94	.94

Figure 1
Schematic representation of path coefficient values for the relationships among critical thinking, teacher self-regulation, and teaching styles (Model 1)

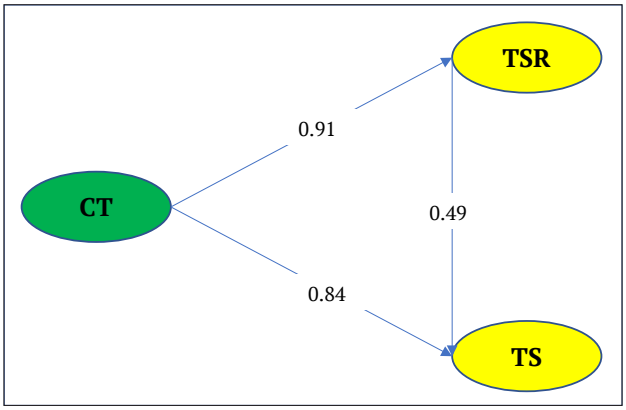
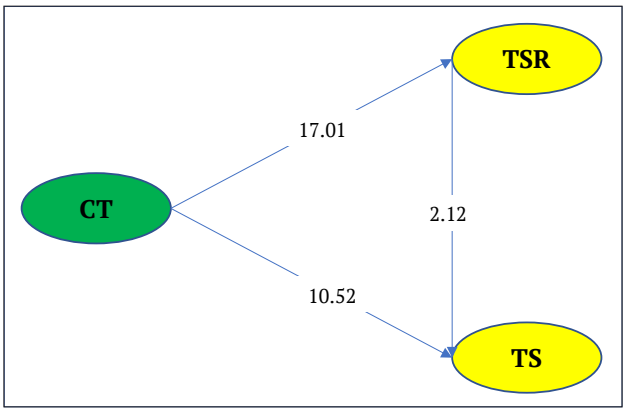


Figure2
T values for path coefficient significance (Model 1)



Based on Figure 1 and Figure 2 (Model 2), CT influences teacher self-regulation as well as teaching style, considering the *t* value, which was greater than 1.96. In addition, Figure 1 and Figure 2 presented the influential role of CT in teaching style. That is to say, CT significantly and positively influences teacher self-regulation (β =.91, t = 17.01) and teaching style (β =.84, t = 10.52). The same is true for the influential role of teacher self-regulation in teaching style (β =.49, t = 2.12). This schematic representation is shown in Figures 3 and 4 in a more detailed way.

Figures 3 and 4 (Model 2) illustrate the significant role of CT and teacher self-regulation on teaching subcomponents. According to these two figures, CT significantly and negatively affects Expert (β = -.26, t = -2.48), Formal Authority (β = -.31, t = -4.59), and Personal Model (β = -.22, t = 2.30). By contrast, it significantly and positively influences Facilitator (β =.68, t = 8.04) and Delegator (β = .64, t = 8.62). It is also worth highlighting that teacher self-regulation significantly and negatively influences Expert (β = -.57, t = -5.42), Formal Authority

(β = -.53, t = -5.55), and Personal Model (β = -.30, t = -4.52). By contrast, it significantly and positively influences Facilitator (β =.77, t = 9.74) and Delegator (β = .64, t = 8.45).

Figure 3

Schematic representation of path coefficient values for the influential role of critical thinking and teacher self-regulation on teaching style' subscales (Model 2)

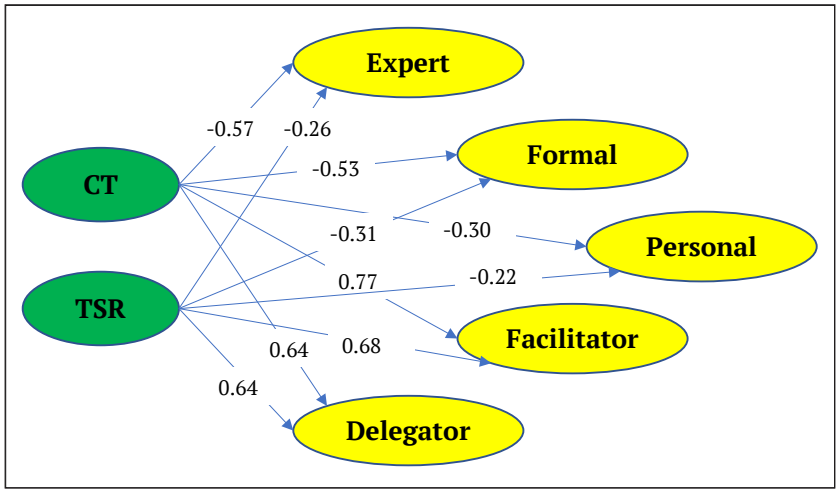
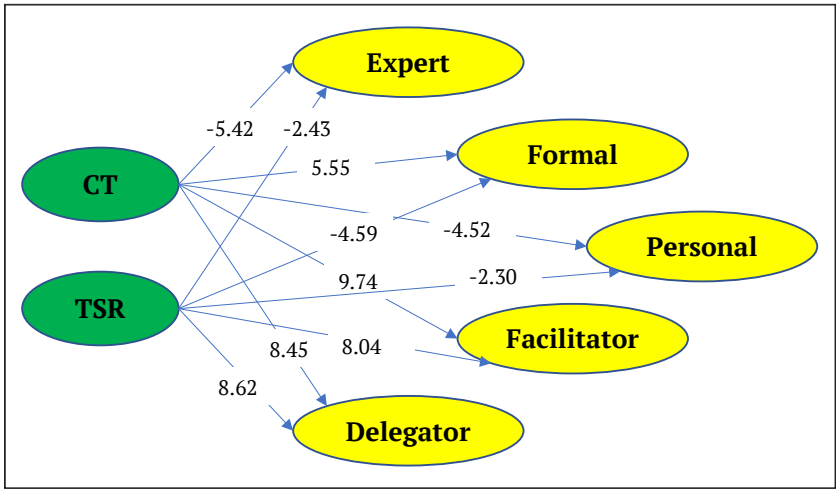


Figure4

T values for path coefficient significance (Model 2)



To determine the relationships among critical thinking, teacher self-regulation, and teaching style preferences, a Pearson product-moment correlation was run.

Table 4
The correlation coefficients among critical thinking, teacher self-regulation, and teaching style

	TS	TSR	CT	expert	formal	personal.model	facilitator	delegator
TS	1							
TSR	.486*	1						
CT	.416*	.457*	1					
expert	-.701*	-.401*	-.410*	1				
formal.authority	-.566*	-.386*	-.328*	.415*	1			
personal.model	-.652*	-.351*	-.436*	.374*	.461*	1		
facilitator	.736*	.420*	.494*	-.438*	-.523*	-.415*	1	
delegator	.713*	.411*	.479*	-.359*	-.429*	-.365*	.380*	1

Note.* Correlation is significant at the 0.05 level (2-tailed).

As Table 4 presents, the relationships between critical thinking and teaching style subscales are as follows: correlated positively with Facilitator (r=.494, p<.05) and Delegator (r=.479, p<.05), but negatively with Personal Model (r=-.436, p<.05), Expert (r=-.410, p<.05), and Formal Authority (r=-.328, p<.05). Furthermore, teacher self-regulation correlated negatively with Expert (r=-.401, p<.05), Formal Authority (r=-.386, p<.05), and Personal Model (r=-.351, p<.05). On the other hand, between self-regulation and Facilitator style of teaching (r=.420, p<.05), and Delegator (r=.411, p<.05), there are positive correlations.

Discussion

As mentioned earlier, the current study intended to shed some light on the possible influences of CT on teacher self-regulation and teaching style, as well as on the effects of teacher self-regulation on teaching style. As the results indicated, CT influences student-centered styles positively and significantly, while the impact of CT on teacher-centered styles is significantly negative. That is to say that the capability to reason, assess, and draw informed conclusions (the skills within the critical thinking realm) help teachers to apply a suitable style of teaching in their classes. Through the lens of CT, teachers are able to find the pros and cons of all procedures and approaches that they employ.

Actually, in the teacher-centered pedagogy, learners are passive and there is no margin for them to develop higher-order thinking skills and critical thinking abilities. This may be due to the fact that the teachers themselves do not reflect and analyze their behavior and actions; they are not critical thinkers or ignore the role of thinking critically. They just put up a set of rigid rules for students in the class to follow; also, group working, collaboration, learners' needs, individual differences, and affective factors are neglected in such classes. The teacher, as the authority in the class, prefers to give a lecture and present his/her knowledge and expertise to the students. In contrast, in the learner-centered pedagogy, raising autonomous and responsible students is a primary goal, which cannot be reached unless the learners have learned and practiced to think critically. Peer to peer interaction and cooperative learning comes to the fore (Condelli, Wrigley, & Yoon, 2009), and indirect instruction, as well as productive learning, is possible via facilitator and delegator styles of teaching (Serin, 2018).

The findings of the current investigation confirm the findings of Zhang et al. (2020), who explored the interplay among critical thinking, teaching, and learning styles, and concluded that CT offers a golden opportunity for both teachers and learners to modify their teaching and learning styles. Moreover, the results of the present investigation are in accord with a recent qualitative study (Elias, Hamzah, & Razak, 2019) indicating that higher-order thinking skills are vital tools that help learners become creative and reflective in a learner-centered context.

The other objective of this study sought to investigate the possible influence of CT on teacher self- regulation. As the findings displayed, CT has a significant positive effect on teacher self- regulation; therefore, the second null hypothesis was rejected. This means that teachers who are empowered with high-level CT skills nurture more self-regulatory strategies in their professional practice. In other words, the ability to think critically raises teachers' awareness and mindfulness in an ever-changing world. This outcome is in accord with

Phan (2010), who presented a theoretical overview of CT and self-regulation in the educational context and inferred there was a close interrelationship between these two concepts. By the same token, Ghanizadeh (2011) confirmed the intertwined relationships between CT and self-regulation. Similarly, Gurcay and Ferah (2018) investigated the relationship between students' critical thinking and their metacognitive self- regulation. Their results indicated that CT correlated positively and significantly with metacognitive self- regulation. In the same vein, Vardi (2015) emphasizes the crucial role of critical thinking on the success of learners with the help of self-regulatory skills. The statistical findings of the present study also support Pintrich's view (2000) that the higher the teachers' involvement in thinking critically, the more they become capable administrators of self-regulatory skills in their actions.

The next objective of this study was to explore the possible effects of teacher self-regulation on the styles teachers adopt in their classes. The findings revealed a significantly influential role of teacher self-regulation on teaching styles; thus, the third null hypothesis was rejected. This finding seems plausible given that self-regulation acts as a magnifying glass trigger for self- awareness, self-reflection, self-monitoring, self-control, and self-evaluation. This self-aid device helps teachers monitor their actions and modify them, when necessary. Thus, they do not consider themselves as the formal authority and expert model in the class, but as a facilitator and delegator to reinforce productive and collaborative learning.

In a nutshell, a cursory look at the existing research concerning critical thinking (Ennis, 1996; Phan, 2010) and theoretical underpinnings on self-regulation (Pintrich, 1999, 2000; Schunk & Zimmerman, 1994; Zimmerman, Bonner, & Kovach, 1996), shows that it can be concluded that the two constructs are two sides of the same coin. Therefore, both of them are necessary for successful and effective pedagogy.

Conclusion

On the whole, this study was an attempt to contribute to the sparse knowledge on teaching style by investigating the relationship between critical thinking, self-regulation, and teaching style preferences among EFL teachers. The findings of the current study illuminated the contribution of CT and teacher self-regulation to the cultivation of learner-centered pedagogy, especially in the EFL teaching context. Furthermore, the benefit of CT for the enhancement of self-regulatory activities was highlighted. The implication is that EFL teachers armed with self-regulatory strategies and critical thinking skills prefer student-centered styles of teaching that enhance the quality of instruction. That is, being empowered by CT and self-regulatory strategies, teachers are capable of breaking the barrier of considering themselves as the dominant authority in the class without any room for student engagement and collaboration. Thus, employing this implication in the pre-service and in-service educational programs may lay the foundations for effective teaching and, consequently, effective learning. This research also hints at some useful implications for policymakers, teachers, and researchers.

Further studies are required to adopt mixed-method approaches and to triangulate data by means of employing other instruments such as class observations and focused interviews. More studies are also suggested to investigate the direction and the strength of the relations between the variables and their subcomponents via structural equation modelling (SEM).

Conflict of interests

The authors declare that they have no conflict of interest.

References

Alvi, E., & Gillies, R.M. (2020). Promoting self-regulated learning through experiential learning in the early years of school: A qualitative case study. *European Journal of Teacher Education, 44*(2),135-157. https://doi.org/10.1080/02619768.2020.1728739.

Bennett, N., Jordan, J., Long, G., & Wade. B. (1976). *Teaching Styles and pupil progress*. Open Books Publishing Limited.

Biggs, J. B. (2001). Enhancing learning: A matter of style or approach? *Perspectives on thinking, learning, and cognitive styles* (pp. 73–102). Lawrence Erlbaum Associates.

Bobkina, J., & Stefanova, S. (2016). Literature and critical literacy pedagogy in the EFL classroom: Towards a model of teaching critical thinking skills. *Studies in Second Language Learning and Teaching*, *6*(4), 677-696. https://doi.org/ 10.14746/ssllt.2016.6.4.6

Condelli, L., Wrigley, H., & Yoon, K. S. (2009). "What works" for adult literacy students of English as a second language. In S. Reder, & J. Bynner (Eds.), *Tracking adult literacy and numeracy skills: Findingsfrom longitudinal research* (pp. 132-159). Routledge.

Cooper, T. C. (2001). Foreign language style and personality. *Foreign Language Annals*, *34*, 301- 316. https://doi.org/10.1111/j.1944-9720. 2001.tb02062.x

Davis, S. G., & Gray, E. S. (2007). Going beyond test-taking strategies: building self-regulated students and teachers. *Journal of Curriculum and Instruction*, *1*(1), 31-47. http://www.joci.ecu.edu/

Delfino, M., Dettori, G., & Persico, D. (2010). An online course fostering self-regulation of trainee teachers. *Psicothema, 22*(2), 299-305. https://doi.org/10.3776/joci.2007.v1n1p31-47

Dembo, M. H. (2001). Learning to teach is not enough - Future teachers also need to learn howto learn. *Teacher Education Quarterly, 28*(4), 23-35.

Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educational process*. Lexington.

Dignath, C. & Büttner, G. (2018). Teachers' direct and indirect promotion of self- regulatedlearning in primary and secondary school mathematics classes – insights from video-based classroom observations and teacher interviews. *Metacognition and Learning, 13*(2), 127-157. https://doi.org/10.1007/s11409-018-9181-x

Elder, L., & Paul, R. (2003). Critical thinking: Teaching students how to study and learn. *Journal of Developmental Education, 27*(1), 36-37.

Elias, F., Hamzah, M. I., & Razak, K. A. (2019). Elements of thinking in teaching student-centered learning on faith: A case study. *International Journal of Academic Research in Progressive Education and Development, 8(3),* 126– 132. http://dx.doi.org/10.6007/IJARPED/v8- i3/6297

Ennis, R. (1996). Critical thinking dispositions: Their nature and assessability. *Informal Logic*, *18*(2–3), 165–182.

Facione, N. C., & Facione, P. A. (1996). Externalizing the critical thinking in knowledge development and critical judgment. *Nursing Outlook, 44*(3), 129-136. http://dx.doi.org/10.1016/S0029-6554(06)80005-9

Fasko, D. (2003). Critical thinking: Origins, historical development, future direction. In D. Fasko (Ed.), *Critical thinking and reasoning: Current research, theory, and practice* (pp.3-20). Hampton Press.

Faruji, L. F. (2012). Teachers' teaching styles at English language institutes in Iran. *International Journal of Social Sciences and Education, 2*(1), 364-373.

Fisher,A.(2001).*Criticalthinking: Anintroduction*. Cambridge University Press.

Flanders, A. (1970). *Analyzing teacher behavior*. Addition-Wesley Publishing Company.

Ghanizadeh, A. (2011). An investigation into the relationship between self-regulation and critical thinking among Iranian EFL teachers. *The Journal of Technology & Education, 5*(3), 213– 221.

Ghanizadeh, A., & Heydarnejad, T. (2015). A cross-contextual analysis of Iranian EFL teachers' attitudes and perceptions of critical thinking. *International Journal of Research Studies in Education, 4*(4), 27-38. http://dx.doi.org/10.5861/ijrse.2015.1138

Giancarlo, C. A., & Facione, P. A. (2001). A look across four years at the disposition toward critical thinking among undergraduate students. *Journal of general education*, *50*(1), 29- 55. http://dx.doi.org/10.1353/jge.2001.0004

Grasha, A. F. (1994). A matter of style: The teacher as expert, formal authority, personal model, facilitator, and delegator. *College Teaching, 42*(4), 142–49.

Grasha, A. F. (1996). Teaching with style: A practical guide to enhancing leaning by understanding teaching and learning style. *Alliance Publishers.*

Grasha, A. F. (2002). The dynamics of one-on-one teaching. *College Teaching*, *50*(4), 139-146. http://dx.doi.org/10.1080/87567550209595895

Gurcay, D., & Ferah, H. O. (2018). High school students' critical thinking related to their metacognitive self-regulation and physics self-efficacy beliefs. *Journal of Education and Training Studies, 6*(4), 125-130. https://doi.org/10.11114/jets.v6i4.2980

Halonen, J. S. (1995). Demystifying critical thinking. *Teaching of Psychology, 22,* 75–81. https://doi.org/10.1207/s15328023top2201_23

Heydarnejad, T., Hosseini Fatemi, A., & Ghonsooly, B. (2017a). An exploration of EFL teachers' teaching styles and emotions. *Journal of Applied Linguistics and Language Research, 4*(2), 26-46. http://www.jallr.com

Heydarnejad, T., Hosseini Fatemi, A., & Ghonsooly, B. (2017b). Emotions and self-regulation: A case of Iranian EFL high school and private language institute teachers. *International Journal of Educational Investigations, 4*(4), 82-100.

Heydarnejad, T., Ebrahimi, M. R., & Najjari, H. (2018). On the associations among critical thinking, reflective thinking, and emotions: A case of Iranian EFL Teachers. *International Journal of Applied Linguistics & English Literature, 7*(6), 97-103. http://dx.doi.org/10.7575/aiac.ijalel.v.7n.6p.97

Ivie, S. D. (2001). Metaphor: A model for teaching critical thinking. *Contemporary Education, 72(*1), 18-23.

Jarvis, P. (2004). *Adult education and lifelong learning: Theory and practic*e. Routledge Falmer.

Jenkins, E. K. (1998). The significant role of critical thinking in predicting auditing students' performance. *Journal of Education for Business*, *73*(5), 274 -279.

Karimnia, A., & Mohammadi, N. (2019). The effects of teachers' gender, teaching experience, and brain dominance on their teaching styles. *International Journal of Research in English Education, 4*(1), 37-46. http://dx.doi.org/10.29252/IJREE.4.1.37

Kaplan, J., & Kies, A. (1995). Teaching styles and learning styles: Which came first? *Journal of Instructional Psychology, 22*, 29–34.

Kavanoz, S., & Akbaş S. (2017). EFL teachers' conceptualizations and instructional practices of critical thinking. *International Online Journal of Education and Teaching, 4*(4), 418-433. http://iojet.org/index.php/IOJET/article/view/251/180

Keeley, B. T., Holland, J., & Watson, M. (2005). Preliminary evidence on the association between critical thinking and performance in principles of accounting. *Issues in Accounting Education*, *20*(1), 33-49. http://dx.doi.org/10.2308/iace.2005.20.1.33

Lowman, J. (1995). *Mastering the techniques of teaching* (2nd ed). San Francisco.

Mayfield, M. (2001). Thinking for yourself: Developing critical thinking skills through reading and writing (5th ed.). Harcourt College.

Minnich, E. (1990). *Transforming knowledge.* Temple University Press.

Monshi Toussi, M. T., & Ghanizadeh, A. (2012). A study of EFL teachers' locus of control and self-regulation and the moderating role of self-efficacy. *Theory and Practice in Language Studies, 2*(11), 2363–2371. http://dx.doi.org/10.4304/tpls.2.11.2363-2371

Nouraey, P., & Karimnia, A. (2016). The impact of rhythm and gender on spelling output of Iranian EFL learners: A cross-sectional study. *Khazar Journal of Humanities and Social Sciences, 19*(2), 53-63. http://dx.doi.org/10.5782/2223-2621.2016.19.2.53

Paris, S. G., & Winograd, P. (1999). The role of self-regulatedlearning in contextual teaching: Principles and practices forteacher preparation. In *Contextual teaching andlearning: Preparing teachers to enhance student success inthe workplace and beyond* (pp. 219–253). ERIC Clearinghouse on Adult, Career, and Vocational Education/ERIC Clearinghouse on Teaching andTeacher Education. https://files.eric.ed.gov/fulltext/ED427263.pdf

Partovi, N., & Tafazoli, D. (2016). On the relationship between Iranian EFL teachers' self-regulation, self-resiliency, gender and teaching experience. *Educational Research, 4*(1), 7- 25.

Pazhoman, H., & Sarkhosh, M. (2019). The Relationship between Iranian English high school teachers' reflective practices, their self-regulation and teaching experience. *International Journal of Instruction*, *12*(1), 995-1010. https://doi.org/10.29333/iji.2019.12164a

Peel, K. L. (2019). The fundamentals for self-regulated learning: A framework to guide analysis and reflection. *Educational Practice and Theory, 41*(1), 23-49. https://doi.org/10.7459/ept/41.1.03

Peel, K. L. (2020). Everyday classroom teaching practices for self-regulated learning. *Issues in Educational Research, 30*(1), 260–282. http://www.iier.org.au/iier30/2020conts.html

Phan, H. P. (2010). Critical thinking as a self-regulatory process component in teaching and learning. *Psicothema, 22*(2), 284-292.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P.R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451-502). Academic.

Randi, J. (2004). Teachers as self-regulated learners. *Teacher' College Record,106*(9), 1825-1853. https://doi.org/10.1111/j.1467-9620.2004.00407.x

Renner, C. E. (1996). *Enrich learners' language production through content-based instruction.* Paper presentedat a National Conference on Lingua e Nuova Didattica, Modena, Italy. (ERIC Document Reproduction Service No. ED 411 694).

Reynolds, A. (1992). What Is Competent Beginning Teaching? A Review of the Literature. *Review of Educational Research, 62*, 1-35. https://doi.org/10.3102/00346543062001001

Safari, E., & Hejazi, M. (2017). Learning styles and self-regulation: An associational study on high school students in Iran. *Mediterranean Journal of Social Sciences*, *8*(1), 463-469. https://doi.org/10.5901/mjss.2017.v8n1p463

Schunk, D. H, & Zimmerman, B. J. (Eds.). (1994). *Self-regulation of learning and performance: Issues and educational applications*. Lawrence Erlbaum Associates.

Scriven, M., & Paul, R. (2004). *The critical thinking community*. http://www.criticalthinking.org/aboutCT/definingCT.shtml

Serin, H. (2018). A Comparison of teacher-centered and student-centered approaches in educational settings. *International Journal of Social Sciences & Educational Studies, 5*(1)*,* 164-167. https://doi.org/10.23918/ijsses.v5i1p164

Shanker, S. (2016). Self-Reg: How to help your child (and you) break the stress cycle and successfully engage with life. *Penguin Random House*.

Skinner, B. F. (1974). *About behavioralism*. Random House.

Solomon, D., & Kendall, A. J. (1979). *Children in classrooms: An investigation of person environment interaction*. Prager Publishers.

Tavakoli, M., & Karimnia, A. (2017). Dominant and gender-specific tendencies in the use of discourse markers: Insights from EFL learners. *World Journal of English Language, 7*(2), 1-9. https://doi.org/10.5430/wjel.v7n2p1

Vardi, I. (2015). *The relationship between self-regulation, personal epistemology, and becoming a critical thinker: Implications for pedagogy*. In M. Davies & R. Barnett (Eds.), The Palgrave handbook of critical thinking in higher education (pp. 193–197). Palgrave Macmillan.

Vaughn, L., & Baker, R. (2001). Teaching in the medical setting: Balancing teaching styles learning styles and teaching methods. *Medical Teacher*, *23*(6), 610– 612. https://doi.org/10.1080/01421590120091000

Watson, G. B., & Glaser, E. M. (1980). *WGCTA Watson-Glaser Critical Thinking Appraisal Manual: Forms A and B*. The Psychological Corporation.

Watson G. B., & Glaser, E. M. (2002). *Watson-Glaser Critical Thinking Appraisal*. Psychological Corporation.

Yesim, C. A., Sungur, S., & Uzuntiryaki, E. (2009). Teacher self-regulation: Examining a multidimensional construct. *Educational Psychology*, *29*(3), 345-356. https://doi.org/10.1080/01443410902927825

Zhang, H., Yuan, R., & He, X. (2020). Investigating university EFL teachers' perceptions of critical thinking and its teaching: Voices from China. *Asia-Pacific Education Researcher*, *29*(5*)*, 483–493. https://doi.org/10.1007/s40299-020-00500-6

Zimmerman, B. J., & Risemberg, R. (1997). Self-regulatory dimensions of academic learning and motivation. In G.D. Phye (Ed.), Handbook of academic learning (pp. 105-125). Academic Press. http://dx.doi.org/10.1016/B978-012554255-5/50005-3

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P.R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp.13-39). Academic Press. https://doi.org/10.1016/B978-012109890-2/50031-7

Zimmerman, B. J., Bonner, S., & Kovach, R. (1996). *Developing self-regulated learners: Beyond achievement to self-efficacy*. American Psychological Association. https://doi.org/10.1037/10213-000

National Research University Higher School of Economics
Journal of Language & Education Volume 7, Issue 1, 2021

Isik, A. (2021). Exploring How ELT Teachers Perceive and Practice English Language Assessment. *Journal of Language and Education, 7*(1), 109-126. https://doi.org/10.17323/jle.2021.10296

Exploring How ELT Teachers Perceive and Practice English Language Assessment

Ali Işık
Istinye University

Correspondence concerning this article should be addressed to Ali Işık, Istinye University, Department of English Language and Literature, Maltepe Mahallesi, Edirne Cirpici Yolu No:9 Zeytinburnu, Istanbul, Turkey. E-mail: isikal@hotmail.com

As a well-designed language education program naturally requires a well-designed assessment component, the pivotal role of assessment in language education needs to be stressed. This study focuses on how English language teaching (ELT) teachers receive training in English language assessment, and how they perceive and practice assessment in Turkey. The study was conducted with the participation of 198 ELT teachers from 24 K-12 level schools and eight universities. A mixed-methods research design was chosen and the data were collected through a questionnaire, follow-up interviews, observations, personal conversations, and sample exam evaluations. The findings indicated that the assessment practices of the teachers were shaped by the teachers' language learning and teaching experiences, their intuition, adherence to assessment traditions, and the emulation of what other teachers conducted to conform to group norms. It was also observed that as the teachers did not receive proper pre-service and in-service assessment training, their assessment knowledge was low. Moreover, it was found that the assessment component of teacher training programs remained peripheral and did not help equip teachers with assessment-related theoretical knowledge and practical skills. Finally, the assessment quality in these schools was found to be low and assessment was taken as a formal requirement to grade students. In the final part of the paper, some suggestions for effective assessment are put forward.

Keywords: Assessment, testing, teacher training, English teaching, English language teachers

Introduction

The demand for efficient language assessment, one of the widespread issues in language education, is on the rise. (Baker, 2016: Fulcher, 2012; Tsagari, 2016). The power of assessment on education programs has started to receive considerable attention and it is widely accepted that assessment affects teaching/learning directly and determines its quality (Menken, Hudson, & Leung, 2014; Tsagari & Banerjee, 2016; Tsagari & Vogt, 2017). Hence, the quest for valid, fair, accountable, and carefully designed assessment compatible with language teaching methodology has gained momentum (Coombe, Vafadar, & Mohebbi, 2020; Levi & Inbar-Lourie, 2020; Malone, 2013). Consequently, language assessment literacy (LAL), what language teachers should know, and how they should apply their conceptual knowledge in assessment have been studied extensively (Hill & McNamara, 2011; Leung, 2009; Rea–Dickins, 2004; Tsagari, 2016). It is argued that assessment is to be considered within the framework of language teaching methodology, not as an entity exclusive from it, and teachers are to be equipped with fundamental theoretical knowledge and practical skills and strategies to design/implement high-quality assessment (Brunfaut, 2014; Douglas, 2010; Jiang, 2017; Vogt, Tsagari, & Spanoudis, 2020). In addition to designing and implementing tailor-made assessment, they are expected to know how to utilize assessment to improve teaching by viewing it as a joint venture and by communicating assessment goals and data to stakeholders (Coombe et al., 2020). In other words, they also need to be qualified enough to exploit assessment data to evaluate a language course and its constituents and revise them both to improve student learning and meet program goals via sharing assessment data with related stakeholders (Kremmel & Harding, 2020; Norris, 2016; Scarino, 2013; Tsagari et al., 2018).

The discussions about the role of assessment in language education eventually direct attention to the quality of assessment training in pre-service and in-service teacher training programs (Coombe et al., 2020; Hill, 2017;

Purpura, 2016). ELT teacher training programs have a broad spectrum ranging from linguistics, literature, pedagogy, and guidance-counselling to approaches and methods of language teaching. With the help of this comprehensive focus, ELT teachers are expected to be equipped with the necessary knowledge and skills to teach English efficiently. However, it can be claimed that the training programs strictly emphasize language teaching methods and fail to give equal importance to assessment. Thus, the fact that a language teaching methodology adopted in a language program naturally creates its assessment system and is only complete with a sound assessment system seems to be neglected (Isik, 2020). Moreover, there prevails a common conceptualization of assessment that considers assessment as a practice following a definite period of teaching. Such a conceptualization deliberates assessment as a means to check the quality of the end-product (Barrios & Milla Lara, 2018; Tsagari et al., 2018). The research also indicates the negligence and the poor quality of language teacher training programs regarding assessment (Davies, 2008; Gan & Lam, 2020; Purpura, 2016; Spolsky, 2008; Vogt & Tsagari, 2014).

However, emphasizing issues related to language teaching methods and neglecting the assessment-related ones leaves professional teacher training programs incomplete. Such an incomplete perspective eventually impedes conceptualizing language education and assessment as the parts of a whole that complement each other (Inbar-Lourie, 2008; Levi & Inbar-Lourie, 2020; Spolsky, 2008). Furthermore, such a misconception may result in hypothesizing assessment as something secondary or something that can be handled by teachers with little or no training. Thus, the role of assessment needs to be highlighted to increase ELT teachers' awareness regarding how essential assessment is in language education. The strong focus on its role could break down the long-lasting, traditional conceptualization of assessment and create an impetus to help teachers get equipped with full-fledged theoretical and practical knowledge and skills to plan, implement, and evaluate sound, valid, and reliable assessment (Baker, 2016; Brown & Bailey, 2008; Tsagari & Vogt, 2017)

As assessment in language education has been studied extensively, it has been a major topic of interest. Ultimately, a general theoretical framework is needed to design, implement, assess, and evaluate effectively by collecting data from different language education contexts. In other words, as pointed out by Inbar-Lourie (2017), Kremmel and Harding (2020), and Vogt et al. (2020a), language teacher assessment competencies and perceived assessment needs must be regarded locally and cross-culturally to create a global body of knowledge on teacher assessment literacy, which can be exploited to design better teacher training programs in assessment. Since the evidence on assessment is scarce in Turkey, this study attempts to provide data from English as a foreign language (EFL) education in the Turkish context to contribute to evolving empirical evidence on assessment by presenting how ELT teachers perceive the notion of assessment and training on assessment. Ultimately, considering how assessment is a vital component of language teaching and education programs, this study aims to shed light on assessment by providing data from a different underexplored language education context. Moreover, this comprehensive study aims to explore how assessment is perceived by ELT teachers working at different levels of EFL education in Turkey. It is likely to contribute to redesigning the content of ELT teacher training programs to increase their assessment literacy. Ultimately, the study attempts to ameliorate the assessment practice, especially in the Turkish EFL context. It also intends to catch the attention of both teacher trainers and other authorities regarding the essential role of assessment in language education and design teacher training accordingly. Finally, it aims at contributing to data on assessment in the Turkish EFL context.

Literature Review

Assessment Training in ELT Teacher Training Programs

Obviously, assessment training offered in language teacher training programs is designed to help teachers perform assessment and evaluation tasks more effectively (Winterbottom et al., 2008). The content of these courses varies concerning the theoretical stance, program goals, and contextual factors (Baker, 2016; Fulcher, 2012; Levi & Inbar-Lourie, 2020; Menken et al., 2014; Purpura, 2016; Taylor, 2014; Vogt et al., 2020a). In other words, although the fundamentals of assessment are covered in the assessment courses, it is hard to talk about a fixed set of conceptual knowledge and practical skills offered to ELT teacher candidates (Brown & Bailey, 2008; Norris, 2016). Thus, the quest for forming the best possible assessment content in teacher training programs has been on-going. Foregrounded topics include assessment-related issues such as the nature and roles, the implementers, and the means and methods of assessment. (Davies, 2008; Hill, 2017; Leung, 2009; Purpura, 2016; Vogt et al., 2020a; Wall & Taylor, 2014).

Inspired by these discussions, researchers investigate English teachers' knowledge and skills in assessment. Hasselgreen, Carlsen, and Helness (2004) studied what teachers, teacher trainers, and experts in European countries knew about assessment and what they needed to know. The results indicated a low level of teacher assessment literacy. The participants lacked proper training, knowledge, and skills in all topics of investigation and needed in-service training in all of them. Vogt and Tsagari (2014) replicated the research conducted by Hasselgreen et al. (2004) and reported that the undergraduate programs did not sufficiently help the teachers develop basic concepts of assessment and practical skills to prepare and implement assessment. They tried to compensate for the insufficiency at work themselves and demanded in-service training on assessment in general. Tsagari and Vogt (2017) released the qualitative part of their study (Vogt & Tsagari, 2014) and indicated that undergraduate programs did not prepare the ELT teachers well enough in terms of assessment literacy and they felt inadequate regarding assessment and thus demanded in-service training.Other studies also attempted to explore the effects of training on language assessment practice. Baker and Riches (2018) showcased the positive impact of assessment training on language teachers' assessment literacy levels in Haitian high schools. Koh, Burke, Luke, Gong, and Tan (2018) studied how training affected the quality of assessment tasks developed by Chinese language teachers and reported that while teachers comprehended various aspects of task design, they had difficulty transferring their knowledge to practice. Vogt, Tsagari, and Csépes (2020) shared the results of a European project, Teachers' Assessment Literacy Enhancement (TALE), encompassing both EFL learners and teachers. They reported that although the teachers adopted a traditional assessment approach to testing, the learners were satisfied with it. They also pointed out low LAL levels for both the teachers and the learners as well as the need for assessment training for both groups.

Contextual Factors

Besides training, researchers also indicated the impact of contextual factors on teachers' assessment practice. Yan, Zhang, and Fan (2018) highlighted the impact of contextual factors on how EFL teachers practiced assessment and perceived their training needs. They found that assessment culture, policies, expected teacher roles, and local resources and constraints affected their practice and perceived needs. Vogt et al. (2020a) also pointed out that contextual factors influenced how teachers perceived the notion of assessment, their LAL levels, and the need for assessment training. In a recent study, Levi and Inbar-Lourie (2020) noted that teachers partially benefited from what they learned in generic assessment training and mostly disregarded language-related construct components when preparing their assessment tasks. Teachers' LAL was found to be formed by generic assessment, language-specific assessment, and contextual factors. Another study underlining the importance of contextual factors was conducted by Vogt et al. (2020a) who studied ELT teachers' perceptions of LAL and their training needs in Germany and Greece. They found that although their perception of assessment was the same, the contextual factors affected how they perceived their LAL levels and training needs.

The Quality and Content of the Assessment Courses

The quality and content of the assessment courses offered for foreign language teachers in universities have also captured the attention of the researchers. Bailey and Brown (1996) and Brown and Bailey (2008) investigated the content of assessment courses and compared the results across twelve years. It was found that the assessment courses covered the fundamentals of assessment, which formed the basis of assessment courses and were evolving gradually. In another study, Heritage (2007) studied the assessment training of EFL teachers in China and found that the EFL teachers were not given enough assessment training in their BA, MA, and Ph.D. programs, and received no in-service assessment training. In another study, Shohamy, Inbar-Lourie, and Poehner (2008) indicated that teachers did not receive enough quality training related to assessment and consequently they lacked fundamentals of assessment and the skills to practice quality assessment. From a different, yet similar, perspective, Xu and Liu (2009) reported that training and experience formed teachers' knowledge base, however, the way they practiced assessment was also controlled by their need to conform to group norms in a given context. Another study conducted by López Mendoza and Bernal Arandia (2009) indicated that teachers in Colombia did not receive enough quality training in assessment and they employed traditional product-oriented tests. Likewise, Jin (2010) found that the assessment courses offered to train teachers were adequate in content but the theoretical knowledge was not put into practice adequately in language classrooms. Scarino (2013) indicated that teachers needed to develop their theoretical knowledge base and be ready to implement context-sensitive assessment practices. Lam (2015) demonstrated that the language assessment training pre-service teachers received in Hong Kong was inadequate and far from bridging

the gap between theory and practice. Sultana (2019) explored the language assessment literacy levels of English teachers in Bangladesh and indicated that the assessment training the teachers received was inadequate, which subsequently prevented them from conducting effective assessment and exploit it to improve language education. In another study conducted in the Chinese context to explore the assessment training needs of university English instructors, Gan and Lam (2020) revealed that the language assessment training teachers received was not enough to meet their classroom-based assessment needs. Nevertheless, they did not want to receive advanced training in assessment due to various personal and contextual factors.

Assessment in the Turkish EFL Context

Although assessment has received global attention, professional associations such as the European Association for Language Testing and Assessment (EALTA) and the International Language Testing Association (ILTA) were founded, and it is a topic studied and promoted professionally by academic journals, it has not yet aroused enough attention in Turkey. Although language testing is a mandatory course in all ELT undergraduate programs, to the knowledge of the researcher, there is only one lecturer who has a Ph.D. degree in language testing. Moreover, research on assessment is scarce in Turkey. In one of those studies, Saricoban (2011) examined the tests prepared by English teachers and pointed out validity problems especially with respect to the target grammar items. In another study, Ozdemir-Yilmazer and Ozkan (2017) found that assessment in university prep schools was dominated by the proficiency exam, which functioned as a gateway for students to continue their academic programs. In another study, Mede and Atay (2017) replicated the study carried out by Vogt and Tsagari (2014) at a Turkish university and found that the assessment literacy of the teachers was limited for classroom-based assessment and assessment-related concepts. Focusing on the student perspective on EFL assessment, the findings of the study conducted by Isik (2000) revealed students' dissatisfaction with the EFL assessment practices at all levels of EFL instruction in Turkey, which they believed to be lacking in terms of assessment. Olmezer-Ozturk and Aydın (2018) suggested the need for a language assessment literacy instrument and shared the psychometric properties of the Language Assessment Knowledge Scale (LAKS) they developed to indicate its validity and reliability. However, they did not report anything about the assessment literacy levels of teachers. Moreover, they collected data only from ELT teachers working at Turkish universities.

Despite an increase in research on language assessment, it has not received enough attention in the Turkish EFL context and subsequently, insufficient data has been accumulated to pinpoint the essential role of assessment to foster EFL education. The research is scarce and the studies carried out so far are not broad enough to provide insights into how assessment is conceptualized and practiced by ELT teachers in Turkey. Although the European-based study conducted by Vogt and Tsagari (2014) included Turkish ELT teachers, it did not report anything exclusive to the assessment practices in Turkey. Furthermore, based on the findings of the study, Isik (2020) suggested the need for comprehensive teacher training in assessment to increase the assessment literacy of ELT teachers. Thus, there is a drastic need to explore and understand assessment in EFL education, in addition to how teachers perceive and practice it in Turkey. Moreover, the data obtained from this study was meant to contribute to evolving global data on assessment from a different EFL context. In short, this study aimed to present a broader perspective about assessment because it included different types of language programs at different institutions of various levels. Thus, it helped conceptualize English teachers' assessment literacy, that is, what they knew about assessment and how they put their theoretical knowledge into practice at institutions at different levels. More specifically the study focused on the following research questions:

1. How do ELT teachers in Turkey perceive training in English language assessment?
2. How is English language assessment in Turkey perceived and practiced by the ELT teachers?

Materials and Methods

Participants

English education is offered at all levels of education, primary to tertiary, in both public and private schools and institutions in Turkey. The ELT teachers in these institutions formed the target population of the study. Upon completing the bureaucratic procedures and receiving the official approval of the authorities, eight

primary schools, eight secondary schools, eight high schools, and eight universities from two major cities constituting about 30% of the population in Turkey were selected using random cluster sampling. The questionnaire was offered to all ELT teachers, 352 from K-12 schools and 157 from the universities, and among those, 125 ELT teachers from K-12 schools and 73 ELT teachers from the universities answered the questionnaire. In other words, 198 ELT teachers provided data about assessment. Two ELT teachers from each institution were selected using random cluster sampling for the follow-up interview. Using convenience sampling, one ELT teacher trainer responsible for planning and implementing in-service training participated in a personal conversation.

The data obtained from the first part of the teacher questionnaire inquiring about the teachers' demographic information revealed that 15% of the teachers had one to five years of teaching experience, 23% of the teachers had six to ten, 18% of the teachers had eleven to fifteen, 23% of the teachers had sixteen to twenty, and 22% of the teachers had 21 and above. As for the schools they teach at, 12% of the teachers were from public primary schools, 11% from private primary schools, 32% from public secondary schools, 16% from private secondary schools, and 28% of the teachers from universities. Regarding academic background, 80% of the teachers had a BA and 8% had an MA in ELT. Finally, 5% received a Certificate in English Language Teaching to Adults (CELTA), and 1% received Diploma in English Language Teaching to Adults (DELTA) training.

Data Collection

Using the triangulation design type of mixed-methods research design, the data was collected with the help of the questionnaire, follow-up interview, and observation from each type of institution (public/private) at each level of education (primary, secondary, high school, and university) in the fall semester of 2018.

Questionnaire

To elicit data about how the teachers perceive assessment, a questionnaire developed by Vogt and Tsagari (2014) was administered to the teachers (see Appendix 1). The teacher questionnaire was composed of two sections. The first comprised eight mixed types of items and covers the demographic information about the ELT teachers. The second consisted of three subsections including a three-point Likert-type scale about assessment-related topics. The first subsection was about classroom-based language testing and assessment (LTA) and included 11 items, the second was about the purposes of testing with eight items, and the third covered the content and concepts of LTA with 16 items. Vogt and Tsagari (2014) found that the internal consistency reliability of individual scales ranged from .80 to 93. For this particular study, the internal consistency reliability of the questionnaire was computed by using Cronbach's alpha and ranged from .78 to .93, which represents a high level of internal consistency.

Follow-up Interview

To get a deeper understanding of the assessment process, two ELT teachers from each type of school at different levels of education were interviewed about the topics covered in the questionnaire. The follow-up interview included eight guiding questions exploited by Vogt and Tsagari (2014). The interviews were recorded and transcribed for analysis.

Observation

The institutions were also visited once a week for one semester to observe the assessment- related practice. The purpose of the observation was to witness who prepared the assessment and how it was planned, prepared, and its data exploited. Three different teachers were observed in three different classrooms, each lasted one class hour (40 to 50 minutes). The observation was conducted using the assessment process flow chart suggested by Coombe, Folse, and Hubley (2011), which includes six major steps: planning, development, administration, analysis, feedback, and reflection. It also functions as a checklist for each step of the assessment process. A haphazardly chosen teacher-made test at the high school level was presented and evaluated using the checklist suggested by Coombe, Folse, and Hubley (2011).

Personal Conversation with an ELT Teacher Trainer

To get information about the in-service teacher training in assessment that ELT teachers received, the Ministry of Education was called to request an appointment. Upon contacting the department accountable for in-service teacher training, a personal conversation (October 12, 2019) was held with one of the ELT teacher trainers.

Data Analysis

SPSS Version 26 was used to analyse the data obtained from the questionnaires. The responses were analysed through descriptive statistics to obtain frequencies and percentages. The data elicited from the interviews were categorized and coded, and the frequencies and percentages were presented. The data obtained through the observations was also categorized and reported. The sample exam was also evaluated to give an overall idea about the exams implemented in the Turkish EFL context.

Results

Questionnaire

The data obtained from the teacher questionnaire were discussed in four parts, namely teacher training in assessment, classroom-focused assessment literacy, the purposes of assessment, and the content and concepts of assessment literacy.

Teacher Training in Assessment

Regarding in-service training in assessment, 4% of the teachers reported that they had received training in assessment and marked the frequency of the training as five years and above. They indicated that 1% of the training was provided by the university ELT departments, 2% by the ELT publishers, and 1% by the schools they worked in.

Classroom-Focused Assessment Literacy

Table 1 summarizes the assessment-related training the participants received in their pre-service education and the need for training in classroom-focused assessment literacy.

Table 1

Classroom- focused assessment literacy and need for training

		Training Received			Training Needed	
		Frequency	Percentage		Frequency	Percentage
Preparing classroom tests	Not at all	86	43.4	None	191	96.5
	A little	101	51.0	Basic	7	3.5
	Advanced	11	5.6	Advanced	0	0.0
Ready-made tests	Not at all	189	95.5	None	193	97.5
	A little	9	4.5	Basic	5	2.5
	Advanced	0	0.0	Advanced	0	0.0
Giving feedback based on assessment	Not at all	195	98.5	None	193	97.0
	A little	3	1.5	Basic	6	3.0
	Advanced	0	0.0	Advanced	0	0.0
Self-/Peer- assessment	Not at all	198	100	None	190	95.1
	A little	0	0.0	Basic	8	4.0
	Advanced	0	0.0	Advanced	0	0.0
Informal assessment	Not at all	198	100	None	174	87.8
	A little	0	0.0	Basic	24	12.1
	Advanced	0	0.0	Advanced	0	0.0
ELP or Portfolio	Not at all	173	87.4	None	154	77.8
	A little	25	12.6	Basic	44	22.2
	Advanced	0	0.0	Advanced	0	0.0

Table 1 indicates that an overwhelming majority of teachers had received no training in ready-made tests, giving feedback on assessment, self/peer assessment, informal assessment, and ELP or portfolio, except for *preparing classroom tests*, which was comparatively evaluated more positively in the classroom-focused assessment sub-section.As 5.6% of the teachers stated that they received advanced training on *preparing classroom tests*, about half of the teachers marked the *A little* option, and the rest (43.4%) *Not at all*. *ELP or portfolio* is the second item that relatively differed from the others. While an overwhelming majority of the

teachers picked the *Not at all* option, 12.6% of them marked the *A little option*. Although the teachers lacked adequate training on classroom-focused assessment, they did not report any significant need for training, excluding ELP or portfolio and informal assessment. Concerning training needed, 22.2% of the teachers indicated that they needed basic training in *ELP or portfolio* and 12.1% demanded basic training on *informal assessment*.

Purposes of Testing

Table 2 illustrates the responses of the ELT teachers regarding the purposes of testing and the need for training in this area.

Table 2

The purposes of testing and the need for training

		Training Received			Training Needed	
		Frequency	Percentage		Frequency	Percentage
Giving grades	Not at all	194	98.0	None	196	99.0
	A little	4	2.0	Basic	2	1.0
	Advanced	0	0.0	Advanced	0	0.0
Finding out what needs to be taught/learned	Not at all	198	100	None	192	97.0
	A little	0	0.0	Basic	7	3.0
	Advanced	0	0.0	Advanced	0	0.0
Placing students	Not at all	193	97.5	None	194	98.0
	A little	5	2.5	Basic	4	2.0
	Advanced	0	0.0	Advanced	0	0.0
Awarding certificates	Not at all	195	98.5	None	198	100
	A little	3	1.5	Basic	0	0.0
	Advanced	0	0.0	Advanced	0	0.0

Table 2 shows that concerning the training received and training needed, the items in this sub-section showed a nearly homogenous distribution. The teachers did not indicate any significant training on the *purposes of testing*, nor did they mention any significant training in this topic.

Content and Concepts of Assessment Literacy:

Table 3 presents the findings of the content and concepts of assessment literacy section of the questionnaire.

Table 3

Content and concepts of assessment literacy and training needed

		Training Received			Training Needed	
		Frequency	Percentage		Frequency	Percentage
Receptive skills (reading/listening)	Not at all	198	100	None	198	100
	A little	0	0.0	Basic	0	0.0
	Advanced	0	0.0	Advanced	0	0.0
Productive skills (writing/speaking)	Not at all	198	100	None	198	100
	A little	0	0.0	Basic	0	0.0
	Advanced	0	0.0	Advanced	0	0.0
Microlinguistic aspects (grammar/ vocabulary)	Not at all	192	97.0	None	198	100
	A little	7	3.0	Basic	0	0.0
	Advanced	0	0.0	Advanced	0	0.0
Integrated language skills	Not at all	198	100	None	192	97.0
	A little	0	0.0	Basic	7	3.0
	Advanced	0	0.0	Advanced	0	0.0
Aspects of culture	Not at all	176	88.9	None	198	100
	A little	22	11.1	Basic	0	0.0
	Advanced	0	0.0	Advanced	0	0.0
Reliability	Not at all	198	100	None	194	98.0
	A little	0	0.0	Basic	4	2.0
	Advanced	0	0.0	Advanced	0	0.0

		Training Received			Training Needed	
		Frequency	Percentage		Frequency	Percentage
Validity	Not at all	198	100	None	198	100
	A little	0	0.0	Basic	0	0.0
	Advanced	0	0.0	Advanced	0	0.0
Using statistics	Not at all	190	95.1	None	164	83.0
	A little	8	4.0	Basic	34	17.1
	Advanced	0	0.0	Advanced	0	0.0

Table 3 illustrates the perceptions of the teachers converge on the *content and concepts of literacy* in terms of training received and training needed. Excluding microlinguistic aspects, aspects of culture, and using statistics, the teachers reported that they received no training in this sub-section. Concerning the *aspects of culture*, 11.1% of the teachers reported that they received a little training. Likewise, 4% of the teachers stated that they received a little training on *using statistics*, and 3% marked the *A little* option for *microlinguistic aspects*. Although their content and concepts of literacy was significantly low, an overwhelming majority did not demand training, except for on using statistics for which 17.1% of the teachers reported the need for basic training.

Teacher Interviews

The data obtained from the interviews are summarized as follows:

Training
The data obtained from the interviews were examined concerning preservice and in-service training.

Preservice. All the teachers indicated that they received an assessment course. Although 92% of the teachers reported that they received an assessment course in ELT, 8% of the teachers said that they took a general course offered to all students of the faculty of education. Reporting the nature of the assessment course, 83% of those teachers said that the course was theoretical and they covered general concepts of assessment. Regarding the practical aspect of the course, 33% of them said that they prepared mostly multiple-choice type of items as practice. Evaluating the effectiveness of the course, 66% of them were not happy with the performance of the lecturer and 93% of the teachers found no relevance between the courses they received at the university and the assessment they practiced in their schools.

In-service. One teacher who participated in the interviews had an MA in ELT. Surprisingly she stated that she had received no courses on assessment in the MA program. Only 17% of the teachers indicated that they received in-service training on assessment and said that the content of the program was like the ones they received in university and had no practical value.

Three teachers with CELTA said that they received no training about assessment during the program. One of them reported that topics on assessment were mentioned theoretically during the training program.

Conceptualization of Assessment
All the ELT teachers stated that assessment was an official requirement and a tool to assess learner performance that is eventually reported in the form of grades. About one-sixth (17%) of the teachers mentioned that it was also a means of improving language programs. Moreover, about one-third (33%) of the teachers reported that the assessment created an extra burden; they would rather deal with language teaching methods and techniques, not assessment.

Function of Assessment
All the teachers considered assessment to be a screening process through which student performance was stratified. The exam scores were utilized to rank students and also decide who passed or failed. Besides, 8% of the teachers believed that assessment provided invaluable feedback about teacher performance and the effectiveness of teaching. Likewise, 8% of the teachers believed that it could function as leverage to revise the English program.

Resources for Assessment
With no exceptions, all the teachers said that they exploited the ready-made tests that accompanied the coursebooks prepared by international publishers, or used the tests they found on the internet. A quarter (25%) of the teachers stated that they obtained texts from the internet but prepared their test items. Only 8% of the teachers mentioned that they referred to the internet to expose themselves to alternatives for assessment and how to put them into practice.

Collaboration for Assessment
The overwhelming majority (92%) of the teachers defined collaboration as sharing the exam preparation process by dividing the number of exam tasks by the number of teachers. Then each teacher prepared the part s/he was responsible for. None of the teachers reported that they got help from their colleagues to check item quality, grammatical accuracy, or the format of the exam.

Means of Assessment
All the teachers considered the traditional pen and paper type of assessment as the common means of assessment. Additionally, 17% of the teachers said that projects affected the grades of learners. A quarter (25%) indicated homework and student participation as a means of assessment, and 8% indicated that what students did inside and outside the classroom mattered.

Exploiting Assessment Data
About one-third (34%) of the teachers said that assessment provided valuable information about student learning and it could be used to improve teaching. About one-sixth (17%) said that it provided feedback about their performance. They also implied that the administration tended to evaluate their performance in accordance with the exam scores of their students. Finally, 8% said that the data obtained from assessment could be used to evaluate the course in general.

The Effects of Assessment
All the teachers said that assessment determined the standing of learners among other students and their passing grades. All the teachers at the tertiary level reported that different types of exams served different purposes. Achievement exam scores determined if students were qualified enough to take the proficiency exam and the scores also served as a practice for their final exam.

Assessment Practice
An overwhelming (92%) percentage of teachers reported that experience, tradition, and habit shaped their assessment practices. They followed the footsteps of others and imitated the exam format adopted in their schools. The majority (83%) of the teachers felt adequate about preparing exams. None of them stated that they got help from their colleagues. Two-thirds (66%) of them stated that they preferred exams prepared by someone else, such as the testing office at the Ministry of Education.

Criteria for Assessment
All the teachers identified the target structures and vocabulary items from their coursebooks and prepared questions accordingly. They identified the items to be assessed and their weight (points allotted for each item) intuitively. At the tertiary level, the exams were developed by the testing office independent of the content of the syllabus. They served as gradual preparation for the proficiency exam. Thus, proficiency set the criteria.

Need for Assessment Training
None of the teachers mentioned any need for in-service training on assessment. Alternative assessment aroused little interest among teachers and only 8% of them reported the need for training in alternative assessment methods.

Observations
Setting
At the schools, the lessons were carried out according to the academic plan that specified the units of the coursebooks taught in each given week. Assessment-related activities were not observed until the week before the exam was scheduled. In the meantime, one teacher from a private secondary school gave quizzes fortnightly. Two teachers from the state schools checked whether the units in the workbooks were done by the students

and pluses and minuses were given for each student and they would be added up at the end of the semester and given as an oral grade. In a private school, the department assigned the assessment tasks to each teacher from each grade two weeks before the exam date. Each teacher shared a part and then they brought their parts together to form the exam. The other private schools did the same one week before the exam. It was common practice in those schools to either gave practice tests or worksheets focusing on the topics that would be covered on the exam. In the state schools, the exam preparation procedures were the same except for the practice tests and worksheets. At the universities, before the exam week, the teachers informed students about the content of the exams in general.

Feedback

At schools, giving learners feedback about the exam took place in the form of announcing their grades. The learners had a chance to examine their exam papers to see their exam performance and how the teachers scored their exam papers. The teachers had to rush when scoring exams to deliver student scores before the parent-teacher association meeting that was held once a semester to share the exam results with the parents.

At the universities, the learners learned their grades by logging into their accounts operated by the student affairs office. They had no chance to see their exam papers unless they officially applied to examine them.

Assessment Tools

The measurements and evaluations were based on pen and paper assessment. The knowledge of grammar, vocabulary, and reading comprehension formed the basis of the exams. The exams employed decontextualized items deprived of their functional use. Discrete items that required little to no learner production mainly dominated the exams. Listening and speaking were not assessed, but writing was assessed in all schools and universities for the learners who were at and above the elementary level. Two of the universities also included listening on the proficiency exams.

A Sample Exam Evaluation

One exam from a high school was chosen randomly and evaluated. The exam was far from meeting the qualities of a well-formed, sound assessment tool (see Appendix 2). First, it did not give any information about the exam. It did not say which class it was prepared for, what topics or units it covered, the date of the exam, any general instructions, or an overall grading scheme. It was poor in terms of face validity. The teacher himself/herself did not prepare the assessment items but found a theme from different sources. S/he cut and pasted them on the exam paper. The exam paper looked like a collage, a mixture of typed and hand-written/drawn sections. The items on the exam were compressed and disorganized. The exam started with a reading text with no instructions. The second task was chart reading. Both reading texts tested grammar and the usage of the simple present tense. Moreover, the learners could have easily found clues from the previous tasks to do task B. In Part C, there were possible answers. Moreover, there were unintentional clues from Text A again, such as "have lunch, watch a film". Part D is contextualized but only testing "am/is/are". The learners could easily find clues from other texts including "am/is/are". Task E tested two things at a time and lacked validity. The prerequisite task was to form the questions and then the next task was to match them with the appropriate answers. There was more than one correct answer on this task. Task F also tested two things at a time, both putting the sentences in the correct order and rewriting them considering the usage of the simple present tense. This task also provided unintentional clues for the other tasks and learners could find clues from other tasks to answer this item. Moreover, the grading was unfair because the teacher awarded two points for each item. Just writing "am, is, are" in Task D and creating the questions and matching them with the correct answers in Task E were of the same weight. Furthermore, the tasks were not organized from simple to complex. Task D was the easiest one and should have been the first task. In short, the exam employed a variety of tasks, but they were all created to test grammar, namely the correct usage of the simple present tense. To sum up, it was not meaningful, valid, or fair.

The Personal Conversation with the ELT Teacher Trainer from the Ministry of Education

To learn about the assessment-related teacher training policy and plans of the Ministry of Education, the top authority responsible for in-service teacher training, a personal conversation was held with a teacher trainer who asked that their name not be disclosed. During the conversation, it was elicited that the Ministry had no general in-service training aiming to train all the ELT teachers in certain periods. The in-service training was provided to only a few of the teachers who were nominated. No needs analysis was carried out to set the content of in-service training sessions. The topics were determined by the teacher trainers or by the academicians

invited from universities but they generally revolved around ELT methodology and techniques. Each year nearly the same in-service training with the same content was offered to ELT teachers.

The overall summary of the results revealed that the assessment training teachers received was inadequate in that they followed a traditional assessment practice either by imitating the existing widespread assessment tradition or emulating their colleagues. They tended to perceive assessment as an official procedure meant to assign grades to their students. Generally, they considered it a practice following a certain amount of teaching to test the quality of the end-product. They did not consider it to be a means of fostering language education. They were satisfied with what had they been practicing and did not mention any need for assessment training. In-service training was offered to a few teachers for a short period but it did not contribute to their assessment knowledge and skills. Put together, both the assessment training they received and the official procedures helped maintain the long-established traditional assessment practices.

Discussion

This research attempted to explore how ELT teachers perceived the training they received on English language assessment. It also sought to explore how they perceived and practiced it in EFL education. The data obtained from the questionnaires, interviews, sample exam evaluation, and observations corroborated one another and helped answer the research questions and draw disciplined inferences about how assessment was perceived and practiced in the Turkish EFL context. The findings showed that the ELT teachers lacked proper training in assessment and perceived assessment as an official procedure to grade their students. It was not the current perspective, but their experience and long-established tradition that informed/determined their assessment practice. However, although their assessment literacy was low, they were satisfied with it and did not feel the need to improve it. Regarding the first research question investigating the assessment background of the ELT teachers, this study indicated that the training they received was insufficient and did not sufficiently cover current assessment theory and practice. It was found that the assessment courses conducted during the pre-service training were organized around theoretical issues and lacked a practicum, which meant that teachers might still not be entirely aware of the conceptualization of assessment, and therefore not be able to efficiently put it into practice. Moreover, the teachers barely received any in-service training on assessment. The findings indicated that only few teachers enjoyed the privilege of receiving in-service training. What was more surprising was that the ones with MAs in ELT reported that they took no courses on assessment in their MA programs. In a similar vein, CELTA and DELTA did not cover assessment; when they did, they only raised some basic theoretical issues. Furthermore, the interview with an ELT teacher trainer from the Ministry of Education also indicated that assessment was not a top priority. The limited in-service training for small number of ELT teachers catered to language teaching methods and neglected assessment. Thus, this finding complied with those of Gan and Lam (2020), Hasselgreen et al. (2004), Heritage (2007), Lam (2015), Scarino (2013), Shohamy et al. (2008), Tsagari and Vogt (2017), and Vogt and Tsagari (2014) who also reported that ELT teachers did not receive enough training and displayed low assessment literacy levels. In short, regarding the preservice and in-service training on assessment, the situation in Turkey was unsatisfactory, which corroborates the findings of Tsagari and Vogt (2017) and Vogt et al. (2020b). Regarding the need for training, the findings in this study were in sharp contrast with the findings in other studies mentioned above as the teachers in this study were content with their current assessment literacy and did not feel the need for assessment training. As they lacked proper training in assessment, they were not aware of their inadequacy and believed that they were assessing effectively. They perceived assessment as a formal procedure, which they performed regularly, whilst meeting the expectations of authorities to grade their students. Moreover, they did not confront any assessment-related problems and criticisms. Ultimately, an overwhelming majority of the teachers did not report any need for assessment training.

As for the second research question, it was observed that the perception and practice of assessment adopted by ELT teachers reflected a traditional understanding. The teachers at all levels of education., adopting a narrow definition and function of assessment, tended to perceive it as an official procedure to assess the performance of students at certain intervals in order to give their grades. Likewise, the teachers did not consider assessment as an ongoing evaluation of each aspect of an English course but as a vehicle to assess student performance. They conceptualized assessment as an officially required post-teaching practice employed at the end of certain periods in the academic program. Similarly, they exploited traditional assessment techniques that verified the findings of Tsagari and Vogt (2017) and Vogt et al. (2020b). They administered pen and paper exams and considered the

homework performance of the students while evaluating their performance. They also had to evaluate term projects that were required by the regulations and assigned to the students volunteering to do a project in English at the beginning of each semester. The exam dates were clearly delineated at the beginning of each academic year, which served as the assessment plan. There was no reference to basic issues such as the principles of assessment, how student performance was evaluated, what kind of assessment types -including alternatives in assessment- were employed, or how assessment was exploited to improve syllabi, teaching/learning, course materials, methodology, teacher development, and the context. At the primary, secondary (middle), and high school levels, nothing was specified about the content of the exams. Thus, administering exams on time and submitting the grades to the system might potentially pose as the main requirement(s) for fulfilling their assessment obligations.

Concerning how the teachers practiced assessment, the quality of assessment and means of exploiting the data obtained from assessment to improve programs and teaching quality did not receive the primary stress. The ELT teachers practiced assessment based on their prior language learning and teaching experience, intuition, and traditions, thus emulating what others had done so far, which was also indicated by Jin (2010) and Xu and Liu (2009). The teachers considered assessment as a kind of forced duty and were not so happy carrying it out. Before preparing the exams, they did not prepare any table of specifications or any guidelines reflecting their objectives. They preferred it to be prepared by the English Department, the testing office, or the Ministry of Education. While preparing the exams, they tended to use coursebook-accompanying tests prepared by international publishers or collected from the internet. They shared the burden of exam preparation with their colleagues and prepared their share of the items. In the private schools, assessment seemed to create frustration among the teachers. While preparing the exams and submitting the final grades of their students, they felt the need to form a fine balance between the academic success of their students and the success-oriented expectations of the administration and parents. This finding complies with the findings of Gan and Lam (2020), Koh et al. (2018), Levi and Inbar-Lourie, 2020, Vogt et al. (2020a), and Yan et al. (2018) who indicated how contextual factors affected assessment practices. At the universities, the exam dates were marked on the academic calendar and carried out accordingly. The EFL teachers were not generally responsible for exam preparation. The exams were prepared and evaluated by the testing office. This forced teachers to be success-oriented because in the report prepared by the testing office the performance of each class was analysed, and indirectly, the ELT teachers were as well. Especially in the university prep classes, all the emphasis was placed on helping the students pass the proficiency exam. Thus, at the university level, teaching for testing dominated the program and the effect of assessment was felt rigorously. This finding on assessment at the university level is in line with that of Ozdemir-Yilmazer and Ozkan (2017).

When regarded concurrently, ELT teacher training in assessment and the perception and practice of assessment showcases that the lack of proper training in assessment directly influenced how the ELT teachers conceptualized and implemented assessment. Oblivious to the current discussions on assessment, the teachers were doomed to depend on their previous EFL learning, teaching, and assessing experience, which contributed to the unproductive application of pre-established traditional testing habits. Moreover, this study showed that the lack of proper training prevented them from adopting a scientific stance as opposed to a traditional one, deeming them susceptible to contextual influence, such as the assessment perspective adopted and practiced by their colleagues. Finally, the insufficiency of training led them to adopt a narrow conceptualization of assessment and show satisfaction towards what they were practicing. In other words, the lack of proper training yielded limited assessment knowledge, which resulted in insufficient assessment practice.

Conclusion and Implications

The issues raised in the purpose of the study were addressed succinctly. First of all, the results of this study are expected to contribute to the evolving global body of knowledge on teacher assessment literacy by providing empirical evidence from a different context of language education, namely Turkish EFL. Considering the local and cross-cultural need for research in language teacher assessment perceptions and competencies, the study also offered a perspective from an underexplored EFL context. Moreover, it pioneered and attempted to portray how assessment is conceptualized and practiced from a wider perspective with the data provided by ELT teachers from different types of schools and universities as well as varying levels of education. Hopefully this encompassing study initiates a pool of local research on teacher assessment literacy to shed light on the perception of assessment practiced in the Turkish EFL context. Finally, by indicating EFL teachers' low LAL

levels, the study is expected to pinpoint a mostly unexplored assessment problem, which has serious detrimental repercussions for their assessment practice. In doing so and in revealing this detrimental lack of awareness, the study is expected to make a noteworthy contribution by drawing the attention of both teachers and authorities to the fossilized assessment problem.

The study also highlighted that the teachers had low LAL levels, which gave rise to the narrow conceptualization of assessment and improper assessment practices. They cared more about preparing and administering the exam on time as indicated in their academic plan and tended to define assessment as an official requirement to evaluate what was taught. The void left by inadequate training in assessment was filled by a series of contributing factors, namely personal language learning/teaching experience, the assessment tradition at schools /universities, the need to conform to the assessment approach and practice adopted by their colleagues, and the assessment courses taken during the pre-service education. The last factor, the assessment courses conducted during the pre-service training, seemed to be the least effective because of their being too theoretical and inefficient to equip teachers with fundamental assessment literacy skills, strategies, and practical skills to prepare and evaluate relevant and sound assessment procedures. Consequently, this gap is filled with tradition, experience, imitating other colleagues, and adopting group norms.

These findings pointed out the need for pre-service and in-service training to align teaching and assessment and consider assessment as a complementary part of the teaching process.Considering the effect of assessment on language education, EFL teachers' weak assessment literacy skills could harm the language education process. However, there seems to be no authority (academic and institutional) to pinpoint the problems and propose solutions for effective assessment. Thus, by collaborating with universities, the Ministry of Education could form an assessment board to plan and implement data-driven teacher training on assessment. Consequently, an EFL assessment model that stems from the realities of that particular EFL context could be realized.

The study attempted to explore one unique aspect of the assessment-related discussions. There remain other issues to be investigated, such as: What means can be employed to help all stakeholders reconceptualize assessment? How can ELT teachers be made aware of their gaps in knowledge on current assessment practices and the need for training?How are context-sensitive pre-service and in-service training programs designed? How can practical, sustainable, career-long in-service training programs addressing all teachers be devised? Hopefully, the study can spark numerous other scholarly studies to focus on these issues to gather a rich amount of data to help make data-driven, disciplined decisions to improve the quality of assessment. Likewise, it is intended that the study contributes to making assessment in English a popular and hot topic among academics and ELT teachers and attract the attention of decision-makers. Such a focus on assessment, hopefully, ignites action to improve both assessment and language education.

This study had several limitations. The data was collected from only two major cities and failed to include all regions in Turkey. Moreover, the data obtained from public and private schools were not analyzed separately to compare these two types of schools. It would have been better to see if there were any differences between these two groups of ELT teachers concerning assessment orientation and practices.

List of Abbreviations

CELTA: Certificate in English Language Teaching to Adults
DELTA: Diploma in English Language Teaching to Adults
ELT: English Language Teaching
EFL: English as a Foreign Language
LAL: Language Assessment Literacy

References

Bailey, K. M., & Brown, J. D. (1996). Language testing courses: What are they? In A. Cumming, & R. Berwick (Eds.), *Validation in Language Testing* (pp. 236–256). Multilingual Matters.

Baker, B. (2016). Language assessment literacy as professional competence: The case of Canadian admissions decision makers. *Canadian Journal of Applied Linguistics/Revue Canadienne de Linguistique Appliquée*, *19*(1), 63-83.

Baker, B. A., & Riches, C. (2018). The development of EFL examinations in Haiti: Collaboration and language assessment literacy development. *Language Testing, 35*(4), 557-581. https://doi.org/10.1177/0265532217716732

Barrios, E., & Milla Lara, M. D. (2018). CLIL methodology, materials and resources, and assessment in a monolingual context: an analysis of stakeholders' perceptions in Andalusia. *The Language Learning Journal*, *48*(1), 60-80. https://doi.org/10.1080/09571736.2018.1544269

Brown, J. D., & Bailey, K. M. (2008). Language testing courses: what are they in 2007? *Language Testing, 25*(3), 349-383. https://doi.org/10.1177/0265532208090157

Brunfaut, T. (2014). A lifetime of language testing: An interview with J. Charles Alderson. *Language Assessment Quarterly, 11*(1), 103-119. https://doi.org/10.1080/15434303.2013.869818

Combee, C., Folse, K., & Hubley, N. (2011). *A practical guide to assessing English language learners.* Michigan University Press.

Coombe, C., Vafadar, H., & Mohebbi, H. (2020). Language assessment literacy: what do we need to learn, unlearn, and relearn?. *Language Testing in Asia, 10*(1), 1-16. https://doi.org/10.1186/s40468-020-00101-6

Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing, 25*(3) 327–347. https://doi.org/10.1177/0265532208090156

Douglas, D. (2010). *Understanding language testing*. Hoddler Education.

Fulcher, G. (2012) Assessment literacy for the language classroom. *Language Assessment Quarterly, 9(*2), 113-132. https://doi.org/10.1080/15434303.2011.642041

Gan, L., & Lam, R. (2020). Understanding university English instructors' assessment training needs in the Chinese context. *Language Testing in Asia, 10*(1), 1-18. https://doi.org/10.1186/s40468-020-00109-y

Hasselgreen, A., Carlsen, C., & Helness, H. (2004). *European survey of language testing and assessment needs. Part one: General findings*. www.ealta.eu.org/resources

Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan, 89*, 140-145. https://doi.org/10.1177/003172170708900210

Hill, K. (2017). Understanding classroom-based assessment practices: A precondition for teacher assessment literacy. *Papers in Language Testing and Assessment, 6*(1), 1-17.

Hill, K., & McNamara, T. (2011). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing, 29*(3), 395–420. https://doi.org/10.1177/0265532211428317

Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing, 25*(3), 385–402. https://doi.org/10.1177/0265532208090158

Inbar-Lourie O. (2017). Language assessment literacies and the language testing communities: A mid-life identity crisis? Paper presented at *39th Language Testing Research Colloquium*, Universidad de los Andes, Bogotá, Colombia.

Isik, A. (2020). Do Students Feel that they are Assessed Properly? *Iranian Journal of Language Teaching Research, 8*(1), 63-92. https://doi.org/10.30466/IJLTR.2020.120808

Jiang, J. (2017). Language assessment: critical issues—An interview with Antony John Kunnan. *Language Assessment Quarterly, 14*(1), 75-88. https://doi.org/10.1080/15434303.2016.1269770

Jin, Y. (2010). The place of language testing and assessment in the professional preparation of foreign language teachers in China. *Language Testing, 27*(4), 555–584. https://doi.org/10.1177/0265532209351431

Kremmel, B., & Harding, L. (2020). Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: Developing the Language Assessment Literacy Survey. *Language Assessment Quarterly, 17*(1), 100-120. https://doi.org/10.1080/15434303.2019.1674855

Koh, K., Burke, L. E. C. A., Luke, A., Gong, W., & Tan, C. (2018). Developing the assessment literacy of teachers in Chinese language classrooms: A focus on assessment task design. *Language Teaching Research, 22*(3), 264-288. https://doi.org/10.1177/1362168816684366

Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing, 32*(2), 169-197. https://doi.org/10.1177/0265532214554321

Leung, C. (2009). Developing formative teacher assessment: Knowledge, practice, and change. *Language Assessment Quarterly, 1*, 19–41. https://doi.org/10.1207/s15434311laq0101_3

Levi, T., & Inbar-Lourie, O. (2020). Assessment literacy or language assessment literacy: Learning from the teachers. *Language Assessment Quarterly, 17*(2), 168-182. https://doi.org/10.1080/15434303.2019.1692347

López Mendoza, A. A., & Bernal Arandia, R. (2009). Language testing in Colombia: A call for more teacher education and teacher training in language assessment. *Profile Issues in Teachers' Professional Development*, *11*(2), 55-70.

Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing, 30*(3), 329-344. https://doi.org/10.1177/0265532213480129

Mede, E. & Atay, D. (2017). English language teachers' assessment literacy: The Turkish context. *Dil Dergisi, 168*(1), 43-60. doi

Menken, K., Hudson, T., & Leung, C. (2014). Symposium: Language assessment in standards-based education reform. *TESOL Quarterly, 48(3)*, 586-614. https://doi.org/10.1002/tesq.180

Norris, J. M. (2016). Language program evaluation. *The Modern Language Journal*, *100*(S1), 169-189. https://doi.org/10.1111/modl.12307

Olmezer-Ozturk, E., & Aydın, B. (2018). Toward measuring language teachers' assessment knowledge: Development and validation of Language Assessment Knowledge Scale (LAKS). *Language Testing in Asia, 8*(1), 1-15. https://doi.org/10.1186/s40468-018-0075-2

Ozdemir-Yilmazer, M., & Ozkan, Y. (2017). Classroom assessment practices of English language instructor. *Journal of Language and Linguistic Studies, 13*(2), 324-345.

Purpura, J. E. (2016). Second and foreign language assessment. *The Modern Language Journal, 100,* 191-208. https://doi.org/10.1111/modl.12308

Rea–Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing, 21*, 249– 258. https://doi.org/10.1191/0265532204lt283ed

Saricoban, A. (2011). A study on the English language teachers' preparation of tests. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, *41*, 398-410.

Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing, 30*(3), 309-327. https://doi.org/10.1177/0265532213480128

Shohamy, E., Inbar-Lourie, O., & Poehner, M. (2008). *Investigating assessment perceptions and practices in the advanced foreign language classroom* (Report No 1108). Center for Advanced Language Proficiency Education and Research (CALPER).

Spolsky, B. (2008). Language testing at 25: Maturity and responsibility? *Language Testing, 25*(3), 297–305. https://doi.org/10.1177/0265532208090153

Sultana, N. (2019). Language assessment literacy: An uncharted area for the English language teachers in Bangladesh. *Language Testing in Asia, 9*(1), 1-14. https://doi.org/10.1186/s40468-019-0077-8

Taylor, L. (2014). General language proficiency (GLP): Reflections on the "Issues Revisited" from the perspective of a UK examination board. *Language Assessment Quarterly, 11*(2), 136-151. https://doi.org/10.1080/15434303.2014.896366

Tsagari, D. (Ed.). (2016). *Classroom-based assessment in L2 contexts*. Cambridge Scholars Publishing.

Tsagari, D., & Banerjee, J. (Eds.). (2016). *Handbook of second language assessment* (vol. 12). Walter de Gruyter GmbH & Co KG.

Tsagari, D., & Vogt, K. (2017). Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects. *Papers in Language Testing and Assessment, 6*(1), 41-63.

Tsagari, D., Vogt, K., Froelich, V., Csépes, I., Fekete, A., Green A., Hamp-Lyons, L., Sifakis, N. and Kordia, S. (2018). *Handbook of Assessment for Language Teachers*. Retrieved from http://taleproject.eu/. ISBN 978-9925-7399-1-2 (digital).

Vogt, K., & Tsagari, D. (2014) Assessment literacy of foreign language teachers: Findings of a European study, *Language Assessment Quarterly, 11*(4), 374-402. https://doi.org/10.1080/15434303.2014.960046

Vogt, K., Tsagari, D., & Csépes, I. (2020). Linking learners' perspectives on language assessment practices to teachers' assessment literacy enhancement (TALE): Insights from four European countries. *Language Assessment Quarterly, 17*(4), 410-433. https://doi.org/10.1080/15434303.2020.1776714

Vogt, K., Tsagari, D., & Spanoudis, G. (2020). What do teachers think they want? A comparative study of in-service language teachers' beliefs on LAL training needs. *Language Assessment Quarterly*, *17*(4), 386-409. https://doi.org/10.1080/15434303.2020.1781128

Wall, D., & Taylor, C. (2014). Communicative language testing (CLT): Reflections on the "Issues Revisited" from the perspective of an examinations board. *Language Assessment Quarterly*, *11*(2), 170-185. https://doi.org/10.1080/15434303.2014.902058

Winterbottom, M., Brindley, S., Taber, K. S., Fisher, L. G., Finney, J., & Riga, F. (2008). Conceptions of assessment: trainee teachers' practice and values. *The Curriculum Journal*, *19*(3), 193-213. https://doi.org/10.1080/09585170802357504

Xu, Y., & Liu, Y. (2009). Teacher assessment knowledge and practice: A narrative inquiry of a Chinese college EFL teacher's experience. *TESOL Quarterly, 43*(3), 493-513. https://doi.org/10.1002/j.1545-7249.2009.tb00246.x

Yan, X., Zhang, C., & Fan, J. J. (2018). Assessment knowledge is important, but …: How contextual and experiential factors mediate assessment practice and training needs of language teachers. *System*, *74*, 158–168. https://doi.org/10.1016/j.system.2018.03.003

Appendix A

Teacher Questionnaire

Part I. General Information

1. Do you work in (country)?	Yes ☐ No ☐
2. What subject(s) do you teach?	
3. What subjects have you studied?	
4. What is your highest qualification?	Please specify:
5. Type of school/institution you teach at:	
6. Average age of pupils:	
7. Your functions at school/institution:	☐Teacher☐Head of department or school☐Mentor ☐Advisory function for authorities (local government, ministry etc.)☐Other? Please specify:
8. During your pre-service or in-service teacher training, have you learned something about testing and assessment (theory and practice)?	☐Yes (please specify:) _____☐No

Part II. Questions About Training in LTA

1. Classroom-focused LTA

1.1. Please specify if you were trained in the following domains

	Not at all	A little (1-2 days)	More advanced
Preparing classroom tests			
Using ready-made tests from textbook packages or from other sources			
Giving feedback to students based on information from tests/assessment			
Using self- or peer-assessment			
Using informal, continuous, non-test types of assessment			
Using the European Language Portfolio, an adaptation of it or some other portfolio			

1.2 Please specify if you need training in the following domains

	None	Yes, basic Training	Yes, more advanced training
Preparing classroom tests			
Using ready-made tests from textbook packages or from other sources			
Giving feedback to students based on information from tests/assessment			
Using self- or peer-assessment			
Using informal, continuous, non-test types of assessment			
Using the European Language Portfolio, an adaptation of it or some other portfolio			

2. Purposes of testing

2.1. Please specify if you were trained in the following domains

	Not at all	A little (1-2 days)	More advanced
Giving grades			
Finding out what needs to be taught/ learned			
Placing students into courses, programs, etc.			
Awarding final certificates (from school/program; local, regional or national levels)			

2.2. Please specify if you need training in the following domains

	None	Yes, basic Training	Yes, more advanced training
Giving grades			
Finding out what needs to be taught/ learned			
Placing students into courses, programs, etc.			
Awarding final certificates (from school/program; local, regional or national levels)			

3. Content and concepts of LTA

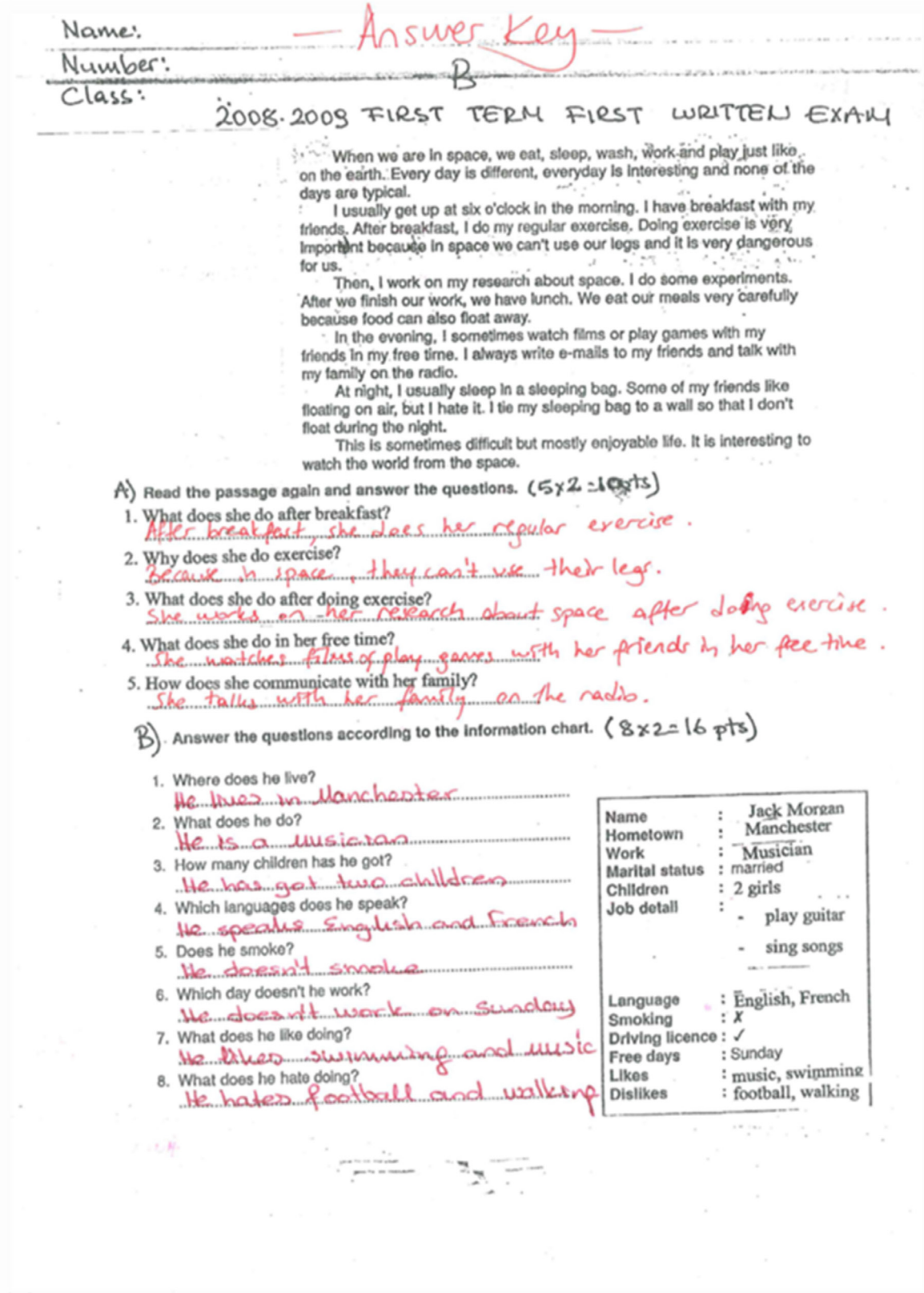
3.1. Please specify if you were trained in the following domains

	Not at all	A little (1-2 days)	More advanced
1. Testing/Assessing:			
a) Receptive skills (reading/listening)			
b) Productive skills (speaking/ writing)			
c) Microlinguistic aspects (grammar/vocabulary)			
d) Integrated language skills			
e) Aspects of culture			
2. Establishing the reliability of tests/assessment			
3. Establishing the validity of tests/assessment			
4. Using statistics to study the quality of tests/assessment			

3.2. Please specify if you need training in the following domains

	None	Yes, basic Training	Yes, more advanced training
1. Testing/Assessing:			
a) Receptive skills (reading/listening)			
b) Productive skills (speaking/ writing)			
c) Microlinguistic aspects (grammar/vocabulary)			
d) Integrated language skills			
e) Aspects of culture			
2. Establishing the reliability of tests/assessment			
3. Establishing the validity of tests/assessment			
4. Using statistics to study the quality of tests/assessment			

Appendix B



Name: — Answer Key —
Number: B
Class:

2008-2009 FIRST TERM FIRST WRITTEN EXAM

When we are in space, we eat, sleep, wash, work and play just like on the earth. Every day is different, everyday is interesting and none of the days are typical.

I usually get up at six o'clock in the morning. I have breakfast with my friends. After breakfast, I do my regular exercise. Doing exercise is very important because in space we can't use our legs and it is very dangerous for us.

Then, I work on my research about space. I do some experiments. After we finish our work, we have lunch. We eat our meals very carefully because food can also float away.

In the evening, I sometimes watch films or play games with my friends in my free time. I always write e-mails to my friends and talk with my family on the radio.

At night, I usually sleep in a sleeping bag. Some of my friends like floating on air, but I hate it. I tie my sleeping bag to a wall so that I don't float during the night.

This is sometimes difficult but mostly enjoyable life. It is interesting to watch the world from the space.

A) Read the passage again and answer the questions. (5x2=10pts)

1. What does she do after breakfast?
After breakfast, she does her regular exercise.
2. Why does she do exercise?
Because in space, they can't use their legs.
3. What does she do after doing exercise?
She works on her research about space after doing exercise.
4. What does she do in her free time?
She watches films of play games with her friends in her free time.
5. How does she communicate with her family?
She talks with her family on the radio.

B) Answer the questions according to the information chart. (8x2=16 pts)

1. Where does he live?
He lives in Manchester.
2. What does he do?
He is a musician.
3. How many children has he got?
He has got two children.
4. Which languages does he speak?
He speaks English and French.
5. Does he smoke?
He doesn't smoke.
6. Which day doesn't he work?
He doesn't work on Sunday.
7. What does he like doing?
He likes swimming and music.
8. What does he hate doing?
He hates football and walking.

Name	:	Jack Morgan
Hometown	:	Manchester
Work	:	Musician
Marital status	:	married
Children	:	2 girls
Job detail	:	- play guitar - sing songs
Language	:	English, French
Smoking	:	✗
Driving licence	:	✓
Free days	:	Sunday
Likes	:	music, swimming
Dislikes	:	football, walking

National Research University Higher School of Economics
Journal of Language & Education Volume 7, Issue 1, 2021

Jebbour, M. (2021). Self-Disclosure and Moroccan EFL Learners' Writing Development: Effects on Complexity, Accuracy, and Fluency. *Journal of Language and Education, 7*(1), 127-140. https://doi.org/10.17323/jle.2021.8620

Self-Disclosure and Moroccan EFL Learners' Writing Development: Effects on Complexity, Accuracy, and Fluency

Mohsine Jebbour
Moulay Ismaïl University

Correspondence concerning this article should be addressed to Mohsine Jebbour, ENCG/ SCALEC Research Laboratory, Moulay Ismaïl University, Présidence, Marjane 2, BP: 298, Meknès, Morocco, 50000.
E-mail: m.jebbour@umi.ac.ma

This article addressed the relationship between students' self-disclosure—that is, sharing social and positive information—and writing development in the English language classroom. The pre-test-post-test research design was adopted to assess whether students' reflections on personal positive experiences including feelings and opinions help improve their writing output as measured by complexity, accuracy, and fluency. The participants, drawn from a convenience sample, were 15 Moroccan students enrolled in the department of English studies at a Moroccan university. These participants were included to establish a homogenous level of English proficiency in writing. The participants completed a pre-test, six positive self-disclosure topics, and a post-test. A paired-sample t-test was computed to determine if a significant mean difference existed between the pre- and post-tests. Although the descriptive statistics suggest that the learners showed relative improvement in complex and fluent language, their overall writing development did not reach a statistically significant difference level. Although differing writing prompts and learners' academic learning experiences influenced the overall findings, this study contributes to the debate about the role of self-disclosure activities in improving certain language components in writing and calls for developing study programs that consider students' personal lives in language arts classes.

Keywords: CAF measures, EFL students, self-disclosure, positivity, writing development

Introduction

In departments of English studies in Morocco, many courses require students to show their work through writing, a widely used vehicle for these learners to practice and enhance language skills in the English as a foreign language (EFL) context (Birhan, 2018; McDonough & Fuentes, 2015). Hence, this study considers writing that involves students' production of complex, accurate, and fluent language (see Norris & Manchón, 2012).

Students' reflections on personal disclosures—including feelings, opinions, and experiences—are anchored to language learning theory, particularly second language (L2) motivation, language teaching methodology, and writing pedagogy (for more information, see Jebbour, 2020). This inclusion supports this study's objective to consider the relationship between self-disclosure and language development in writing.

Reflecting on self-disclosures can support learners' writing performances, in that writing about personal information fosters engagement in the task, increases the proportion of topic-relevant main clauses, allows students to pay close attention to language form, and provides access to well-known information (Bento, 1991; Dörnyei, 1994; Foster & Skehan, 1996; Guilloteaux & Dörnyei, 2008; Skehan & Foster, 1997). Consequently, students become familiar with the content of the target task, particularly the level of sentence structures, vocabulary, and grammatical rules. Assumably, "as students incorporate this understanding and awareness into their own writing, their writing performance, particularly the complexity [, accuracy] and fluency of their essays, will also improve" (Barrot, 2018, p. 188).

Aims and Significance of the Study

This study explores whether students' reflections on personal disclosure contribute to their writing development in university-level EFL classrooms. Specifically, this investigation aimed to determine whether

undergraduate EFL learners' reflections on personal positive experiences including feelings and opinions aid writing development in terms of three research questions:

1. Does reflecting on positive self-disclosure help improve EFL learners' production of complex language in writing?
2. Does reflecting on positive self-disclosure help improve EFL learners' production of accurate language in writing?
3. Does reflecting on positive self-disclosure help improve EFL learners' production of fluent language in writing?

Results from this study will inspire language educators, including university teachers of English who consider writing a venue for language development. The results will also increase these teachers' understanding of the role of personal tasks in language arts classes.

Literature Review

Classroom Self-Disclosure

Self-disclosure was first a subject of research in interpersonal relationships in humanist psychology and subsequently in communication studies. In psychology, self-disclosure refers to "the act of revealing personal information to others" (Jourard, 1971, p. 2), where sharing personal information with people is an underlying criterion of a healthy personality. In communication studies, self-disclosure means "any message about the self that a person communicates to another" (Wheeless & Grotz, 1976, p. 338).

Students' personal disclosure in the classroom is highly affected by others' responses. If responses are positive and supportive, students will feel safe to self-disclose (Wakefield, 2009). However, a negative reaction, particularly from teachers and classmates, may cause students not to reveal their personal disclosures (Berman, 2001; Derlega et al., 2010).

Students may share different topics when engaging in self-disclosure. Specifically, social penetration theory (SPT) posits that individuals may communicate superficial, social, and/or intimate information to others. Superficial information involves sharing biographical information such as names, hometown, and hobbies. Social information pertains to disclosing opinions, feelings, experiences, attitudes, etc.; and intimate information emerges when the discloser deeply shares private secrets, such as family problems (Altman & Taylor, 1973; Cayanus & Martin, 2016; Derlega et al., 2008).

Students' self-disclosures can take different dimensions: relevance, negativity, and positivity (Cayanus & Martin, 2016). Relevance occurs when disclosing information relating to the course content, like sharing an experience of doing voluntary work when discussing community service (Jebbour, 2018; Jebbour & Mouaid, 2019). Negativity involves sharing 'bad' information in the classroom like lying to teachers, whereas positivity entails disclosing desirable aspects of one's life like getting an 'A' in class (Cayanus et al., 2009).

Although different topics and dimensions govern the process of self-disclosure, the operational definition chosen for inclusion in this study is self-disclosure, which involves students' communication of social and positive information in writing. It is essential to note that sharing intimate disclosures is socially undesirable and inappropriate in the classroom setting and is likely to deteriorate the professional aspect of the teacher-student relationship (Lannuti & Straumann, 2006; Myers, 1998; Zhang et al., 2009). Hence, this study supports the sharing of social information in the classroom. Examples of such social information include personal feelings, opinions, and experiences (Guilloteaux & Dörnyei, 2008). In addition, because self-disclosing desirable information helps foster interpersonal relationships (see West & Turner, 2010), this justifies considering the positivity dimension in this research. The implication is that students' sharing of personal positive information in the writing classroom may help them establish a healthy relationship with the reader (i.e., classmates and/or teachers).

The use of self-disclosure activities in writing is grounded in the expressive/process movement in English language pedagogy. This approach considers writing to be a means for self-discovery, thereby giving students

an opportunity to express their personal feelings through journal writing and narratives (see Lam, 2015; Reid, 2001), which facilitate revision as an essential component of the writing process (Stotsky, 1995). Tasks involving self-disclosure may also function as a learning practice (Moon, 2004), inasmuch as students have "opportunities to explore [their] own lives and to discover things about themselves and others that they may not otherwise have learnt" (Stotsky, 1995, p. 758). Finally, expressing self-disclosures allows students to enhance reflective thinking (Hoadley-Maidment, 2000), in that personal writing "can be among the most intellectually rigorous genres, demanding self-discipline and self-criticism" (Berman, 2001, p. 7).

Complexity, Accuracy, and Fluency in Writing

Complexity, accuracy, and fluency (CAF) measures are considered essential components of language performance in writing (see Norris & Manchón, 2012). Experts in the field of L2 acquisition have devoted close attention to the study of the dimensions of CAF to understand language performance (see Wolfe-Quintero et al., 1998; Barrot & Gabinete, 2019).Yet, researchers use different parameters to observe the development of the three CAF measures in learners' spoken and written language. Thus, considering the different parameters of the CAF measures is relevant.

First, complexity refers to the use of varied and elaborated lexical and syntactic structures (Barrot & Gabinete, 2019). Previous research used parameters, such as word frequency, sentence length (Peng & Wang, 2018), simple, compound, complex, and compound-complex sentence types, subordination/coordination, word per finite-verb, and complex noun structures to measure linguistic and syntactic complexity (Rosmawati, 2014).

Second, accuracy involves the ability to produce error-free language in writing (Barrot & Gabinete, 2019) and is measured by different parameters, including holistic scales, error-free units, and the number of errors with/without classification (see Polio & Shea, 2014). These measures have a degree of reliability and validity, and each measure is used depending on the purpose of the study. For instance, holistic scales measuring accuracy in terms of vocabulary, spelling, punctuation, syntax, idiom use, etc. are usually used for placement purposes (Polio, 2001).

Third, fluency involves the "ability to produce written words and other structural units in a given time" (Barrot & Gabinete, 2019, p. 2). That is, the more a language learner is fluent, the faster he/she can retrieve vocabulary and linguistic chunks to construct a written text. The literature on this parameter indicates that fluency in L2 writing is measured from the product or process perspectives or both. When looking at writing as product, measures, composition scholars count the number of words, clauses, and T-units produced in a final text (Larson-Freeman, 2006; Polio, 2001). However, when examining L2 writing through a process lens, researchers consider other measures, including the mean number of characters, mean number of characters during rewriting, mean pause time length between words, and standard deviation of characters (Van Waes & Leijten, 2015).

CAF measures are "dimensions for describing language performance, most frequently used as dependent variables to assess variation with respect to independent variables such as acquisition level or task features" (Palloti, 2009, p. 590). Therefore, the type of task that teachers assign in the writing class affects students' content of writing, their degree of engagement in the production process, and the learning outcomes achieved in the end (Norris & Manchón, 2012). Tasks involving self-disclosures play a crucial role in increasing language students' interest and engagement in the learning process (Dörnyei, 1994; Guilloteaux & Dörnyei, 2008; Stotsky, 1995). This is well attested by the fact that Personalised Language Use – that is, writing about self-disclosure – was found to increase motivation to use the foreign language (Bernard, 2010) and guarantee meaningful language learning (Iida, 2016), which are essential for interlanguage development (Skehan & Foster, 1997).

Research exploring the impact of self-disclosure tasks on students' language development in the language arts classroom remains sparse in the currently available literature. Writing about self-disclosures in a computer-mediated communication environment increases learning autonomy and improves students' writing in English (Serag, 2011). Specifically, when students talk about personal information, they produce a larger amount of accurate output (Foster & Skehan, 1996; Skehan & Foster, 1997). In addition, when students write about personal experiences, they tend to produce lengthy pieces of writing with accurate sentences and suitable diction (Komalasari, 2013). Such insights are a point of departure to explore whether reflecting on positive self-disclosures is a vehicle through which learners can develop their language skills by producing complex, accurate, and fluent language in writing.

Materials and Methods

Research Design

A number of studies on L2 writing development used cross-sectional research designs without having control groups (see Polio & Park, 2016). This study adopted the same methodological approach to test if university students' reflections on their positive self-disclosures aided writing development in terms of CAF in the EFL classroom. Data were collected and coded using a quantitative approach and submitted to statistical analyses to answer the three research questions formulated at the onset of this study.

Assessments and Measures

The measures adopted to assess writing development in the participants' pre- and post-tests were complexity: counting the number of clauses per T-unit (C/T)–a T-unit refers to an independent clause and its dependent clauses (Barrot & Gabinete, 2019, p. 4)–, accuracy: counting the number of error-free T-units per T-unit (EFT/T), and fluency: calculating the number of words per T-unit (W/T). These CAF measures served as a tool to assess the participants' writing performance before participating in this study.

Participants

This study was limited to undergraduate students from the department of English studies at a Moroccan university. Upon receipt of the approval from the graduate office in charge of research activities at the university, the researcher recruited 19 participants who were willing to take part in this study because self-disclosure is considered a voluntary activity (Antaki et al., 2005).

The 19 participants completed a writing activity before they were officially included in the study. The results indicated that 15 participants' writing output showed no statistically significant difference in the production of C/T, F (2, 12) = .19, p = .82, EFT/T, F (2, 12) = .78, p = .48, and W/T, F (2, 12) = .93, p = .42. Based on this result, these 15 students (7 males and 8 females), with ages from 19 to 25 years, were recruited. As for their education level, five were first-year students, five were second-year students, and five were third-year students.

The anonymity of the participants was guaranteed by giving each student a code (e.g., Participant 1, Participant 2, etc.). To ensure confidentiality, the participants' names are not mentioned throughout the manuscript.

Instrumentation

This study employed a pre-test-post-test instrument to assess whether the EFL students' reflections on their positive self-disclosures aided the development of CAF in their writing output (i.e., pre-test and post-test). Prior to collecting actual data from the participants, the researcher, in consultation with a university professor of English with more than 30 years of teaching experience, developed eight writing prompts (six self-disclosure topics, a pre-test, and a post-test).

Four of the eight writing prompts were given to five undergraduate students of English. Three participants expressed unfamiliarity with the topics, which resulted in developing new writing prompts that were again distributed to the same students, who finally affirmed familiarity with the topics. Given that familiarity reduces working memory demands (Klingberg, 2010) and enhances students' writing performance (Graham et al., 2015), this justified the validity of the writing prompts.

Materials

Participants responded to the following prompts:

Pre-test
In one coherent paragraph, tell the reader about your favourite movie, book, or short story, and why people should watch it or read it.

Post-test

The university should organise a visit to the United States or England for students of English as a part of their learning experience. Do you agree or disagree? Explain your answer and use examples to support your opinions.

Self-Disclosure Writing Prompts

1) Write one coherent paragraph in which you define the characteristics of a positive friendship. Explain who your best friend is and express your feelings about him/her. 2) Write one coherent paragraph in which you recount a positive learning experience at university. Describe the experience and give your opinions and feelings about it. 3) In some countries, young adults have jobs while they are still students. Do you think this is a good idea? Support your opinion by using specific reasons and examples. 4) Write one coherent paragraph in which you define the characteristics of a good teacher. Tell the reader who your best teacher is and express your feelings about him/her. 5) Write a coherent paragraph in which you tell the reader about an exciting travel experience you have had alone or with a friend or family member. Specify the time and place and your feelings about the experience. 6) People go to university for different reasons (for instance, to expand their knowledge, to be important people in society, etc.). According to your experience as a student, tell the reader positive reasons for going to university. Use specific reasons and examples to support your answer.

Procedure

Initially, the participants completed the writing pre-test. After collecting the pre-test scripts, the participants completed six writing topics that elicited their positive personal disclosures. Next, the researcher distributed the writing post-test after the completion of the self-disclosure topics. The participants received no feedback on their writing performance throughout the study. Overall, the students had one week to complete each writing activity. They completed the writing prompts outside the classroom because attendance is not mandatory at the university, and the participants had different schedules. Such circumstances required giving the participants enough time to complete the writing tasks at their convenience. Hence, the writing prompts were collected on different days each week because meeting the participants at the same time was elusive. The participants' writing output was coded after ensuring that they did not deviate from the purpose of each writing task.

Coding

The data were coded for the three CAF measures. For the complexity construct, the total numbers of clauses (C) were divided by the total numbers of T-units (T) to obtain C/T. As for accuracy, the total numbers of error-free T-units (EFT) were divided by the total numbers of T-units (T) to have EFT/T. Regarding fluency, the total numbers of words (W) were divided by the total numbers of T-units (T) to obtain W/T. Table 1 summarises the coding scheme used in this study.

Table 1

Calculation of CAF measures

Component	Code	Calculation
Complexity	C/T	Clauses per T-unit
Accuracy	EFT/T	Error-free T-units per T-unit
Fluency	W/T	Words per T-unit

The participants' means and standard deviations in the three CAF measures in the pre-test were as follows: C/T: *M = 3.73, SD = 1.63*, EFT/T: *M = .37, SD = .20*, W/T: *M = 19.75, SD = 3.95*. Their means and standard deviations in the post-test were as follows: C/T: *M = 4.06*, *SD = 1.56*, EFT/T: *M = .31, SD = .32*, W/T: *M = 23.06, SD = 7.68*.

Inter-Rater Reliability

The pre-test and post-test texts (a writing post-test sample is provided in Appendix A) were submitted to three raters. The author was rater 1. Rater 2 was a university professor of English with more than 15 years of experience in teaching. Rater 3 was a university professor of English who had more than 30 years of teaching experience. The raters employed the three parameters (i.e., C/T, EFT/T, and W/T) for tracking CAF measures

to analyse the participants' writing output (the analysis of the writing post-test sample is provided in Appendix B). A meeting was held between the author and each of the raters to discuss the accuracy of the coding.

Percent agreement between raters was employed to measure reliability. The rationale for using the percentage of overlapping positive cases was that it is considered a relatively better measure of reliability since it yields the cases confirmed by the raters (Rosmwati, 2014). In the current research, the percent agreement reached 90% for complexity (C/T), 76% for accuracy (EFT/T), and 88% for fluency (W/T). Such results are considered consistent and trustworthy (van Geert & van Dijk, 2003).

Data Analysis

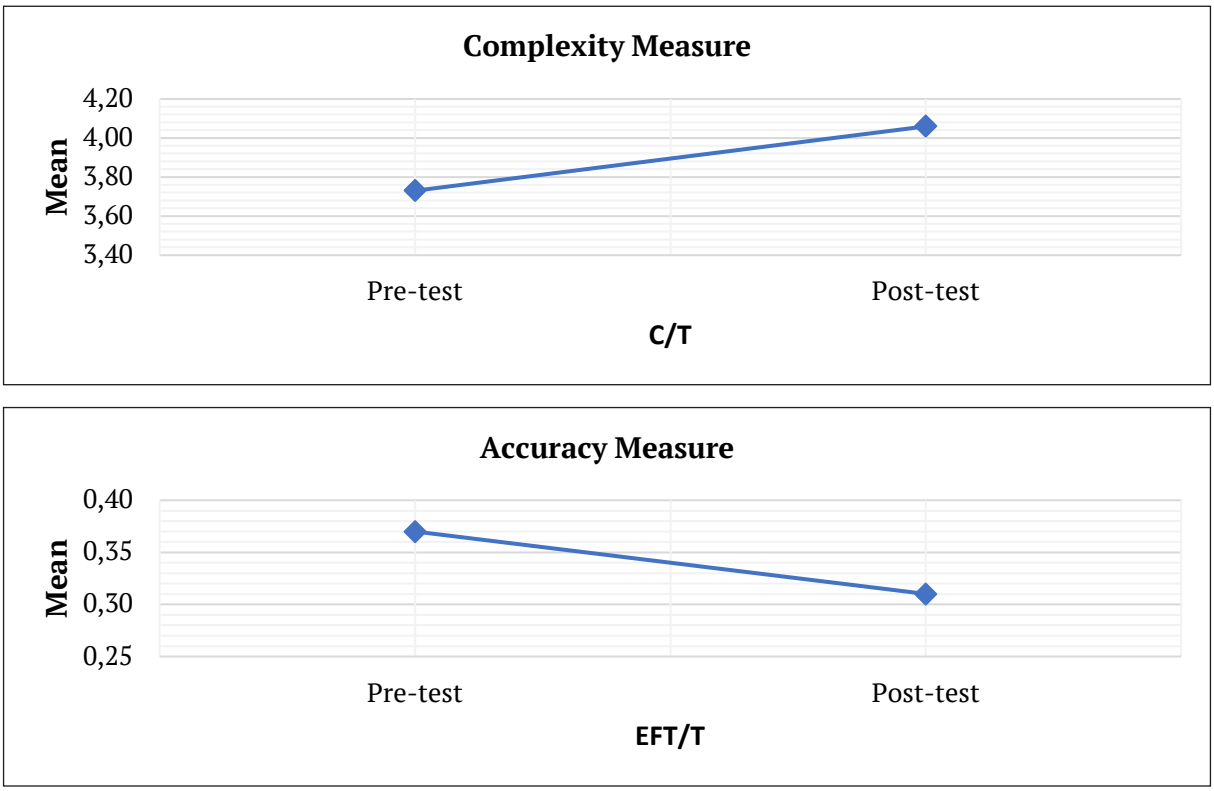
A paired-sample *t*-test was conducted to analyse the data gathered from the participants; $p < .10$ was required for significance as the recommended set alpha in L2 research (see Larson-Hall, 2010). Specifically, three paired-sample *t*-tests were computed to explore if a significant mean difference existed between the participants' pre- and post-tests in an effort to provide empirical evidence supporting whether the EFL students' reflections on their positive self-disclosures aided the development of CAF in their writing output.

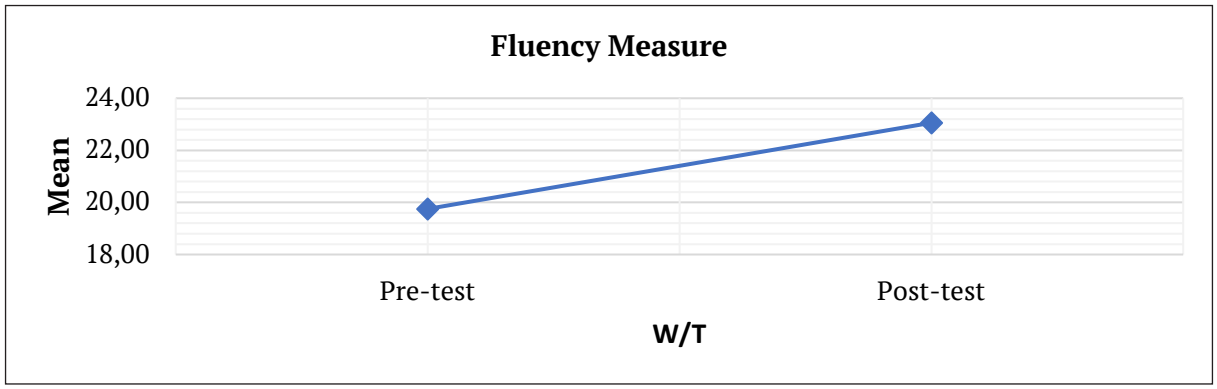
Results

Descriptive Statistics

The participants' mean scores on CAF measures, shown in Figure 1, indicated improvement in the production of C/T and W/T and a breakdown in the production of EFT/T. Over the eight-week period of this study, the EFL students' writing performance was more complex and fluent but less accurate after completing the six self-disclosure prompts.

Figure 1
The participants' mean scores on CAF measures in the pre- and post-tests





Main Findings

The first research question asked whether reflecting on positive self-disclosures helps improve EFL learners' production of complex language in writing as measured by C/T. The results indicated that the students' production of C/T did not reach a statistically significant difference in the scores between the pre-test (*M = 3.73, SD = 1.63*) and the post-test (*M = 4.06*, *SD = 1.56*); *t* (14) = -.71, *p* > .10 *(p* = .49) (see Table 2).

Table 2
Paired-sample t-test for complexity in the students' pre- and post-tests

Complexity	Paired Differences		t	df	Sig. (2-tailed)
	Mean	SD			
C/T	-.33	1.82	-.71	14	.49

Note.*p < .10

The second research question asked whether reflecting on positive self-disclosures helps improve EFL learners' production of accurate language in writing as measured by EFT/T. The results showed that the participants' production of EFT/T did not reach a statistically significant difference in the scores between the pre-test (*M = .37, SD = .20*) and the post-test (*M = .31, SD = .32*); *t* (14) = .87, *p* > .10 (*p* = .39) (see Table 3).

Table 3
Paired-sample t-test for accuracy in the students' pre- and post-tests

Accuracy	Paired Differences		t	df	Sig. (2-tailed)
	Mean	SD			
EFT/T	.06	28	.87	14	.39

Note.*p < .10

The third research question asked whether reflecting on positive self-disclosures helps improve EFL learners' production of fluent language in writing as measured by W/T. The findings showed that the participants' production of W/T did not reach a statistically significant difference in the scores between the pre-test (*M = 19.75, SD = 3.95*) and the post-test (*M = 23.06, SD = 7.68*); *t* (14) = -1.7, *p* > .10 (*p* = .106) (see Table 4).

Table 4
Paired-sample t-test for fluency in the students' pre- and post-tests

Fluency	Paired Differences		t	df	Sig. (2-tailed)
	Mean	SD			
W/T	-3.30	7.4	-1.7	14	.106

Note.*p < .10

Discussion

This study is the first to assess whether undergraduate learners' reflections on positive self-disclosures aided writing development in terms of CAF in the EFL classroom, especially in Morocco. Although one previous study (Serag, 2011) examined self-disclosure and writing, the author used a different research design and had dissimilar objectives from the current study. Hence, results from this manuscript are incomparable with those of Serag's study.

The participants were relatively homogeneous in their level of English proficiency before completing the six self-disclosure activities (i.e., the treatment). Although descriptive statistics revealed that the participants' production of complex and fluent language improved after the treatment, the main findings indicated that the learners' writing development did not reach a statistically significant difference in terms of complexity (C/T), accuracy (EFT/T), and fluency (W/T). With regard to complexity, although the participants' mean score on the post-test (M = 4.06) increased from the mean score on the pre-test (M = 3.73), the learners showed no significant development in C/T, indicating that simple and compound sentences were prominent in their pieces of writing. This conclusion supports the claim that writing about personal matters encourages the production of simple linguistic forms (Foster & Skehan, 1996). Thus, the learners' use of complex sentences including one or more dependent clauses (i.e., T-unit) in writing was weak. As for accuracy, the students produced more errors when producing complex sentences throughout writing, and the descriptive statistics confirmed this result, as the test takers' mean score in EFT/T on the post-test (M = .31) decreased from the mean score of the pre-test (M = .37). With respect to fluency, the descriptive statistics showed that the learners' mean score in W/T on the post-test (M = 23.06) was higher than that of the pre-test (M = 19.75). However, the t-test was found to be statistically insignificant, suggesting that the participants' length of complex sentences was modest because of having a lower amount of vocabulary.

The absence of significant differences regarding the test takers' writing development reiterates that the CAF measures "do not operate in complete independence from each other, and that findings obtained by CAF measures depend crucially on the participants involved and on the context in which the data are collected" (Housen & Kuiken, 2009, p. 469). The participants did not show significant development in writing due to a number of factors.

First, the participants' inability to produce complex sentences might have been affected by the different task prompts that elicited differing performances. While the pre-test task was a narrative and persuasive piece of writing, the post-test task was an argumentative piece of writing. Perhaps the argumentative type of writing was more difficult than the narrative and persuasive type of writing. Second, the students' low grammar levels and the non-use of dictionaries during the writing process justify the breakdown in producing accurate language on the post-test. Although the study did not classify the errors made by the participants, the frequent mistakes identified in their writing were concerned with spelling, word choice, and tense use. The learners completed the writing topics outside the classroom, where they had enough time to consult dictionaries and grammar books to construct an accurate piece of writing. However, writing in an EFL context requires learning to express thoughts while simultaneously learning the target language; as such, "grammatical errors are unavoidable because of the cognitive demand of performing both tasks at the same time" (Liao, 2016, p. 77). Third, the learners' deficiency in using complex language without mistakes negatively affected their production of a higher number of words per T-unit in return, indicating that they need to enrich their language repertoire with vocabulary. Fourth, the students' academic learning experiences are the last factor that best explains the overall findings. First-year students were in the process of integrating into the university atmosphere after graduating from high school, and hence their encounter with writing for the first time outside the classroom was a new task that required skills that they might have not developed yet. For instance, teachers expect novice learners to produce simple language in their first academic year because they will study advanced grammar in the second year. Second-year students were in the process of studying the complex sentences during the implementation of this study. While participation in this research was an opportunity for these students to practice their language, they may have lacked the skill of transferring what they had learnt in the advanced grammar class to the writing pre- and post-tests. However, third-year students were beginning to write a research paper for the first time and were obliged to submit the theoretical part of their monographs to their supervisors before the end of the first semester to fulfil graduation requirements. These factors coupled with preparation for final examinations created frustration among the participants, who decided to devote less

effort to complete the writing tasks, thereby displaying insufficient writing performance after they completed the self-disclosure topics.

As a final note, the main obstacle to assessing the role of writing activities in most departments of English in Morocco involved students' perceptions of writing as a boring and demanding activity and the belief that homework was not a course requirement. Although some professors assign writing activities, many students object to doing the task in or out of class. This perception explains the limited number of participants, whose decision to take part in this study should be understood as an act of kindness to help the researcher collect data and an opportunity to become familiar with the types of prompts that they may have in summative examinations. Being aware of the challenges affecting L2 writing research in higher education in Morocco explains the emergence of limitations relating to internal and external validity.

The first limitation was concerned with the low number of participants who took part in this investigation. Inclusion of a larger number of participants in the study could have yielded significant scores between the pre- and post-test. The second limitation involved the absence of examining the effects of extraneous variables, such as the participants' degree of motivation on writing development. Another methodological limitation was related to the unlimited time set for the participants to finish the writing activities. If time for completing writing tasks had been controlled, the analysis could have yielded different results and conclusions. The last limitation had to do with external validity. That is, conclusions could be drawn only about the target participants who participated in this study since EFL students' overall language abilities differ even within one grade level due to the non-use of placement tests to diagnose high school graduates' level of English proficiency when enrolling in most university English departments in Morocco (for further details, see Jebbour, 2019). Thus, the sample cannot be considered representative and may not provide generalizable findings for the total population of EFL learners in the same setting.

Conclusion

This study concluded that although descriptive statistics revealed that the EFL students showed relative development in C/T, and W/T as well as a breakdown in EFT/T after they completed the six self-disclosure prompts, the main findings showed that the learners' writing production in terms of CAF did not reach a statistically significant difference level, indicating that the test takers' writing development did not improve after the treatment.

The current results suggest theoretical and practical implications for language educators, teachers, researchers, and learners. At the theoretical level, this study contributes to the debate over the role of self-disclosure activities in improving certain language components by providing empirical data from the EFL context of Morocco in an effort to enrich the theory of L2 writing pedagogy. At the practical level, this study may inform educational decisions about developing study programs that consider students' personal lives in language arts classes. Furthermore, the findings from this manuscript may inform teaching practices by providing language instructors with essential implications for encouraging students to use personal information with the aim to improve certain language components. First, for teachers whose students aim to continue their studies abroad, self-disclosure writing activities may function as a rehearsal strategy to support students' oral performance by talking about their personal experiences while engaging in friendly interaction with foreigners. Second, if teachers involved in language teaching are interested in helping their students produce simple language illustrated in independent clauses, self-disclosure topics may be meaningful for meeting this objective. Accordingly, the inclusion of self-disclosure prompts in a 'paragraph writing' course could be an effective choice in departments of English studies. Third, the increased numbers of errors in the learners' written output require composition teachers to include grammar activities (e.g., text conversion) in their classes and put consistent emphasis on the use of dictionaries to improve students' syntactic knowledge and minimise the production of less accurate language forms and words in the language production stage. Finally, participation in this study may increase the participants' awareness of the benefits of reflecting on self-disclosures through diaries or journals, which might give support to their language skills.

Future directions for research are necessary to enrich discussions about the effects of self-disclosure tasks on students' writing development in the language classroom. Accordingly, conducting an experimental study where

participants complete self-disclosure and non-self-disclosure writing prompts is necessary. Such a study requires the use of the same task prompt for a pre- and post-test and a control group to better determine the effects of self-disclosure writing activities on writing development. Similarly, a longitudinal study can better increase understanding of the role of positive self-disclosures and their effect on students' language proficiency in the written mode because improving writing skills cannot be described based on completing only six self-disclosure topics in a relatively short period of time. Lastly, it would be interesting to carry out a study on students' self-disclosures in relation to one or two measures maximum (e.g., accuracy and fluency), using other aspects of the measures, such as error-free clauses per clause and the number of clauses in a text, and errors with classification.

Acknowledgment

I would like to thank the participants for devoting time to participate in this study. I would also like to thank the anonymous reviewers and editor whose comments contributed to the development of the quality of this manuscript.

References

Antaki, C., Barnes, R., & Leudar, I. (2005). Self-disclosure as a situated interactional practice. *British Journal of Social Psychology*, *44*(2), 181-199. https://doi.org/10.1348/014466604x15733

Altman, I., & Taylor, D. (1973*). Social penetration: The development of interpersonal relationships*. Holt, Rinehart & Winston.

Barrot, J. S. (2018). Using the sociocognitive-transformative approach in writing classrooms: Effects on L2 learners' writing performance. *Reading & Writing Quarterly*, *34*(2), 187-201. https://doi.org/10.1080/10573569.2017.1387631

Barrot, J., & Gabinete, M. (2019). Complexity, accuracy, and fluency in the argumentative writing of ESL and EFL learners. In P. Jordens & R. Leah (Eds.), *International review of applied linguistics in language teaching* (pp. 1-24). DeGruyter Mouton.

Bento, S. (1991). [Review of the book *Learning and instruction: European research in an international context*, Volume 2:2. Analysis of complex skills and complex knowledge domains, by H. Mandl, E. De Corte, S. N. Bennett & H. F. Friedrich]. *Computers & Education*, *17*(2), 169-171.

Berman, J. (2001). *Risky writing: Self-disclosure and self-transformation in the classroom*. University of Massachusetts Press.

Bernard, J. (2010). Motivation in foreign language learning: The relationship between classroom activities, motivation, and outcomes in a university language-learning environment [Unpublished Master's thesis]. Carnegie Mellon University. http://repository.cmu.edu

Birhan, A. (2018). Effects of mastery learning instruction on engineering students' writing skills development and motivation. *Journal of Language and Education*, *4*(4), 20-30. https://doi.org/10.17323/2411-7390-2018-4-4-20-30

Cayanus, J. L., & Martin, M. M. (2016). Teacher self-disclosure. In P.L. Witt (Ed.), *Handbook of communication science* (vol: 16. Communication and Learning, pp. 241-258). DeGruyter Mouton.

Cayanus, J. L., Martin, M. M., & Goodboy, A. K. (2009). The relationship between teacher self-disclosure and student motives to communicate. *Communication Research Reports*, *26*(2), 105-113. https://doi.org/10.1080/08824090902861523

Derlega, V. J., Anderson, S., Winstead, B. A., & Greene, K. (2010). Positive disclosure among college students: What do they talk about, to whom, and why? *Journal of Positive Psychology*, *6*(2), 119–130. https://doi.org/10.1080/17439760.2010.545430

Derlega, V. J., & Winstead, B. A., & Greene, K. (2008). Self-disclosure and starting a close relationship. In S. Sprecher, A. Wenzel & J. Harvey (Eds.), *Handbook of relationship beginnings* (pp. 153-174). Psychology Press.

Dörnyei, Z. (1994). Motivation and motivating in the foreign language classroom. Modern Language Journal, 78(3), 273-284. http://www.jstor.org/stable/330107?origin=JSTOR-pdf

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, *18*(3), 299-323. https://doi.org/10.1017/S0272263100015047

Graham, S., Harris, K. R., & Santangelo, T. (2015). Research-based writing practices and the common core: Meta-analysis and meta-synthesis. *The Elementary School Journal*, *115*(4), 498–522. https://doi.org/10.1086/681964

Guilloteaux, M. J., & Dörnyei, Z. (2008). Motivating language learners: A classroom-oriented investigations of the effects of motivational strategies on student motivation. *TESOL Quarterly, 42*(1), 55-78. https://doi.org/10.1002/j.1545-7249.2008.tb00207.x

Hoadley-Maidment, E. (2000). From personal experience to reflective practitioner: Academic literacies and professional education. In M. Lea & B. Stierer (Eds.), *Student writing in higher education*: *New contexts* (pp. 165–78). Society for Research into Higher Education & Open University Press.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, *30*(4), 461-473. https://doi.org/10.1093/applin/amp048

Iida, A. (2016). Poetic identity in second language writing: Exploring an EFL learner's study abroad experience. *Eurasian Journal of Applied Linguistics, 2*(1), 1-14. https://www.ejal.info/index.php/ejal/article/view/131

Jebbour, M. (2020). Developing a student self-disclosure measure: A pilot study. *Journal of English Language Teaching and Linguistics, 5*(1), 17-27. http://dx.doi.org/10.21462/jeltl.v5i1.374

Jebbour, M. (2019). English language teaching in Morocco: A focus on the English department. *The Journal of North African Studies*. https://doi.org/10.1080/13629387.2019.1681267

Jebbour, M. (2018). University students' perceptions of the effects of teacher self-disclosure in the English language classroom. *Journal of English Language Teaching and Linguistics*, *3*(3), 275-285. http://dx.doi.org/10.21462/jeltl.v3i3.166

Jebbour, M., & Mouaid, F. (2019). The impact of teacher self-disclosure on student participation in the university English language classroom. *International Journal of Teaching and Learning in Higher Education*, *31*(3), 424-436.

Jourard, S. (1971). *The transparent self.* Van Nostrand.

Komalasari, A. (2013). Employing experiential learning to teach writing for English as a foreign language learners through a reflection project. *International Journal on Education, 1*(1), 24-30.

Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences*, *14,* 317–324. https://doi.org/10.1016/ j.tics.2010.05.002

Lam, R. (2015). Convergence and Divergence of Process and Portfolio Approaches to L2 Writing Instruction: Issues and Implications. *RELC Journal*, *46*(3), 293-308. https://doi.org/10.1177/0033688215597119

Lannutti, P. J., & Strauman, E. C. (2006). Classroom communication: The influence of instructor self-disclosure on student evaluations. *Communication Quarterly 54*(1), 89-99. https://doi.org/10.1080/01463370500270496

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. Routledge.

Liao, C. (2016). Enhancing the grammatical accuracy of EFL writing by using an AWE-assisted process approach. *System, 62,* 77-92. https://doi.org/10.1016/j.system.2016.02.007

McDonough, K., & Fuentes, C. G. (2015). The effect of writing task and task conditions on Colombian EFL learners' language use. *TESL Canada Journal*, *32*(2), 67-79.

Moon, J. (2004). *A handbook of reflective and experiential learning*. Routledge Falmer.

Myers, S. (1998). Students' self-disclosure in the college classroom. *Psychological Reports, 83*(7), 1067-1070. https://doi.org/10.2466%2Fpr0.1998.83.3.1067

Norris, J., & Manchón, R. M. (2012). Investigating L2 writing development from multiple perspectives: Issues in theory and research. In R.M. Manchón (Ed.), *L2 writing development: Multiple perspectives* (pp. 221–244). de Gruyter Mouton.

Peng, J., & Wang, Ch. (2018). Effect of the linguistic complexity of the input text on alignment, writing fluency, and writing accuracy in the continuation task. *Language teaching Research, 00*(0), 1-18. https://doi.org/10.1177/1362168818783341

Polio, C. (2001). Research methodology in second language writing: The case of text-based studies. T. Silva & P. Matsuda. (Eds.), *On second language writing* (pp. 91-116). Erlbaum.

Polio, C., & Park, J. H. (2016). Language development in second language writing. In R. M. Manchon´ & P. Matsuda (Eds.), *Handbook of second and foreign language writing* (pp. 287–306). Routledge.

Polio, C., & Shea, M. C. (2014). An investigation into current measures of linguistic accuracy in second language writing research. *Journal of Second Language Writing, 26*, 10–27. https://doi.org/10.1016/J.JSLW.2014.09.003

Rosmawati (2014). Second language development dynamics: How dynamic systems theory accounts for issues in second language learning. *The Australian Educational and Development Psychologist*, *31*(1), 66-80. https://doi.org/10.1017/edp.2013.22

Serag, A. (2011). Self-disclosure in EFL writing by computers. In A. Stewart (Ed.), *JALT2010 Conference Proceedings* (pp. 551-564). JALT.

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, *1*(3), 185-211. https://doi.org/10.1177%2F136216889700100302

Stotsky, S. (1995). The uses and limitations of personal or personalized writing in writing theory, research, and instruction. *Reading Research Quarterly*, *30*(4), 758-776. https://doi.org/10.2307/748197

van Geert, P., & van Dijk, M. (2003). Ambiguity in child language: The problem of inter-observer reliability in ambiguous observation data. *First Language*, *23*(3), 259–284. https://doi.org/10.1177/01427237030233001

Van Waes, L., & Leijten, M. (2015). Fluency in writing: A multidimensional perspective on writing fluency applied to L1 and L2. Computers and Composition, *38*, 79-95. https://doi.org/10.1016/j.compcom.2015.09.012

Wakefield, C.R. (2009). *Self-disclosure and cohesion in the religious classroom* [Unpublished Master's thesis]. Southern Utah University. https://www.suu.edu

West, R., & Turner, L. (2010). *Introducing communication theory: Analysis and application* (4th ed.). McGraw Hill.

Wheeless, L. R., & Grotz, J. (1976). Conceptualization and measurement of reported self- disclosure. *Human Communication Research*, *2*(4), 338-346. https://doi.org/10.1111/j.1468-2958.1976.tb00494.x

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

Zhang, S., Shi, Q., Tonelson, S., & Robinson, J. (2009). Preservice and inservice teachers' perceptions of appropriateness of teacher self-disclosure. *Teaching and Teacher Education*, *25*(8), 1117-1124. http://dx.doi.org/10.1016/j.tate.2008.07.011

Appendix A

A writing post-test sample

It is very necessary for English students to get a close and a direct contact with an English speaking country like England, Australia, and the United States of America. Because of that, the university must organize a study trip to one of those countries. The main reason that makes this trip very important is that students will be face to face with the English culture. There is no way to learn a language without understanding its culture. By doing so, students will have a better awareness and consciousness about what they're thinking, saying, and writing. Another good reason that makes the study trip a great idea is that students will use it to practice their language skills, especially the skills of understanding and speaking. Additionally, it will make them curious about both the culture and the language, which will make them highly motivated, passionate, and excited for their education. A study trip is always a positive thing to do. It has a lot of benefits that will make students do better in their studies, also their job career.

Appendix B

T-units in the writing post-test sample

1. The main reason that makes this trip very important (DC) is that students will be face to face with the English culture (IC).
2. Students will have a better awareness and consciousness (IC) about what they're thinking, saying and writing (DC).
3. It will make them curious about both the culture and the language (IC), which will make them highly motivated, passionate, and excited for their education (DC).
4. It has a lot of benefits (IC) that will make students do better in their studies, also job career (DC).
5. Another good reason that makes the study trip a great idea (DC) is that students will use it to practice their language skills (IC).

Errors in the writing post-test sample

1. consciousness about
2. excited for
3. also their job career

Calculation of CAF measures of the writing post-test sample

Parameter and calculation	Coding
C/T = 14/5	2.8
EFT/T = 2/5	0.4
W/T = 100/5	20

National Research University Higher School of Economics
Journal of Language & Education Volume 7, Issue 1, 2021

Lambovska, M., & Todorova, D. (2021). 'Publish and Flourish' instead of 'Publish or Perish': A Motivation Model for Top-quality Publications. *Journal of Language and Education, 7*(1), 141-155. https://doi.org/10.17323/jle.2021.11522

'Publish and Flourish' instead of 'Publish or Perish': A Motivation Model for Top-quality Publications

Maya Lambovska[1], Daniela Todorova[2]

[1] University of National and World Economy

[2] Todor Kableshkov University of Transport

Correspondence concerning this article should be addressed to Maya Lambovska, University of National and World Economy, complex Students' Town "Hristo Botev", 1700 Sofia, Bulgaria.
E-mails: *mlambovska@abv.bg*

Although the 'publish-or-perish' principle has spread globally, many authors believe that it is a negative reinforcer (motivator) and harmful. With this paper, we have tried to help overcome the growing pressure of negative reinforcers on researchers. The paper aimed to propose a model for factors influencing researchers to publish in WoS/Scopus journals, based mainly on positive reinforcement and a combination of concepts including theories of control, management, stakeholders, and psychology. The model was intended for Bulgarian universities. It covered 17 motivational drivers and 29 potential features of internal university stakeholders directly involved in the topic. Factor ranking was not incorporated in the model. The research methodology covered the methods of expert evaluation, analysis/synthesis, induction/deduction, and the toolkit consisted of a comprehensive survey and Kendall's rank concordance coefficient. The model was implemented at a Bulgarian state university. The empirical study was conducted among 120 researchers. It resulted in factor rankings by university internal stakeholders. The highest-ranked motivational driver was reputation, and the lowest-ranked was the publish-or-perish pressure reducing. The highest-ranked potential features were university prestige and potential and support for promotion. We believe that this model contributes to the theory of behaviour control. The model will also improve university research management by enriching its tools.

Keywords: higher education, research management, academic publishing, WoS & Scopus journals, publish-or-perish, motivation, potential, stakeholders

Introduction

Research plays a crucial role in social progress (Daumiller & Dresel, 2020). In turn, the fate of research depends on its quality and successfully being published (Bedeian, Van Fleet, & Hyman, 2009). "Research quality is considered a theoretical construct that encompasses a multitude of different facets" (Daumiller, Siegel, & Dresel, 2019, p. 4). In our view, one key criterion for the quality of research is publishing in prestigious academic journals. It contributes to the dissemination and practical application of the most valuable ideas, and, as a result, increases the public benefit of research. Additionally, the current approach to universities as stakeholder organizations (Osterloh & Frey, 2014) predetermines the importance of such publishing for all stakeholders. For researchers, there are influences on their tenure, career advancement, professional recognition, and rewards (Miller, Taylor, & Bedeian, 2011). For universities, it increases institutional rankings, their funding, and quality/accreditation scores (Blaginin, Volkova, & Strielkowski, 2019; Karimi & Asadnia, 2020). For society, "high-quality publications remain the primary tool for describing national and individual contributions to science and society" (Lambovska & Yordanov, 2020, p. 188). Thus, publishing in prestigious journals contributes to the prosperity of all stakeholders.

Two questions arise about publishing in prestigious journals: 1. Which academic journals are prestigious? 2. What motivates researches to publish in prestigious journals?

Concerning prestigious journals, most authors identify them with those indexed by Web of Science (WoS)/ Scopus (Blaginin et al., 2019; Raitskaya & Tikhonova, 2020a). We share this view, as these databases "offer

some objectivity" (Blaginin et al., 2019, p. 361) by evaluating journals "based on peer review, citations, h-index and impact factor" (Osterloh & Frey, 2014, p. 83).

Concerning the second question, a globally spread approach is one based on the 'publish-or-perish' principle as a negative motivator (reinforcer). Following it, researchers "are forced to publish as much as possible" (Binswanger, 2015, p. 19) "mainly in international journals" (Raitskaya & Tikhonova, 2020a, p. 4). In most cases, this principle is classified as a negative reinforcer due to its predominantly aggressive application, which puts strong pressure on researchers. Many authors consider it harmful, "even dangerous" (Blaginin et al., 2019, p. 357), due to its negative effects, including higher stress, nonsense research, intense competition among peers, an increased market for predatory journals, communication problems of non-native writing, wider national scientific communities communicating only in native languages, and the promotion of 'cartels' among scientific communities (Beavitt & Popova, 2020; Behl, Chavan, Dutta, & Sheorey, 2020; Hung & Baranovskaya, 2020; Miller et al., 2011; Raitskaya & Tikhonova, 2020b). Also, "the 'publish-or-perish' principle crowds out the intrinsic motivation of researchers, which is a necessary condition for true excellence" (Binswanger, 2015, p. 30). We fully support these views. We also consider the 'publish-or-perish' principle to be a negative reinforcer. It should not be neglected, as the literature brings it out as an important motivator for researchers.

Another approach to motivate researchers is by combining positive and negative reinforcers. It is applied in several of the models we found in the WoS/Scopus literature on this topic. Two models (Daumiller & Dresel, 2020; Karimi & Asadnia, 2020) are analytical. They cover motivational drivers derived from one or two theories. Results of their testing are country- and/or community-specific. The third model (Lambovska & Yordanov, 2020) is based on a systematic literature review. It is general and descriptive. None of the models classified the factors influencing researchers by stakeholders. Our literature review led us to three main conclusions. First, "few studies have explicitly examined researchers' motivations" (Daumiller & Dresel, 2020, p. 1). Second, this is a novel and relatively under-investigated topic. Third, there are no models in the scientific literature to influence researchers to publish in prestigious journals, which are comprehensive enough, theoretically sound due to an appropriate combination of concepts from social theories, and take into account stakeholder motivation and potential.

With this paper, we would attempt to bridge these gaps. Following reinforcement theory (Hamner, Luthans, & Krietner, 2015) and our belief in the higher efficiency of positive motivation (Stoyanov, 2014), we suggest an approach to influence researchers based primarily on positive reinforcers. Another novelty of this approach is that we look at the influence on researchers in a complex way - as a function of the stakeholder motivation and potential. In this regard, a set of concepts from theories of control, stakeholders, organizational behaviour, psychology, as well as general and performance management are at the core of our approach.

The *research aim* of this paper is to propose a model for factors influencing researchers to publish in WoS/Scopus journals, based mainly on positive reinforcement and a combination of concepts from social theories. It is intended for Bulgarian universities. The model covers only factors that apply to internal stakeholders. By factors, we mean motivational drivers and potential features.

We raise three research questions (RQs):

- RQ1: Which drivers and features typify, respectively, the motivation and potential of stakeholders to contribute to publications published in WoS/Scopus journals?
- RQ2: What are the features of the models for influencing researchers to publish in WoS/Scopus journals available in the scientific literature?
- RQ3: In the current Bulgarian context, what set of stakeholder motivational drivers and potential features can be proposed to increase publications in WoS/Scopus journals?

The *research methodology* we applied covers the methods of expert evaluations, analysis/synthesis, induction/deduction, as well as the comprehensive survey of the entire population and Kendall's rank concordance coefficient.

We believe that this model will contribute to the theory of behaviour control, in particular to the control of researchers' behaviour. It will enrich the tools of university research management and make it possible to

decrease the pressure and effects of negative reinforcers on researchers. By creating it, we try to support current efforts to encourage Bulgarian researchers to publish in prestigious journals. The model can also be used by other research communities working in a similar context including reputational systems (Vogel & Hattke, 2018), educational regulations (Vorontsova, Shvindina, Mayboroda, Mishenina, & Heiets, 2020), and publication, organizational, and national culture (Ghinea, Mihaylova, & Papazov, 2015; Raitskaya & Tikhonova, 2020c).

We pursue the following *strategy* to achieve our research aim. In the first part of the paper, we conduct a systematic review of the WoS & Scopus literature. Thus, we identify achievements and gaps in the scientific literature and systematize them, answering RQ1 and RQ2. These results are the starting point for substantiating and creating the model we are proposing. We chose the concepts underlying the model based on gaps in the literature. We also included some motivational drivers and potential features described in the answer to RQ1 (Figure 1). In the second part, we outline the theoretical framework and context of the proposed model. Then we clarify the nature and specifics of the model. Thus, we answer RQ3. In the third part, we present the results of the model after it was implemented at a Bulgarian university.

Literature Review

Papers found in WoS & Scopus before November 2020 are reviewed here, using a structured approach (Kitchenham et al., 2009), methods of content analysis, and deduction. We searched for papers through five combinations of keywords and derivatives, including 'stakeholder', 'influence', 'motivation', 'potential', 'publish-or-perish', 'academic', 'researcher', 'publication', 'journal', 'Scopus', and 'WoS'. Initially, 623 papers were found. Of these, only 15 papers, which we will refer to as the 'sample', were eligible for this study. Based on the sample analysis (Figure 1), we answered RQ1 and RQ2.

In response to RQ1, we discovered and systematized motivational drivers and potential features of stakeholders to contribute to publications published in WoS/Scopus journals (Figure 1). Then we ranked the drivers and features by frequency of occurrence. We did not classify them by stakeholders, as the papers lacked such a grouping.

Regarding motivational drivers, we found 14 main ones (Figure 1a). We grouped them as either extrinsic or intrinsic. The results showed that extrinsic drivers were cited more compared to intrinsic ones. The most cited drivers were collaboration/co-authorship (1st); promotion/tenure and job mobility (2nd); financial assets and publish-or-perishing pressure reducing (3rd); researcher's reputation (5th). Competition and university reputation were cited the least, 13th and 14th.

Regarding features of the stakeholder potential, we found 15 main ones (Figure 1b). According to the results, the most cited features were pressure (publish-or-perish and cultural) (1st); collaboration/co-authorship potential and material support (2nd); promotion/tenure (4th); research funding (5th). The least cited were autonomy (14th) and other types of publications (15th).

In response to RQ2, we discovered three models for influencing researchers available in the literature. We analysed them and systematized their features (Table 1).

The first model (Daumiller & Dresel, 2020) is based on achievement goal theory. Accordingly, Daumiller and Dresel (2020) developed a model of ten classes of achievement goals to describe the researchers' motivation, including task approach and avoidance, learning approach and avoidance, appearance approach and avoidance, normative approach and avoidance, work avoidance, and relational. It was tested among 824 German university researchers. They ranked mastery (task and learning standards) and performance (appearance and normative strivings) the highest. Relational and work avoidance goals were also mentioned. The highest-ranked results are of mixed nature (positive and negative drivers). They are integrated into row 7 of Figure 1. Our results partially confirmed those of Daumiller and Dresel. In Figure 1a, the researcher's reputation and personal development ranked 5th and 7th, respectively. We found five deficiencies of Daumiller's and Dresel's model. First, it does not literally cover our topic, as it addresses the researcher motivation to research but not specifically to publish in WoS/Scopus journals. Second, the model is based on a single theory. This one-sided approach does not allow a variety of motivational drivers to be taken into account. Third, the model does not classify

Figure 1

Results of the literature sample analysis on the factors influencing researchers

a) Motivational drivers to publish in WoS/Scopus journals

	Reference	Motivational drivers – type and frequency of occurrence (times)													
		Extrinctive									Intrinsic				
		Collaboration and co-authorship	Competition	Financial assets	Funding of research	Promotion/tenure, job mobility	Publish-or-perish pressure reducing	Researcher's reputation	Working conditions (incl. autonomy)	University reputation	Contribution to science and	Contribution to society	Challenging/creative work	Job satisfaction	Personal development
1	Alfonso (2019)			1											
2	Banks (2018)														
3	Barreto, Roa and Velasquez (2019)			1			1		1			1			
4	Behl, Chavan, Dutta, and Sheorey	1		1		1	1		1						
5	Binswanger (2015)						1								
6	Blaginin, Volkova and Strielkowski				1		1								
7	Daumiller and Dresel (2020)						1	1							1
8	Johnston (2016)										1				1
9	Karimi and Asadnia (2020)					1	1	1							
10	Lambovska and Yordanov (2020)	11	3	8	8	8	4	7	4	2	5	8	5	5	5
11	Miller, Taylor and Bedeian (2011)			1		1	1	1			1				
12	Raitskaya and Tikhonova (2020a)	1				1	1								1
13	Richard, Plimmer, Fam and Campbell							1	1					1	
14	Romani-Dias, Carneiro and Barbosa	1			1										
15	Vogel and Hattke (2018)	1				1			1						1
	Rank	1	13	3	6	2	3	5	9	14	10	7	12	11	7

b) Features of the stakeholder potential to influence researchers to publish in WoS/Scopus journals

	Reference	Potential features														
		Affinity for research/scientific	Authonomy	Collaboration, co-authorship potential	Command of foreign languages	Funding of research	Intangible support and awards	Intellectual abilities	Material support/incentives	Methodological training	Other types of publications	Pressure (cultural and publish-or-	Professionalization of scientific activity	Promotion, tenure, carier abroad	Time	University potential and prestige
1	Alfonso (2019)				1				1	1						1
2	Banks (2018)				1							1	1			
3	Barreto, Roa and Velasquez (2019)					1	1		1	1						
4	Behl, Chavan, Dutta, and Sheorey			1		1	1		1			1		1		
5	Binswanger (2015)								1			1		1		
6	Blaginin, Volkova and Strielkowski											1				
7	Daumiller and Dresel (2020)											1	1		1	
8	Johnston (2016)		1													
9	Karimi and Asadnia (2020)											1		1		
10	Lambovska and Yordanov (2020)	3		13	4	9	5	3	10	5	1	11		9	9	5
11	Miller, Taylor and Bedeian (2011)								1			1		1		
12	Raitskaya and Tikhonova (2020a)			1	1		1			1		1	1	1		
13	Richard, Plimmer, Fam and Campbell						1			1						
14	Romani-Dias, Carneiro and Barbosa			1	1				1	1			1		1	
15	Vogel and Hattke (2018)		1			1	1									
	Rank	12	14	2	9	5	7	12	2	7	15	1	11	4	6	10

motivational drivers by stakeholder. Fourth, the results of its testing are country-specific. Fifth, the model uses complex formulations of drivers and is difficult to apply in practice.

The second model (Karimi & Asadnia, 2020) is based on Dornyei's motivational self-system theory (2009) and Markus and Nurius's theory of possible selves (1986). It clarifies how researchers' possible selves (the ideal, the ought-to, and the feared) affect their productivity. It was tested for Iranian applied linguistic researchers. Empirically, four "main triggers to publish in top-tier journals were identified: desires to gain visibility in the research community, to become prolific researchers, to preserve their international researcher positions, fear of the publish-or-perish principle" (Karimi & Asadnia, 2020, p. 215). These results are included in the literature sample (row 9 in Figure 1). They are also of mixed nature and are close to ours. As shown in Figure 1a, promotion and publish-or-perish pressure reduction are ranked 2nd and 3rd, respectively. In our view, the model has three deficiencies. First, the main drawback is that researchers' motivation is examined in the context of a single psychological aspect. Second, the results of its testing are community- and country-specific. Third, the model refers to a single stakeholder - the researcher.

Table 1
Features of the models for influencing researchers found in WoS & Scopus

Model	Daumiller & Dresel, 2020	Karimi & Asadnia, 2020	Lambovska & Yordanov, 2020
Type	analytical	analytical	descriptive
Theory (number)	yes (1)	yes (2)	no (0)
Nature (general/specific)	general	specific	general
Nature of test results	specific	specific	-
Motivational drivers covered	10	14	14
Potential features covered	0	0	12
Type of highly ranked drivers (row in Figure 1)	mixed (7)	mixed (9)	positive (10)
Classification by stakeholders	no	no	no

The third model (Lambovska & Yordanov, 2020) is general and descriptive, as it is derived from a systematic literature review. The model covers two types of factors: motivational drivers of researchers and features of the potential to influence researchers. According to its results, the most cited motivational drivers were collaboration, research funding, financial assets, promotion, contribution to society, and researcher's reputation. The most cited potential features were collaboration, material support, research funding, promotion, and time. The results are integrated into our literature sample (row 10 in Figure 1). They largely determine our results. We found four deficiencies of this model. First, there is no theory at the core of the model. Second, it is not tested. Third, the drivers/features are not classified by stakeholders. Fourth, the 'publish-or-perish' principle is not a search term, and, thus, its representation is artificially restricted.

The main results of our literature review could be summarized as follows:

- Very few papers eligible for this study were found in the WoS & Scopus literature. Most of them identify motivational drivers and potential features in the context of other topics.
- Only three models for influencing researchers were found: two analytical and one descriptive.
- The analytical models are based on one or two theories and tested for a specific community/country. One analytical model is community- and country-specific.
- Only the descriptive model covers the features of the stakeholders' potential to influence researchers.
- The maximum number of motivational drivers covered by the models is 14.
- All models lack systematization by stakeholders.

Solution

Theoretical Framework

Concepts from theories of control, stakeholders, organizational behaviour, psychology, general and performance management are at the root of the proposed model.

The notion of influence on behaviour and its interpretation in various scientific theories is of paramount importance for this model. In *general management*, "the ability to influence behaviour is defined as power" (Yordanov, 1997, p. 32). In *control theory*, influence is known as control influence. It refers to the concept of control as a critical social attitude based on the 'domination-subordination' dependence. This attitude is "determined by the differences in the socio-economic potentials of the subjects of control (stakeholders) to influence" (Simeonov, 1997, p. 18). The control influence is embedded in the *concept of control relations*. In it (Simeonov, 1997), control influence (CI) is mathematically described (Formula 1) as a product of the stakeholder's motivation (M) and its potential to influence (PI). Control influences of the stakeholders on a target result form the so-called 'picture of control relations' of the organization at some point (Bourne & Walker, 2005). It shows the stakeholders who ultimately control the achievement of the target result at this point. This 'picture' is a starting point for redesigning the control relations of the organization, aimed at

achieving its goals. The concept of control relations underlies our model. The model also applies the *principle of self-control* to researchers as an object of control.

$$CI=M\times PI \quad (1)$$

The notion of human resource performance is also used here. *Performance management* establishes seven primary criteria of the organizations' performance (Kirova, 2010; Sink, 1986): effectiveness, efficiency, quality, profitability, productivity, quality of work-life, and innovation. Of these, effectiveness, "degree of goals-achieving" (Cizmas, Feder, Maticiuc, & Vlad-Anghel, 2020, p. 8), and productivity, "output/results per resource" (Sink, 1985, p. 20), directly address our topic. We are not interested in productivity, as it is based on effectiveness and is community/country-specific in this context. Thus, human resource performance refers primarily to *human resource effectiveness* here and it is subject matter of both psychology and organizational behaviour. In *psychology*, "human resource effectiveness is described mathematically as a product between motivation and ability" (Sink, 1985, p. 303). According to *Vroom's expectancy theory* (1985), the result of multiplying ability and motivation is performance. This wording is very close to the definition of control influence in control theory (Formula 1).

The idea of increasing effectiveness through behaviour control based on motivation leads us to apply *reinforcement theory*. This is a behavioural theory that uses "the principle of learning through reinforcement to influence and modify human behaviour to desired ends" (Hamner et al., 2015, p. 115). It identifies four types of strategies: positive and negative (avoidance learning) reinforcement, extinction, and punishment. "The first two serve to strengthen the desired behaviour, and the others to weaken undesired behaviour" (Hamner et al., 2015, p. 116). In our model, only behaviour-strengthening strategies are applied. "Reinforcement is conceptually related to motivation, but it is something different" (Luthans, Luthans, & Luthans, 2015, p. 133). Motivation is the internal explanation of behaviour, and reinforcement is external (Hamner et al., 2015).

Behaviourists view *motivation* as an important hypothetical construct to explain behaviour. Luthans et al. (2015, p. 132) define it as "a basic psychological process that starts with a need that activates a behaviour or a drive that is aimed at a goal or incentive". We share this opinion. Here we also apply the concepts of *Murray's personality theory, Vroom's expectancy theory,* and *Ryan and Deci's types of motivation*. We support Murray's (1936) idea of the lack of a single hierarchy of needs for all individuals. Therefore, we do not introduce a hierarchy of motivational drivers in the model and do not classify them into primary/secondary or physiological/psychological. We base our model on Vroom's concept of performance (here called 'control influence', Formula 1). By motivational drivers, we mean possible outcomes, following Vroom's terminology (1985). Depending on the way they are generated, motivational drivers are classified as either intrinsic ("internally generated, driven by internal rewards") or extrinsic ("externally generated, dependent on external rewards") (Luthans et al., 2015, p. 134). Ryan and Deci (2000, p. 54) define "intrinsic motives as reflections of the natural human propensity to learn and assimilate, based on the needs of competence, autonomy and relatedness". They believe that external motives reflect either external governance or true self-regulation (Ryan & Deci, 2000). We rooted these views in the model, also classifying motivational drivers as either intrinsic or extrinsic (Figure 2).

The control relations in this model are considered according to the context *of stakeholder theory*. Here, both the object and the subjects of control are stakeholders. Also, all of the factors influencing researchers are classified by the stakeholders in the model (Figure 2). According to this theory, "organizations must bear in mind the different perspectives and expectations of a variety of constituents, called stakeholders, who can influence the outcome of the organization" (Freeman, 2015, p. 32). The influence of the stakeholders depends to a large extent on their potential. *Stakeholder potential* is defined as "the strength and ability to influence others" (Simeonov, 1997, p. 16). It consists of three groups of elements (Simeonov, 1997; Stoyanov, 2012): personal strength (creativity, culture, expertise, intellect, material status, skills, etc.), position power (formal authority; control over rewards, punishments, information, etc.) and political power (coalitions, institutionalization, co-option, etc.). Stakeholders are usually divided into internal and external stakeholders. "Academic staff, non-academic staff and students are the main internal stakeholders at the university" (Turan, Cetinkaya, & Ustun, 2016, p. 745). Here we are interested in the academic staff divided into three stakeholders - researcher, academic department head, and top university management.

Methods

The methods of analysis/synthesis and induction/deduction were used to create this model.

The Bulgarian Context

The importance of the Bulgarian researchers’ publications in WoS/Scopus journals stems mainly from the current regulations in higher education. According to the requirements of the Bulgarian Ministry of Education and Science (MES RB, 2018), these publications are an indicator for calculating research grants for state universities. They are also a very important indicator of the Bulgarian University Ranking System (MES RB, 2010). This ranking is considered an official state assessment of the quality of university education and is, therefore, a key factor in student influx. Additionally, these publications are a key prerequisite for a high university accreditation score, which is another indicator of the Ranking System. Moreover, the per-student subsidy of state universities is based on their ranks and accreditation scores.

Other important features of the current context in which Bulgarian researchers work should also be noted. As for finances, their salaries are fixed and relatively low compared to researchers from other European countries. Publications in WoS/Scopus journals are not a prerequisite for a higher salary. However, many state universities pay financial bonuses for them. As for work/promotion, Bulgarian researchers have permanent employment contracts that do not depend on these publications. Most Bulgarian state universities require such publications only in competitions for the academic position of professor. This is not the case for most private universities. Also, many researchers teach at several universities or have other employment contracts. They research either for promotion or for prestige. As for awareness, many researchers are not informed about the normative significance of top achievements, as they have not had to comply with it so far. As for the scientific position of Bulgaria, its current ranking in WoS/Scopus is not high. In 2019, Bulgaria ranked 64th in Scopus and 63rd in WoS by citable documents, including articles.

The current Bulgarian context gives us reason to conclude: 1. Publications in WoS/Scopus journals are very important for the survival of Bulgarian state universities and are a prerequisite for researchers’ reputations. 2. These publications are considered “a tool to control the relevance of public spending” on science at state universities (Yordanov, 1999, p. 40). 3. The ‘publish-or-perish’ principle puts strong pressure on Bulgarian university management, but not on researchers.

Model

In response to RQ3, we created a model for factors influencing researchers to publish in WoS/Scopus journals (Figure 2). It is based on the following *limitations*:

1. Target community - Bulgarian researchers
2. The model covers internal university stakeholders directly involved in this topic.
3. Factor ranking is not incorporated in the model.

The main *peculiarities* of the model are as follows:

- It is country-specific (limitation 1).
- Due to limitation 3, it applies to the Bulgarian research community as a whole.
- It covers 46 factors influencing researchers. They are classified into motivational drivers (17) and potential features (29). Vroom’s expectancy theory and the notion of control influence (Formula 1) underlie this classification. Per Ryan and Deci’s concepts (2000), the motivational drivers are grouped as either intrinsic or extrinsic. Based on the theory, the potential features are divided into three groups: personal strength, position, and political power.
- The factors are systematized by stakeholders based on stakeholder theory and our sample.
- It addresses academic staff (limitation 2) divided into three stakeholders - researcher, academic department head, and top university management (Rector, Vice-rectors, Deans).
- All but one of the motivational drivers are positive reinforcers in line with behaviour strengthening strategies of reinforcement theory.

Concerning *intrinsic* motivational *drivers*, job satisfaction is the only one enrolled in the model for all stakeholders. In our view, it is currently the key intrinsic driver for Bulgarian researchers.

For the *researcher stakeholder*, the model covers five *extrinsic drivers*. Four drivers match those ranked 2nd to 5th in the results of the literature sample analysis (Figure 1a). The rationale of the drivers:

- Financial benefits: The driver is 1. highly ranked in the sample (3rd, Figure 1a); 2. "a major positive reinforcer of work motivation" (Luthans et al., 2015, p. 352); 3. very important for Bulgarian researchers due to their low pay. 4. Bulgarian state universities reward researchers with financial bonuses for articles in WoS/Scopus journals.
- Promotion to academic and administrative positions: The driver is 1. highly ranked in the sample (2nd, Figure 1a); 2. a major positive reinforcer (Luthans et al., 2015). 3. Currently, the requirement to publish in WoS/Scopus journals is one of the most difficult for acquiring a professorial academic position in most Bulgarian state universities. 4. The promotion results in higher pay and raises the researcher's reputation.
- Researcher's reputation: The driver is 1. ranked 5th in the sample (Figure 1a). 2. Reputation is considered "an effective social positive reinforcer" due to its nature of "informal feedback and achievement motive" (Luthans et al., 2015, pp. 134, 352).
- Publish-or-perish pressure reducing: This is the only negative reinforcer in the model. The publish-or-perish principle is 1. highly ranked in the sample (3rd, Figure 1a); 2. already globally-spread (Raitskaya & Tikhonova, 2020b). 3. Although it is still not applied fiercely in Bulgaria, its importance is constantly growing due to higher education regulations.
- Participation in groups/juries, projects: It combines the drivers of collaboration and research funding, which are highly ranked in the sample (1st and 6th, Figure 1a). Although "collaboration is an important affiliation motive" according to behaviourists (Luthans et al., 2015, p. 134) and the most cited driver found in the sample (Figure 1a), it is not a separate driver in this model. There are two reasons for this. First, cooperation is not currently vital for Bulgarian researchers due to the non-application of the 'publish-or-perish' principle for their promotion and periodic attestation at most universities. Second, we believe that cooperation is a potential feature in this context, but not a motivational driver.

For the *department head stakeholder*, the model covers four *extrinsic drivers*. Two (recognition by top university management and department reputation) are feedback. They are considered prerequisites for the third driver - promotion to administrative positions. Publish-or-perishing pressure reducing is the fourth driver. The importance of these drivers is explained above.

For the *top university management stakeholder*, five *extrinsic drivers* are included in the model. Three are similar to the previous stakeholder: promotion to administrative positions, university reputation, and recognition by academic/administrative staff. The others are extremely important to Bulgarian universities, especially during the COVID-19 crisis. These are 'increases in university accreditation/rating scores' and 'increases in university subsidies (for state universities only)'. Student influx depends on the former indicator and the finances of state universities on the latter.

Features of stakeholder potential are identified based on our experience as university researchers and the literature review results (Figure 1b). In the model, the publish-or-perish pressure is not a potential feature, as we consider it is a motivational driver.

For the *researcher stakeholder*, the model covers nine potential features. The features of personal strength are an affinity for research/scientific work, command of foreign languages, and intellectual abilities. In our view, the researcher's position power is best described by two potential features: promotion to academic positions and time. Although these features are not highly ranked in the literature sample (4th and 6th, Figure 1b), they provide great opportunities for researchers, especially in the current Bulgarian context. The researcher's political power group is formed by four potential features. Of these, three are effects connected to researchers' collaborations (2nd, Figure 1b), including opportunities for scientific trips abroad, participation in research projects, scientific contacts abroad. We recorded them as separate features due to their difference in nature and significance for researchers. The fourth feature is university potential and prestige (10th, Figure 1b).

Figure 2
Authors' model for factors influencing researchers to publish in WoS/Scopus journals

The *department head* and *top university management* stakeholders have similar potential features in the model. We believe that personal strength features are not particularly important for the control influences of these stakeholders. That is why we did not include them in the model. The opposite is true for the features of the other two groups. In the position power group of these stakeholders, there are five features that are the same. They are related to support (for promotion to academic degrees/position - 4th in Figure 1b and for department scientific activity) and nominations (for intangible awards - 7th in Figure 1b and for scientific/editorial boards). In this group, there is another feature for the department head (material support/incentives, 2nd in Figure 1b), and two other features for the top university management - initiation of material awards for scientific achievements and nomination for administrative positions. They are included in the model for two reasons: 1. the high rank in the sample results and 2. the current situation in which Bulgarian researchers work. In the political power group, there is a coincidence of three features: opportunities for scientific trips abroad, scientific contacts abroad, and support for joining large-scale research projects. To take into account the department head's potential to support subordinates in their research, the collaboration feature is added to the political power group of this stakeholder.

Implementation of the Model

Here we present the results from the model implementation at the Todor Kableshkov Transport University (TKTU), Sofia in 2020 (Figure 3). TKTU is a Bulgarian state university. In 2019, it ranked 31st out of 41 Bulgarian universities in terms of articles indexed by WoS (MES RB, 2010).

Research Design

The method of expert evaluations and a comprehensive study of the entire population were used to test the model. Respondents had to rank motivational drivers and potential features in descending order of importance. The population of TKTU researchers covered 120 people, including 24.17% professors, 33.33% associate professors, and 42.5% assistants. It was distributed by *internal stakeholders* as follows: researchers – 101, department heads – 12, top university management - 7. Researcher stakeholders were divided into five groups (General Engineering, Geodesy & Construction, Natural Sciences, Social Sciences, and Technical Sciences), based on the Scopus and WoS subject areas in which TKTU researchers published most between 2014 and 2019.

Kendall's rank concordance coefficient (W) was applied as a tool to measure stakeholder consensus. This coefficient is very often used to measure the consistency of expert evaluations in rankings. Its value varies in the range from 0 to 1 ($W \in [0,1]$). If its value is greater than or equal to .5 ($W \geq .5$), it is assumed that there is sufficient consensus (David, Kendall, & Stuart, 1951). Here we adopted the following scale for concordance assessment (Table 2): low – $W \in (0, .3)$, moderate - $W \in [.3, .5)$, significant - $W \in [.5, .7)$, strong - $W \in [.7, .9)$, and very strong concordance - $W \in [.9, 1]$ (where $\in$,(,[are designations of affiliation, open and closed interval).

Results and Discussion

The model implementation at TKTU resulted in stakeholder rankings of the factors influencing researchers to publish in WoS/Scopus journals (Figure 3). The results were as follows.

Rankings of motivational drivers at TKTU (Figure 3a):

- The researcher stakeholder - The highest-ranked drivers were the researcher's reputation (1st), job satisfaction (2nd), and promotion to academic positions (3rd). Publish-or-perish pressure reducing was the lowest-ranked (9th).
- The department head stakeholder – Department reputation and job satisfaction were the highest-ranked (1st). Promotion to administrative positions was 3rd. Publish-or-perish pressure reducing was the lowest-ranked (5th).
- The top university management stakeholder – The highest-ranked were university reputation (1st), increases in university subsidies (2nd), and increases in university accreditation/rating scores (3rd). Promotion to administrative positions took the lowest place (6th).

Rankings of potential features at TKTU (Figure 3b):

- The researcher stakeholder – University potential and prestige (1st), affinity for research (2nd), and promotion to academic positions (3rd) were the highest-ranked. Time was the lowest (9th).
- The department head stakeholder – The top-ranked features were the support for promotion to academic degrees (1st), support for promotion to academic positions (2nd), and scientific contacts abroad (3rd). Collaboration potential was the lowest-ranked (10th).
- The top university management stakeholder – The two top-ranked features were the same as for the department head stakeholder. The nomination for administrative positions was ranked 3rd. Support for the department scientific activity was the lowest-ranked (10th).

The results of TKTU ranking concordance showed that a strong concordance assessment prevailed (Table 2). It was typical of half of the stakeholder rankings and 70% of the researcher group rankings. Regarding motivational drivers, the department head and top university management stakeholders were exceptions, with moderate and very strong concordance respectively. Regarding potential features, there were four exceptions.

We ignored them, as the concordance of researcher stakeholder was above .5 (.5967). Thus, we concluded that the TKTU rankings were concordant.

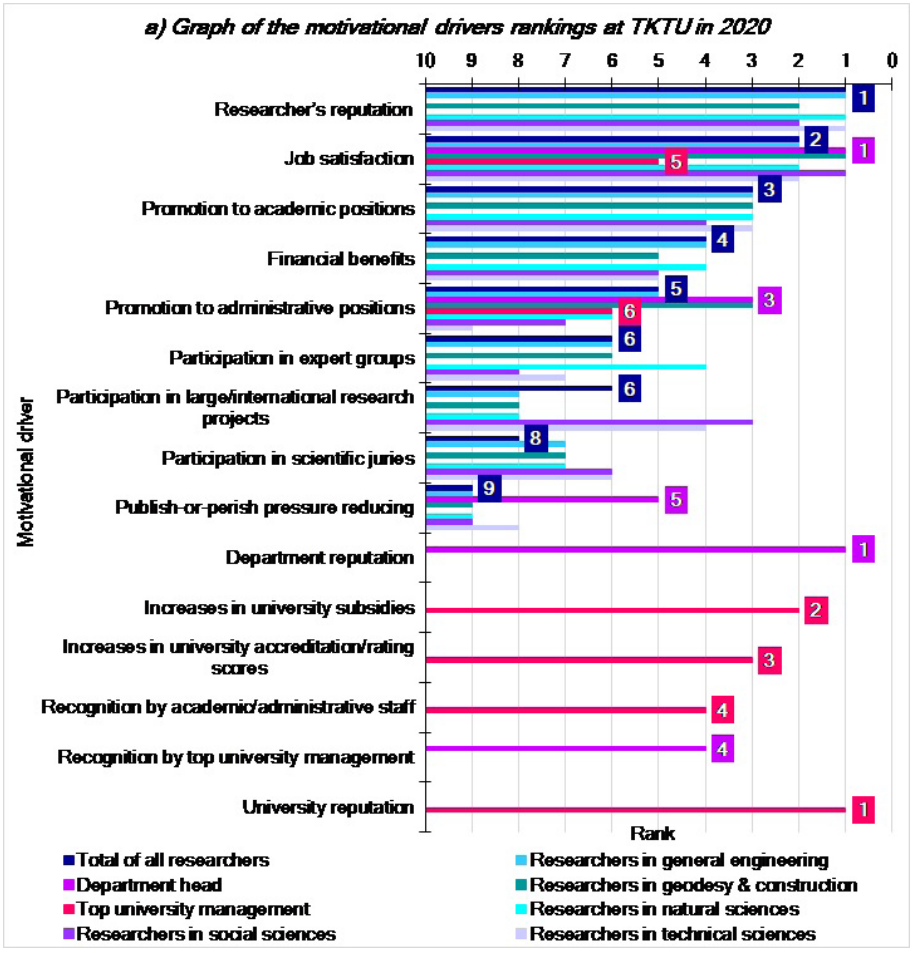
The experimental results gave us reason to draw the following main conclusions:

- For motivational drivers: The TKTU rankings did not confirm the results of models presented in our review. There were two exceptions: reputation and the ‘publish-or-perish’ principle. Reputation was ranked highest in the TKTU rankings and highly in the models of Daumiller and Dresel (2020) and Karimi and Asadnia (2020). The ‘publish-or-perish’ principle was the lowest-ranked in the TKTU rankings. Many authors oppose its use (Binswanger, 2015; Blaginin et al., 2019; Raitskaya & Tikhonova, 2020b, etc.). In a large-scale study, Richard et al. (2015) proved that it demotivates rather than motivates researchers.
- For potential features: The TKTU rankings differed largely from the sample results.
- For TKTU: Statistically, the experimental results are representative of TKTU due to the scope of the study (entire population). Therefore, they can be used as an element of a comprehensive model for research management of TKTU.
- The experimental results did not confirm our expectations for the highest ranking of material incentives/benefits as both a motivational driver and a potential feature.

The experimental results cannot be considered representative of the Bulgarian research community. Moreover, factor ranking is not incorporated in this model (limitation 3). Therefore, we expect variations in the results of its implementation at Bulgarian universities, depending on the university ownership, size, organizational and publication culture, community research field, etc.

Figure 3

Ranking graphs of the stakeholder motivational drivers and potential features at TKUT in 2020



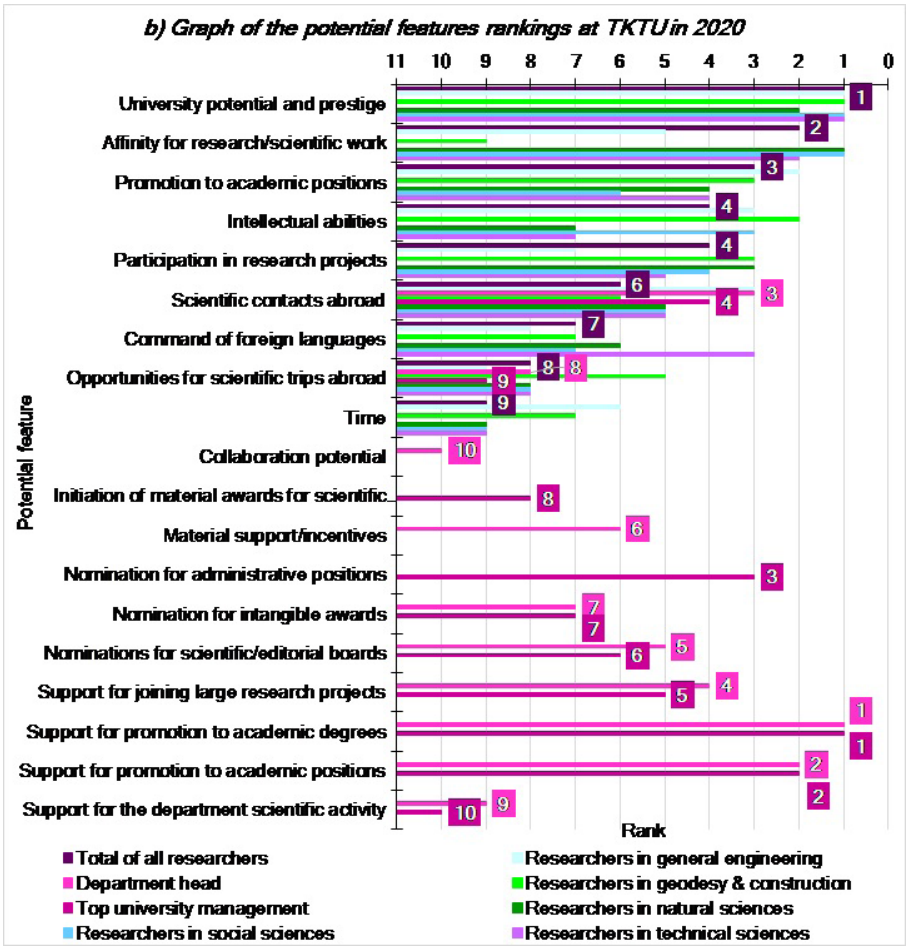


Table 2
Concordance of the TKTU rankings for motivational drivers and potential features in 2020

Stakeholders	Kendall's W rank concordance coefficient for			
	Motivational drivers		Potential features	
	Quantitative value	Qualitative level	Quantitative value	Qualitative level
Researcher group				
General Engineering	.7866	strong	.4413	moderate
Geodesy & Construction	.8102	strong	.1269	low
Natural Sciences	.8733	strong	.7707	strong
Social Sciences	.7912	strong	.7889	strong
Technical Sciences	.7071	strong	.4905	moderate
Total of all researchers	.8133	strong	.5967	significant
Department head	.3911	moderate	.8138	strong
Top university management	.9016	very strong	.8828	strong

Conclusion

This paper presents a model for factors influencing researchers publishing in WoS/Scopus journals, based mainly on positive reinforcement and a combination of concepts from social theories (control, management, stakeholder theory, psychology, etc.). The model is intended for Bulgarian universities. It covers motivational

drivers and potential features of internal university stakeholders directly involved in the topic. All but one of the motivational drivers are positive reinforcers based on strategies to strengthen behaviour. Factor ranking was not incorporated in the model.

Results from the model implementation at Todor Kableshkov Transport University in Sofia are also presented in the paper. They can be used for developing a comprehensive model for research management of this university. In our view, the model is also suitable for all universities (Bulgarian and foreign) operating in a similar context in terms of culture (national, organizational, and publication), reputational systems, educational regulations, etc.

We believe that the proposed model is a contribution to the theory of behaviour control. For future research, the model could be integrated into models for managing the influence on researchers to publish in WoS/Scopus journals. In terms of practice, we try to support efforts to encourage Bulgarian researchers to publish in prestigious journals. We also expect the pressure of negative reinforcers on researchers to be reduced as a result of the model's implementation. Overall, we hope that this paper will contribute to the improvement of university research management.

Conflict of interests

The authors declare that they have no conflict of interest.

References

Alfonso, A. G. (2019). Exogenous factors, implicit theories and linguistic-discursive procedures in scientific writing. *Enunciacion*, *24*(2), 184-198. https://doi.org/10.14483/22486798.15199

Banks, D. (2018). Thoughts on publishing the research article over the Centuries. *Publications*, *6*(1), Article 10. https://doi.org/10.3390/publications6010010

Barreto, I. B., Roa, E. R. A., & Velasquez, T. D. M. (2019). Scientific production of teachers in Peruvian Universities: motivations and perceptions. *Revista San Gregorio 35*, 86-96.

Beavitt, T. A., & Popova, N. G. (2020). The role of the English article system in developing dialogical context: A view from Russian science. *Journal of Language and Education*, *6*(3), 52-68. https://doi.org/10.17323/jle.2020.10360

Bedeian, A. G., Van Fleet, D. D., & Hyman, H. H. (2009). Scientific achievement and editorial board membership. *Organizational Research Methods*, *12*(2), 211-238. https://doi.org/10.1177/1094428107309312

Behl, A., Chavan, M., Dutta, P., & Sheorey, P. A. (2020). Benchmarking publication metrics for Indian business researchers. Exploring the role of collaboration. *Benchmarking-an International Journal*, *27*(2), 571-591. https://doi.org/10.1108/bij-04-2019-0184

Binswanger, M. (2015). How nonsense became excellence: Forcing professors to publish. In *Incentives and performance: Governance of research organizations* (pp. 19-32). Springer International Publishing. https://doi.org/10.1007/978-3-319-09785-5_2

Blaginin, V., Volkova, M., & Strielkowski, W. (2019). Measuring academic leadership using research performance. In W. Strielkowski (Ed.), *Sustainable leadership for entrepreneurs and academics, Esal2018* (pp. 355-362). Springer. https://doi.org/10.1007/978-3-030-15495-0_37

Bourne, L., & Walker, D. H. T. (2005). Visualising and mapping stakeholder influence. *Management Decision*, *43*(5), 649-660. https://doi.org/10.1108/00251740510597680

Cizmas, E., Feder, E.-S., Maticiuc, M.-D., & Vlad-Anghel, S. (2020). Team management, diversity, and performance as key influencing factors of organizational sustainable performance. *Sustainability, 12*(18), 7414. https://doi.org/10.3390/su12187414

Daumiller, M., & Dresel, M. (2020). Researchers' achievement goals: Prevalence, structure, and associations with job burnout/engagement and professional learning. *Contemporary Educational Psychology*, *61*, Article 101843. https://doi.org/10.1016/j.cedpsych.2020.101843

Daumiller, M., Siegel, S., & Dresel, M. (2019). Construction and validation of a short multidisciplinary research performance questionnaire (SMRPQ). *Research Evaluation*, *28*(3), 241–252, https://doi.org/10.1093/reseval/rvz009

David, S. T., Kendall, M. G., & Stuart, A. (1951). Some questions of distribution in the theory of rank correlation. *Biometrika*, *38*(1-2), 131-140. https://doi.org/10.1093/biomet/38.1-2.131

Dornyei, Z. (2009). The L2 motivational self system. In Z. Dornyei, & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 9-42). Multilingual Matters Ltd.

Freeman, R. E. (2015). *Strategic management: A stakeholder approach*. Cambridge University Press. https://doi.org/10.1017/CBO9781139192675

Ghinea, V., Mihaylova, L., & Papazov, E. (2015). Organizational culture dynamics: Complex systems dynamics. *Quality - Access to Success*, *16*(141), 99-105.

Hamner, C., Luthans, F., & Krietner, R. (2015). Operant behavior and organizational behavior modification. In *Organizational behavior 1: Essential theories of motivation and leadership* (pp. 114-133). Taylor and Francis.

Hung. B. P., & Baranovskaya, T. A. (2020). Second language teaching and teacher education in diverse contexts. *Journal of Language and Education*, *6*(3), 4-6. https://doi.org/10.17323/jle.2020.11455

Johnston, A. (2016). Motivation and the academic - where the drivers sit. *Journal of Management Development*, *35*(6), 765-777. https://doi.org/10.1108/jmd-09-2015-0140

Karimi, M. N., & Asadnia, F. (2020). Applied linguistics researchers' possible selves as motivational resources guiding their research productivity. *Research in Post-Compulsory Education*, *25*(2), 214-235. https://doi.org/10.1080/13596748.2020.1742994

Kirova, M. (2010). Effective combination of the performance characteristics of the innovations. In *The Proceedings of the 6th International Scientific Conference "Business and Management"* (vol. 1-2, pp. 399-406). VGTU. https://doi.org/10.3846/bm.2010.054

Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering - A systematic literature review. *Information and Software Technology*, *51*(1), 7-15. https://doi.org/10.1016/j.infsof.2008.09.009

Lambovska, M., & Yordanov, K. (2020). Motivation of researchers to publish in high-quality journals: A theoretical framework. *TEM Journal*, *9*(1), 188-197. https://doi.org/10.18421/tem91-27

Luthans, F., Luthans, B. C., & Luthans, K. W. (Eds.). (2015). *Organizational behavior: An evidence-based approach* (13th. ed.). Information Age Publishing.

Markus, H., & Nurius, P. (1986). Possible selves. *American Psychologist*, *41*(9), 954-969. https://doi.org/10.1037/0003-066x.41.9.954

Miller, A. N., Taylor, S. G., & Bedeian, A. G. (2011). Publish or perish: Academic life as management faculty live it. *Career Development International*, *16*(4-5), 422-445. https://doi.org/10.1108/13620431111167751

Ministry of Education and Science of the Republic of Bulgaria [MES RB]. (2010). *Bulgarian University Ranking System*. https://rsvu.mon.bg/rsvu4/#/

Ministry of Education and Science of the R Bulgaria [MES RB]. (2018). *Rules for monitoring and evaluation of the research activity of scientific organizations, the research fund and higher schools* [Act]. http://www.mon.bg›pravilnik_ocenka-NIDejnost_290618

Murray, H. A. (1936). Basic concepts for a psychology of personality. *Journal of General Psychology*, *15*(2), 241-268. https://doi.org/10.1080/00221309.1936.9917922

Osterloh, M., & Frey, B. S. (2014). Academic rankings between the "republic of science" and "new public management". In A. Lanteri , & J. Vromen (Eds.), *Economics of economists: Institutional setting, individual incentives, and future prospects* (pp. 77-103). Cambridge University Press. https://doi.org/10.1017/CBO9781139059145.005

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

Raitskaya, L., & Tikhonova, E. (2020b). Pressure to publish internationally: Scholarly writing coming to the fore. *Journal of Language and Education*, *6*(1), 4-7. https://doi.org/10.17323/jle.2020.10631

Raitskaya, L., & Tikhonova, E. (2020c). Seven deadly sins: Culture's effect on scholarly editing and publishing. *Journal of Language and Education*, *6*(3), 167-172. https://doi.org/10.17323/jle.2020.11205

Richard, J. E., Plimmer, G., Fam, K. S., & Campbell, C. (2015). Publishing success of marketing academics: Antecedents and outcomes. *European Journal of Marketing*, *49*(1-2), 123-145. https://doi.org/10.1108/ejm-06-2013-0311

Romani-Dias, M., Carneiro, J., & Barbosa, A. D. (2019). Internationalization of higher education institutions: The underestimated role of faculty. *International Journal of Educational Management*, *33*(2), 300-316. https://doi.org/10.1108/ijem-07-2017-0184

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*(1), 54-67. https://doi.org/10.1006/ceps.1999.1020

Simeonov, O. (1997). *Theoretical issues of control*. Akadem Consult.
Sink, D. S. (1985). *Productivity management: Planning, measurement and evaluation, control and improvement*. Wiley.
Sink, D. S. (1986). Performance and productivity-measurement - The art of developing creative scoreboards. *Industrial Engineering*, *18*(1), 86-90.
Stoyanov, I. (2014). Expectations, attitudes and behavior of the human factor in the organizational environment. *KSI Transactions on KNOWLEDGE SOCIETY*, *7(1)*, 45-49.
Stoyanov, I. (2012). Power and behaviour of the human factor in management. *Dialog E-journal*, *3*, 23-55.
Turan, F. K., Cetinkaya, S., & Ustun, C. (2016). A methodological framework to analyze stakeholder preferences and propose strategic pathways for a sustainable university. *Higher Education*, *72*(6), 743-760. https://doi.org/10.1007/s10734-015-9973-8
Vogel, R., & Hattke, F. (2018). How is the use of performance information related to performance of public sector professionals? Evidence from the field of academic research. *Public Performance & Management Review*, *41*(2), 390-414. https://doi.org/10.1080/15309576.2017.1400986
Vorontsova, A., Shvindina, H., Mayboroda, T., Mishenina, H. & Heiets, I. (2020). The impact of state regulation in a sphere of education on sustainable development of national economy. *Problems and Perspectives in Management*, *18*(4), 275-288. https://doi.org/10.21511/ppm.18(4).2020.23
Vroom, V. H. (1985). Citation Classic - Work and Motivation. *Current Contents/Social & Behavioral Sciences, 27*, 20-20.
Yordanov, K. (1997). Control as state power. *Economic Archive, 1*, 31-38.
Yordanov, K. (1999). Control on public finances. *Economic Archive, 1*, 36-41.

National Research University Higher School of Economics
Journal of Language & Education Volume 7, Issue 1, 2021

Pospelova, T. (2021). The Collaborative Discussion Model: Developing Writing Skills through Prewriting Discussion. *Journal of Language and Education, 7*(1), 156-170. https://doi.org/10.17323/jle.2021.10748

The Collaborative Discussion Model: Developing Writing Skills through Prewriting Discussion

Tatiana Pospelova
National Research University Higher School of Economics

Correspondence concerning this article should be addressed to Tatiana B. Pospelova, Higher School of Economics, 11, Pokrovskiy b-rd, Moscow, Russia, 109028. E-mail: tpospelova@hse.ru

This study aims to investigate the effect of peer-assisted prewriting discussion on second language (L2) academic writing and its benefits for students with different proficiency levels. While there is a significant body of research exploring the positive impact of collaboration on L2 writers' written performance and the ways it could be organised, there is little practical consideration on how to formulate explicit instruction. The rationale for this research lies in designing and arranging explicit instruction that could lead to L2 learners producing a higher quality writing output. Based on both qualitative and quantitative methods, and drawn on students' written texts and data analysis, the current study was conducted to devise and test a proposed model, which the author will term the 'collaborative discussion model' (the CDM). The control and experimental groups of Russian EFL students (n = 48) were engaged in written assignments after naturally occurring discussions and then the latter group was involved in an instructor-led discussion. The practice writing tasks were rated with the analytic rubric used in IELTS, assessing task response, coherence and cohesion, lexical resource, and grammatical range. The findings suggest that collaborative prewriting tasks, accomplished in the experimental group of students with different levels of L2 proficiency, may encourage students to engage more in reflection about the content and language of the text. As the texts produced after introducing the CDM were scored higher, especially on the criteria of task response and lexical resource, it is suggested that scaffolding prewriting discussions can potentially augment the writing skills of learners and the CDM can be used as a complementary activity to address the challenges associated with academic writing. The results of the questionnaire can imply that there are benefits of explicit instruction for students with different levels of L2 proficiency, although in nuanced ways and different degrees.

Keywords: prewriting discussions, collaborative activities, scaffolding, collaborative discussion model

Introduction

Academic writing purports to be an integral aspect of students' academic and professional training across the disciplines in the majority of higher institutions all over the world. University curricula are intensively using academic writing in English as a second language as a compulsory component of students' education. By and large, writing is seen as a complex language-focused process that consists of multiple interactive stages based on strategic functions (e.g., idea generation, outlining, drafting, revising, and editing) and, thus, requires from the user a range of embedded skills that must be combined in order to contribute to high-quality output. Not only does it embrace language-related elements, but it represents a complex social and cognitive process accompanied by an emotional component that affects learners' motivation and engagement (TEAL Center, 2012).According to the Common European Reference Framework for languages (CEFR), an advanced learner can write clear, well-structured expositions on complex subjects, underlining the relevant salient issues, expand and support points of view at some length with subsidiary points, reasons, and relevant examples (Council of Europe, 2011). Thus, academic English teaching and learning at the tertiary education level can be acknowledged as one of the most challenging tasks for both teachers and students.

In order to facilitate the process of academic writing acquisition, classroom practitioners have started to incorporate a variety of collaborative activities based on the sociocultural theory positing that language and cognitive skills develop through interactions with others. In the multifaceted process of academic writing,

collaborative learning classroom practices, with learners interacting in order to solve academic problems, have manifested a number of benefits. Besides enhancing the overall writing skills of learners, such activities can help them become more flexible, adapt to various language contexts and studying environments, and synergise literacy development in both speaking and writing in conjunction with listening and reading (Fernandez Dobao, 2012; Kim, 2008; Neumann & McDonough, 2014, 2015).

However, it seems necessary to pinpoint that although collaborative activities appear to be a widely utilised instructional technique in the learning environment, an array of factors can circumscribe the desired effect of collaboration. Not all interactions can be implicitly effective in achieving the goals of mastering a variety of linguistic features (Chen & Yu, 2019; McDonough & De Vleeschauwerb, 2019; Rahimpour et al, 2011; Zambrano et al, 2019). In content-focused collaborations, for example, language forms necessary for the language acquisition can be undervalued and overlooked while learners should be encouraged to reflect on their language usage, such as lexical choice, grammar, mechanics, etc. (Swain, 2010; Swain & Lapkin, 2001; Yazdi Amirkhiz et al, 2013). Therefore, this limitation necessitates designing and arranging collaborative classroom activities that could lead to L2 learners producing high-quality writing output. Some researchers have emphasised the vital importance of explicit instruction to scaffold developing writers in the process of planning, drafting, and revising the text as well as to boost their self-efficacy and motivation for writing (De Smedt et al, 2020; McDonough & De Vleeschauwerb, 2019). According to Mascolo (2009), it seems a rather challenging task for the teacher to devise and then apply in the classroom the model of instruction with pedagogically meaningful aims. It is also paramount to organise effective collaborative groups so that both learners of lower and higher proficiency levels could benefit from collaboration (Zambrano et al, 2019).

The present study corroborates and extends the body of research on collaborative prewriting discussions by supplementing it with a proposed instructional model, which the author will term 'the collaborative discussion model' (the CDM). It can scaffold learners while they acquire needed writing strategies and skills as well as confidence and positive motivation for writing. The main aim of the study is to investigate the effect of peer-assisted collaboration scaffolded by the CDM on Russian students. Based on both qualitative and quantitative methods, and drawn from the students' written texts and data analysis, this study addresses the following research questions:

(1) How does the proposed CDM affect the writing performance of Russian students?
(2) What stages of the discussion based on the CDM, if any, contribute the most to high-quality writing output?
(3) Do students with higher and lower L2 proficiency levels benefit equally from collaboration?

Literature Review

To date, scaffolding collaborative discussions as a prewriting strategy has been substantially implemented and studied in the field of academic writing teaching. Guided participation of the learners in the process of language acquisition implies not only their active involvement but interactive participation, which means that the learner appropriates skills and constructs knowledge from cooperation with the teacher. Only through experience conflicting with their own comprehension and in conflicting situations can a learner fill in the gaps in their knowledge (Mascolo, 2009). Recent studies have analysed effective educational situations when learners can support each other through actively partaking in the process of critical thinking during the discussion, with some of them being assigned the role of being more knowledgeable others (Kim, 2008, 2009; Leeser, 2004; Swain & Lapkin, 2001; Suzuki & Itagaki, 2007; Watanabe & Swain, 2007, 2008).

While communicative tasks are highly valued as they provide opportunities to communicate in the L2, a host of prerequisites should be taken into consideration to make interactions more successful. Effective collaboration can be administered in deliberately developed collaborative groups. In recent research, there has been a narrower focus on how scaffolding occurs when learners collaborate, in dyads, triads, or other small-size groups, in order to co-construct written texts (Di Camilla & Anton, 2012; Edstrom, 2015; Storch, 2005; Wigglesworth & Storch, 2009, 2012), where the benefits of such collaboration were well-documented, providing persuasive arguments in favour of this. Fernandez Dobao (2014) compared pairs and groups of four learners and found that the latter were more successful at solving language-related problems. Zambrano et

al. (2019) concluded that learners can take advantage of collaborative work when they receive guidance on how to collaborate but not when collaborative groups develop naturally. Previous research has also indicated that it is particularly explicit instruction, organized in various ways, that translates into better learner performance after peer-assisted collaboration (De Smedt et al., 2020; McDonough & De Vleeschauwer, 2019; Rahimpour et al., 2011; Zambrano et al., 2019). In a study by Ellis and Yuan (2003), the effects of different types of prewriting conditions, namely pre-task planning as a form of instruction, unpressured on-line planning, and no planning, on the writing performance of Chinese learners were examined, with pre-task planning resulting in greater fluency and syntactic variety of the output. Another study to report is the one by McDonough and Neumann (2014), who found that when students were given explicit instructions and visual tools facilitative of brainstorming ideas during preparation for writing they had more time to focus on linguistic features while writing. Teng and Zhang (2019) went further and proved that self-regulated learning strategies-based writing instruction contributed to increased levels of both linguistic and performance self-efficacy in students.

In order to elaborate on the findings of the above studies, it is necessary to highlight how collaborative discussions are organized. They are mainly presented in the form of 'languaging', described by Swain as 'the process of making meaning and shaping knowledge and experience through language' (Swain, 2006, p. 96). That means that while discussing the task, learners produce meaningful and comprehensible output in terms of the language. The concept of languaging is usually implemented into practice through language-related episodes (LREs), defined as segments of collaborative dialogue where the students reflect on what language forms to use (lexical choices, grammatical constructions, and mechanics) or correct themselves and others while performing the task (Swain & Lapkin, 1998, 2001; Yazdi Amirkhiz et al, 2013). It has been concluded that the participants' higher proficiency levels result in a greater number of LREs produced in the course of collaboration (Williams, 2001, 2008). Consequently, the L2 proficiency levels of learners and the way how they are organised in groups can be considered as imperative factors that can influence the procedure of LREs, their intensity, quality, and eventually the written outcome (Fernandez Dobao, 2012, 2014; Kim, 2008, 2009; Leeser, 2004; Neumann & McDonough, 2014, 2015; Watanabe & Swain, 2007, 2008).

In some cases, empirical studies have provided contradictory findings concerning the form-meaning construct of writing. Given that concentrating on every aspect of writing has been proved to be demanding due to the limitation of human attention, there could be trade-offs between content and form (Rahimpour et al., 2011). This means that, depending on the structure of the prewriting task, learners could either focus on the meaning of the ideas or how to formulate them, and this can have an effect on the written output. Thus far, some researchers have concluded that LREs have contributed to enhanced fluency and syntactic complexity, detecting no significant effect of collaborative planning on accuracy and giving little credit to the grammatical aspect (Ellis & Yuan, 2003; Kang & Lee, 2019; Rahimpour et al, 2011). However, a cluster of studies has come to the opposite conclusion, that collaboration has a positive impact on linguistic accuracy. Learners working together fulfilled the tasks more competently, in some cases producing shorter but grammatically more accurate texts (McDonough & Neumann, 2014, 2015; Storch & Wigglesworth, 2009, 2012; Williams, 2001, 2008). Storch (2005) revealed that collaboratively produced texts are not only linguistically more accurate and complex but also the quality of content improved; when students wrote a text collaboratively in dyads, the theses of their texts were more accurate and relevant.

To this end, an array of studies has explored what students take into account in their discussions while preparing for writing. Due to their nature, LREs imply that students tend to focus on form in terms of grammar, lexicon, mechanics, or discourse (Edstrom, 2015; Kim, 2008). Williams (2001) found that in meaning-focused classrooms, learners showcase the tendency to concentrate on lexis and, as a result, lexical LREs are far more frequent than grammatical LREs. At the same time, specifically devised grammar-focused tasks can enhance attention to grammar prior to writing (Fernandez Dobao, 2012; Leeser, 2004; Swain & Lapkin, 1998). Yet, it has been revealed that in LREs learners also discuss such components of written texts as their content or organisation (Elola & Oskoz, 2010; Storch, 2005). Students feel more confident expressing their ideas and analysing the ideas of their peers about the content and organisation of texts in the course of such prewriting sessions.

The current study builds on the findings of the previous research, positing the positive correlation between learner-learner interaction before writing and the quality of the produced texts. It aims to look into the

collaborative discussion method as a pre-writing strategy for the individual writing tasks of Russian students in an EAP course and, most importantly, focuses on practical considerations in the design and implementation of the collaborative discussion model. To achieve this, the study investigated what students tend to discuss during spontaneous collaborative activities before writing and what stages of the scaffolding discussions are considered the most or least useful by the students, which can further be a benchmark for improving prewriting instruction.

Materials and Methods

Participants

The focal participants were 48 EFL students (26 males, 22 females) with the same linguistic background, from different academic groups enrolled in the bachelor degree programme in the Department of Economics at the National Research University Higher School of Economics, Moscow, Russia. Their mean age constituted 19.2 years (SD = 1.6) and they had studied English previously for a mean of 10.1 years (SD = 1.9), with some of them starting in primary school, the others in secondary school, and all of them continuing in their first year at university. Upon entering the university, the students had been tested on their level of English language proficiency by the English Department. At the time of the experiment, they fell within the B1-B2 levels according to the Common European Framework of Reference. The participants were enrolled in the EAP course, which was required for their degree, and aimed to develop all four academic English language skills with no special emphasis on writing skills. At the end of the course, the students were supposed to take an IELTS mock examination. In terms of writing, the course targeted essay-level writing skills. The focus was on different essay types, i.e., opinion, discussion, advantages and disadvantages, problem and solution, and mixed question essays. The classes were conducted for two class periods per week, which also included Business English classes, in a 16-week semester. The participants were divided equally into the control group (CG) and the experimental group (EG). All the participants were engaged in the experiment on a voluntary basis and were guaranteed that in no case would the results of the experiment have any impact on their grades for the course.

Materials and Procedure

At the outset of the study, the students were involved in a naturally occurring face-to-face discussion, as one of the pre-writing activities. Before this, they had read and listened to texts on the corresponding topics in the students' theme-based EFL textbook. Then they were asked to accomplish a writing task (pre-test) in the form of an essay corresponding to the standard of the IELTS examination. In the pre-test stage, the students of both the experimental (EG) and the control (CG) groups were organised into smaller subgroups of four students according to their L2 proficiency level and each group included a more knowledgeable student to scaffold the peers with lower levels. The discussions in each group, which were audio-recorded, lasted for approximately 20-25 minutes before the students proceeded to the classroom-based individual writing. They were given 30 minutes to accomplish the task, a shorter time than that allocated in IELTS because the stage of planning had been done collaboratively. The submitted assignments were assessed independently by two raters. Inter-rater reliability constituted .79, which was calculated as the two-way mixed average measures intraclass correlation coefficient (in SPSS). The raters were specifically trained to evaluate essays according to the analytic rubric in the form of IELTS writing band descriptors for Task 2 (public version). The rubric included such components as task response with the requirements from candidates to formulate and develop a position in relation to a given prompt in the form of a question or statement; coherence and cohesion with the overall clarity and fluency of the message, how the response organises and links information, ideas, and language; lexical resource with the range of vocabulary the candidate has used and its accuracy and appropriateness; grammatical range and accuracy at the sentence level (Appendix C). The four components in each text were scored separately and constituted the average total score.

The post-test took place eight weeks after the pre-test. Based on the findings from the previous part of the experiment, the experimenter designed a structured prewriting discussion task, the collaborative discussion model (the CDM). Its implementation and effectiveness were tested further (Appendix A). The instructions and materials used for the discussions were devised in the way to encourage student reflection on content,

organization, and language, particularly, understanding the task, brainstorming main ideas, generating supporting details, providing examples, evaluating and organising ideas, together with eliciting useful vocabulary and grammar. This time, at the post-test stage, the students in the same experimental subgroups were given the CDM in the form of handouts. However, the control subgroups participated again in naturally occurring discussions. The same amounts of time of approximately 20-25 minutes and 30 minutes were allocated for discussion and writing, respectively.

At the final stage of the experiment, the EG students were asked to express their perceptions on the usefulness of organising instruction-led scaffolding activities, the CDM, in a specially designed questionnaire (Appendix B). The questionnaire asked the students to evaluate the different components of the prewriting session in terms of their practicality: understanding task requirements, brainstorming main ideas, generating supporting details, providing examples, evaluating ideas, organising ideas, and eliciting useful vocabulary and grammar. It also included questions about students' preferences for individual or collaborative types of prewriting activities and open questions about reasons why this type of activity (individual or collaborative) is preferred and how students felt while collaborating with the others and fulfilling the assignment. During the final stage of the study, all data and results were processed and described.

Data Analysis

The students' individually written texts were rated using the analytic rubric applied in IELTS, with equally weighted subscores for each component, which were composed average total scores for overall quality. The subscores included task response, coherence and cohesion, lexical resource, and grammatical range and accuracy and could be scored from 1 (poor) to 9 (excellent). The analysis of the written assignments was based on the number of scores from 4 (as the minimum given the proficiency level of the participants) to 9 (as the maximum) received by the students in both the control and the experimental groups as a total and also separately for each component of the analytic rubric before and after using the CDM. Task response was checked with a focus on the quality of addressing the task, expressing a clear position on the topic, developing ideas and supplementing them with details and examples, and the relevance of these ideas to the topic. Coherence and cohesion were gauged for the layout of the texts, presenting the central idea within a paragraph, sequencing the information with an overall progression, and using cohesive devices within and between sentences. The assessment of lexical resource included linguistic accuracy, the control of lexical features, word choice, collocations, word formation, spelling, and register. Grammatical accuracy was rated for the usage of the tenses, subject/verb agreement, subordinate clauses, word order, articles, prepositions, and punctuation. The collected data were afterwards subjected to statistical analysis, namely to a dependent *t*-test for paired samples, to check whether there were any statistically significant differences between the two tasks completed after naturally occurring discussions and after instruction-led discussions in both groups. The students' *t*-distribution under the null hypothesis, implying no differences between pre-test and post-test results, was calculated for both groups.The effects of the introduced CDM on the total scores and subscores for each component of the writing tasks were examined. Since the sample size was relatively small, the data collected via the questionnaire were analysed manually. The results of the study are presented in the next section.

Results

The first research question addressed the effect of the devised collaborative discussion model on the writing performance of students. As can be seen from Figure 1, at the pre-test stage of the study the most frequent scores for essays were 6 out of 9 (n = 11; n = 9), according to the IELTS assessment system, in both the control and experimental groups, which corresponded with the proficiency level of the participants. At the post-test stage, the results of the writing tasks changed in both groups. The majority of final essays in the EG earned a score of 7 (n = 11), while in the CG the most frequent score remained 6 (n = 9), with the number of scores 7 and 8 increasing by 1.

In order to determine if the pre-test and post-test results differed statistically significantly, they were compared using Student's*t*-values for the experimental and control groups. In the EG (Table 1), the differences between the pre-test and post-test total scores could be determined as highly significant ($t = 4.412$, $p < .001$), with the

critical values for *t*-tests on 23 degrees of freedom (n-1) being 3.768 (p = .001). The total average essay scores changed positively with a mean difference of .45 (SD = .228).

Figure 1

Number of scores for the students' essays after pre-test and post-test discussions for the experimental group (EG) and the control group (CG)

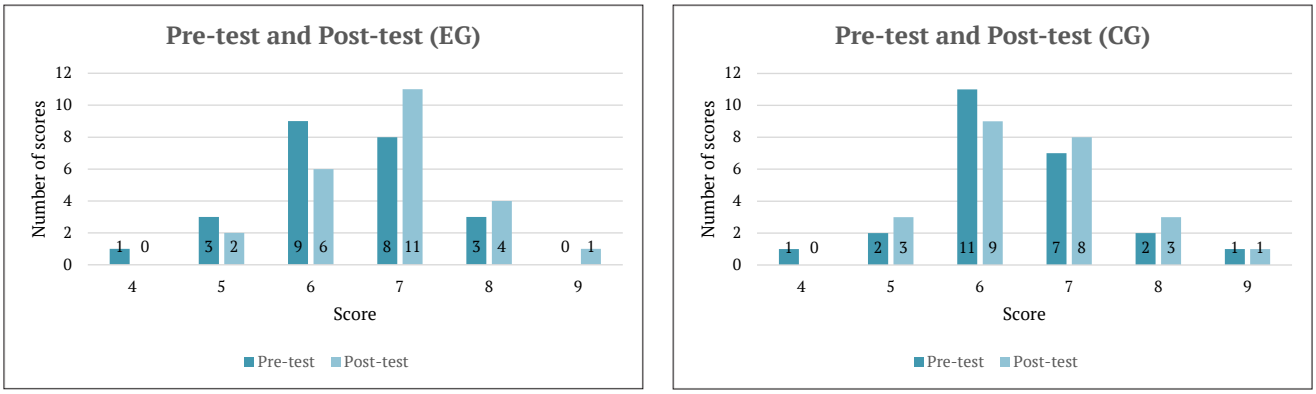


Table 1

The results of paired t-test for the experimental group (n = 24)

Assessment criteria	Mean		Mean Difference	t-emp	p-value
	pre-test	post-test			
Task response	6.29	7.04	.75	7.474	.001***
Coherence and cohesion	6.33	6.58	.25	2.769	.011*
Lexical resource	6.46	6.96	.50	4.796	.001***
Grammatical range and accuracy	6.46	6.63	.17	2.145	.043*
Total	**6.38**	**6.83**	**.45**	**4.412**	**.001*** **

Note. * - p < .05 ** - p < .01 *** - p < 0.001

In the CG (Table 2), marginally significant differences were found between the pre-test and post-test total results (t = 2.145, p < .05), with the critical values from the t distribution on 23 degrees of freedom (n-1) being 2.069 (p = .05). The average total score alongside the average subscores for the four components of the essays increased in their mean value, with the mean difference varying from .33 to .21 (SD = .062), which showcases consistent improvement. Thus, the total scores represent larger effect sizes as the t-values move further away from zero.

Table 2

The results of paired t-test for the control group (n=24)

Assessment criteria	Mean		Mean Difference	t-emp	p-value
	pre-test	post-test			
Task response	6.46	6.79	.33	3.391	.003**
Coherence and cohesion	6.38	6.63	.25	2.304	.031*
Lexical resource	6.25	6.50	.25	2.769	.011*
Grammatical range and accuracy	6.29	6.50	.21	2.460	.022*
Total	**6.42**	**6.58**	**.16**	**2.145**	**.043***

Note. * - p < .05 ** - p < .01 *** - p < .001

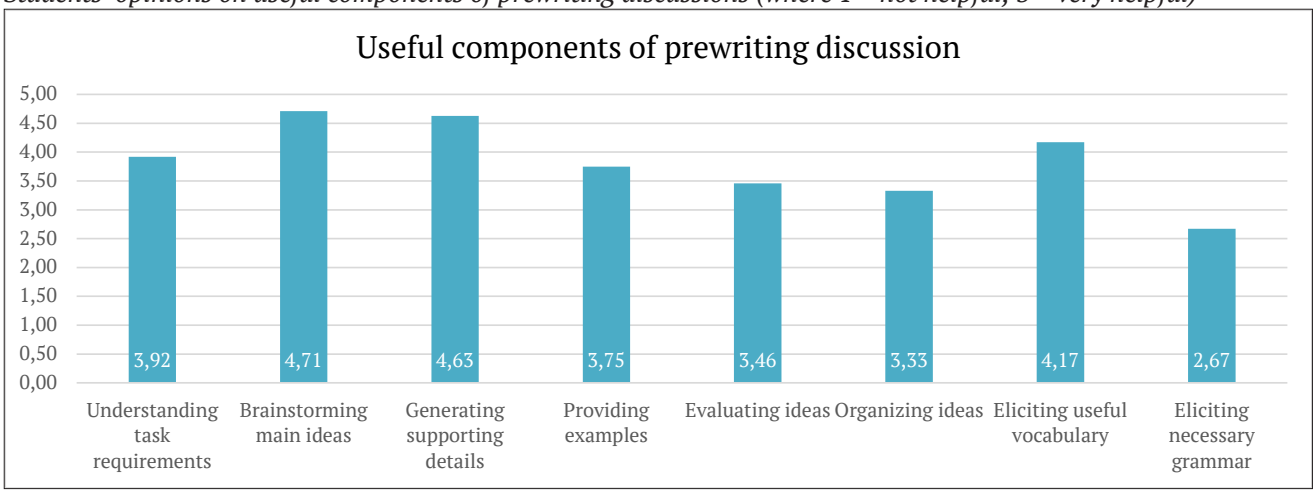
In sum, in relation to the first research question, the post-test texts written by both the experimental and control groups were of higher quality. In the EG, the total scores for the final essays were significantly higher than those obtained during the pre-test stage. In the CG, the results also improved but less significantly.

The second research question asked what stages of discussion supported by the collaborative discussion model contributed most to the quality of the writing output, provided that it improved. Based on the previous analysis, it could be suggested that different parts of the CDM had a different impact on the results of the essays. In the experimental group, the differences across the subscores for task response and lexical resource could be viewed as highly significant (t = 7.474, t = 4.796). The subscores for task response improved the most, with a more noticeable mean difference (MD = .75), which implies that more students received higher scores for this part of the task. The results of the pre-test and post-test differed significantly with a confidence level of 95 percent in relation to the subscores for coherence and cohesion (t = 2.769, p = .011) and also showcased a marginal significance for grammatical range and accuracy (t = 2.145, p = .043). In the control group, concerning the criterion of task response, the difference was significant with 99 percent confidence (t = 3.391, p < .01). However, across the criteria of coherence and cohesion, lexical resource, and grammatical range and accuracy, marginally significant differences were found between the pre-test and post-test results (t = 2.304, 2.769, 2.460, correspondingly), where the calculated t-emp did not exceed the critical values (p < .05).

In the questionnaire addressing the learners' attitudes toward the components of the CDM and the usefulness of working with peers in class (Appendix B), the students reported a positive perception of the structural prewriting discussions, with the highest rankings given to brainstorming, eliciting supporting details, and eliciting useful vocabulary. They believed the tasks were helpful for understanding the assignment and for selecting relevant examples. They found evaluating which ideas were appropriate for the assignment requirements and organizing ideas into a prewriting plan to be less useful. Their lowest rating was for choosing necessary grammatical structures. When asked if they preferred to work alone or with peers to carry out prewriting activities, the majority of the students claimed that they preferred to work in groups because they could generate more ideas, exchange different viewpoints, clarify information they had not understood, and get suggestions from their peers. Those five students who preferred to work alone mentioned that if they performed individually, they would have greater concentration and less dependence on others as well as less stress without being in the focus of their peers.

Figure 2

Students' opinions on useful components of prewriting discussions (where 1 = not helpful; 5 = very helpful)



When elaborating on the aspects of the prewriting discussions that were helpful, 21 of 24 students mentioned the general topics of ideas and content most often, specifically gathering more ideas to write about, selecting examples, developing a better understanding of the topic, and being exposed to critical views that lead them to find more relevant examples. Only eight students mentioned organisation, with five of them stating that the prewriting discussions helped them improve the organisation of their ideas. However, two students claimed that the organisation section of the prewriting discussions was not helpful and that this component of prewriting could be best done individually. Responding to the question about feelings, all the students pointed to the pleasant atmosphere of the classroom collaborative activity and positive feelings. In sum, based on higher subscores for task response and lexical resource in the post-test essays and on the students' opinions, it could be concluded that the parts of the CDM that focus on brainstorming, eliciting supporting details, and eliciting useful vocabulary produced a larger effect on the final result.

The third research question asked if students with higher and lower L2 proficiency levels benefited equally from collaboration. The results of the questionnaire suggest that practically all of the learners benefited somewhat from the task. Most of the students showed a positive attitude towards the collaborative discussions and felt a positive influence from this on their written assignments. They claimed a greater variety of ideas, knowledge, and creativity to share, and therefore more possibilities for language development. Almost half of them felt the positive impact of collaboration on the lexical variability for their texts. Five of these respondents found the discussions helpful for vocabulary learning, explaining that they were able to memorise a number of words they had learned from their more advanced peers. Four of the weaker students also pointed out that working in small groups provided motivation and a fun atmosphere, which made them feel comfortable. It might seem that the weaker students contributed little to the collaborative discussions but they were in fact actively involved in the process, not only as listeners and observers. Enjoying the comfortable atmosphere, they participated in the collaborations by asking questions, thus facilitating the discussion, and eliciting new ideas. Yet, it is also necessary to highlight that six of 24 learners felt the collaborative writing activities could not help them enhance their lexical or grammatical knowledge because of the similar proficiency level of their peers. These more knowledgeable learners saw opportunities to improve their fluency and overall speaking skills but did not think they could learn grammar or vocabulary while working with learners of the same proficiency level. Practically all the respondents mentioned the favourable and pleasant atmosphere while working with their peers, which allowed for arousing positive emotions accompanying the learning process. Therefore, it can be concluded that scaffolding activities proved to be beneficial for students with different levels of L2 proficiency although in nuanced ways and to different degrees, which is supported by the results of the study.

Discussion

The findings of this study are in line with those of previous research studies that established a positive relationship between the patterns of collaborative discussion and learners' performance in writing (McDonough & Neumann, 2015; Storch, 2005; Wigglesworth & Storch, 2009). As a whole, with the collaborative discussion model used as an explicitly instructed prewriting strategy inside small groups of peers with different levels of English language proficiency, the quality of essays of the students improved. The findings indicate that the texts produced following the instruction-led prewriting tasks were scored higher than the texts produced following the naturally occurring discussions. At the beginning of the experiment, during the naturally occurring discussions the learners spent time allocated for discussion chatting on the topic and content in general and even digressing from the topic. It was clear that the students did not make the best use of the time given for planning and ignored the criteria used in assessing their written assignments in the format of IELTS, although they were fairly aware of them. After the CDM was introduced and applied in the experimental subgroups, the learners tended to be more focused on the components of the assignments using the writing prompts and bearing in mind what they should discuss. As a result, the quality of the essays improved, especially relative to some particular criteria. Juxtaposing the initial essays and the final ones, it could be said that the most impactful part of the CDM appeared to be understanding the task, brainstorming main ideas, generating supporting details, providing examples, and also eliciting useful vocabulary since the quality of the essays concerning the corresponding criteria of task response and lexical resource improved significantly. Cohesion and coherence and grammatical range and accuracy improved marginally. This study has shown that during the instruction-led discussions the students discussed the content, organization, and vocabulary, which is in line with the previous research where the learners' higher degree of attention to the structure of the text and lexical choices rather than accuracy was attributed to the meaning-focused nature of the tasks (Storch & Wiglesworth, 2009; Williams, 2001, 2008). But these results contradict those of Storch (2005) and DiCamilla and Anton (2012), whose students focused their efforts on inventing content and also on addressing both grammatical and lexical issues. In the second experiment, the section of the CDM eliciting necessary grammar proved to be the least helpful, which could be explained by time constraints and difficulty for struggling students to advance the aspect of grammar mastering in such restricted conditions. Research into grammar instruction in the planning stage (TEAL, 2012) also showcases its negative results in terms of students' overall writing quality. Although teaching grammar purports to be an essential part of the language acquiring process, it seems more productive to use practice-oriented grammar approaches and their integration with other writing activities. Likewise, coherence and cohesion are attributed to such constructs of academic writing that tend to be rather problematic to be achieved comprehensively in a short period of time, although understanding the proper organisation of a text and using appropriate linking devices contribute much to the process.

The questionnaire data largely confirm the results of the study, as students reported that they talked about content more frequently during the prewriting discussions and that the prewriting discussions were the most helpful for brainstorming the main ideas, choosing the most relevant ones, and generating supporting details alongside eliciting the necessary task vocabulary. The students talked about the organisation of the ideas less frequently during the prewriting discussions, and fewer students reported that the prewriting discussions were helpful for eliciting the necessary grammar, although two of the stronger students mentioned that this section prompted them to revise and modify the grammatical structures in terms of complexity and variability. Given the fact that students' weak points concerning the formal aspect of academic writing appeared to be the organisation of the text, distinguishing between the topic sentence and supporting ideas, and well-balanced argumentation (Pospelova, 2016), the CDM can be an effective tool for fostering academic skills in writing. This refers both to macro-skills development like accomplishing the communicative function of the text, organising ideas logically, soliciting and using peer feedback for producing a written output, and micro-skills like using appropriate words and word combinations, grammatical patterns, and cohesive devices.

Notwithstanding the obvious benefits of instruction-led practices based on scaffolding, it seems necessary to elucidate the limitations of such scaffolding. The present study was implemented in groups of four students. When discussing the role of the learner-learner interactions in collaborative writing, we must consider the fact that not all small groups work equally effectively. These differences arise due to a plethora of factors such as topics for discussion, the size of the group, the personality types of participants, and their learning styles, confidence, experience, and goals (Storch, 2005), which could have influenced the results of this study. Identifying students' individual abilities and background knowledge can be a challenge to the organisation of successful scaffolding. Therefore, it is crucial that teachers be able to determine a more-knowledgeable-other (MKO) and allocate students in groups in a way so as to achieve the effect of matching learners with higher and lower proficiency levels without depriving them of the motivation to communicate (Fernandez Dobao, 2014; Leeser, 2004; Watanabe & Swain, 2007, 2008). In order to foster collaborative interactions, teachers must monitor group activities and use their observations to place passive or disengaged students effectively in future groups. The optimal number of participants in a group is yet to be explored.

Although it was not the main focus of the study, the students' perceptions about the collaborative discussions were elicited through the questionnaire administered at the end of the experiment. Practically all the respondents mentioned the favourable and pleasant atmosphere while working with the peers, which allowed for arousing positive emotions accompanying the learning process.

Conclusion

The present study attempted to investigate the effects of explicit instruction on the quality of writing, what students focus their attention on in the process of structured prewriting collaborative activities, and to design and test a collaborative discussion model that could scaffold effective learner-learner interactions based on the practicality of the components of the discussion. The study also addressed the question of benefits to students with higher and lower L2 proficiency levels from collaboration as a prewriting strategy. The analysis of the data confirmed that active-learning classroom practices aimed at articulating the steps in the complex process of academic writing were beneficial. In the course of prewriting collaborative discussions, with more advanced students scaffolding the less accomplished peers, learning occurs through a gradual process of the internalisation and employment of regulatory language used by others, as the weaker learner moves from assisted to independent performance (Vygotsky, 1978). Research into collaborative writing has shown that learners discuss content as well as the organisation of the text and language, and their interaction is positively associated with the text quality (McDonough & Neumann, 2015).

The current study indicates that collaborative prewriting tasks encouraged students to engage in reflection about their own and their peers' ideas. As the texts produced following collaborative prewriting tasks were scored higher, it is suggested that prewriting discussions can potentially augment learners' knowledge of the language (Wigglesworth & Storch, 2012).Inquiry learning engages students in critical thinking via a focused investigation into the assigned topic. The findings suggest that scaffolding may enhance inquiry and performance (Simons & Klein, 2007), especially when students are intentionally instructed on how to accomplish the task. Additionally, encouraging students to reflect on their own performance is the key to

perceiving progress not only in terms of the language, but in terms of communicative objectives. Active engagement in prewriting discussions may also contribute to learners' motivation and confidence.

The results of the study, although limited with the size of the sample, further extend the body of research into the learner-learner interactions in language-related episodes during prewriting discussions by supplementing existing data with the practical collaborative discussion model. The CDM was applied as an instrument to guide a reflective, language-focused conversation by defining a specific focus on the components of academic writing assignments.Being explicitly instructed, the learners were able to concentrate better on the task and ultimately were more successful in solving both conceptual and language-related problems. However, the grammatical aspect and cohesion improved marginally, thus leaving the door open for further research. Hence, it is proposed that the designed collaborative discussion model can be used as a tool to enhance in-class active discussions and as an alternative or a complementary activity to address the challenges associated with academic writing. The CDM suggests tasks that focus on the necessary elements of writing. Based on peer-peer scaffolding, it can engage students to be actively involved in the process of learning, by constructing knowledge rather than passively receiving it. It can develop the ability to comprehend and analyse issues, articulate thoughts, produce a higher quality of written output, and help students be more responsible for the learning process as a whole. The CDM can also be used as a sample for teachers to design similar models depending on the purposes of the task, the needs of learners, their individual characteristics, their motivation and readiness to cooperate, and other factors. Thus, teachers can adjust the instruction to reflect specific class learning conditions.

One of the essential constraints that collaborative discussions may face is the time required, which can explain why these types of activities are not commonly used in L2 writing classrooms. It should also be noted that the teacher's role in organising effective concurrent discussions and providing valuable feedback to students is paramount (Swain & Lapkin, 1998).The results have shown that unless students are given a specifying instruction, they tend to only express their ideas on the assigned topic without evaluating them in terms of the quality, considering alternatives, justifying them, and organising them into a writing plan. Learners may also need assistance establishing appropriate roles and expectations when working together.

This study can help teachers better analyse their students' writing and guide them on how to better instruct students in order to improve their formal writing competence. To this end, teachers need to focus on the parts of the language that are significant for the writing process and present some difficulties for students when evaluating the effectiveness of teaching materials and teaching techniques, revealing what parts of the syllabus have been inadequately learned or taught and require further attention.

Acknowledgements

Many thanks to the students who participated in this study as well as my colleagues who rated the written assignments and devoted their time and effort to the task.

References

British council. (n.d.). IELTS Task 2 writing descriptors. https://takeielts.britishcouncil.org/sites/default/files/ielts_task_2_writing_band_descriptors.pdf

Chen, W., & Yu, S. (2019). A longitudinal case study of changes in students' attitudes, participation, and learning in collaborative writing. *System*, *82*, 83–96. https://doi.org/10.1016/J.SYSTEM.2019.03.005

Council of Europe. (2011). *Common European Framework of Reference for Languages: Learning, teaching, assessment* (12th ed). Cambridge University Press.

De Smedt, F., Graham, S., & Van Keer, H. (2020). "It takes two": The added value of structured peer-assisted writing in explicit writing instruction. *Contemporary Educational Psychology*, *60*, 101835. https://doi.org/10.1016/j.cedpsych.2019.101835

DiCamilla, F., & Anton, M. (2012).Functions of L1 in the collaborative interaction of beginning and advanced second language learners. *International Journal of Applied Linguistics, 22*(2), 160–188. https://doi.org/10.1111/j.1473-4192.2011.00302.x

Edstrom, A. (2015). Triads in the L2 classroom: Interaction patterns and engagement during a collaborative task. *System*, *52,* 26–37. https://doi.org/10.1016/j.system.2015.04.014

Ellis, R., & Yuan, F. (2003). The effects of pre-task and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, *24*(1), 1–27. https://doi.org/10.1093/applin/24.1.1

Elola, I., & Oskoz, A. (2010). Collaborative writing: fostering foreign language and writing conventions development. *Language Learning and Technology*, *14*(3)*,* 51–71.

Fernandez Dobao, A. (2012). Collaborative writing tasks in the L2 classroom: comparing group, pair, and individual work. *Journal of Second Language Writing*, *21,* 40–58. https://doi.org/10.1016/j.jslw.2011.12.002

Fernandez Dobao, A. (2014). Attention to Form in collaborative writing tasks: Comparing pair and small group interaction. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes, 70*(2)*,* 158–187. https://doi.org/10.3138/cmlr.1768

Kang, S., & Lee, J.-H. (2019). Are two heads always better than one? The effects of collaborative planning on L2 writing in relation to task complexity. *Journal of Second Language Writing*, *45,* 61–72. https://doi.org/10.18823/asiatefl.2018.15.4.4.945

Kim, Y. (2008). The contribution of collaborative and individual tasks to the acquisition of L2 vocabulary. *Modern Language Journal*, *92*(1), 114–130. https://doi.org/10.1111/j.1540-4781.2008.00690.x

Kim, Y. (2009). The effects of task complexity on learner–learner interaction. *System*, *37,* 254–268. https://doi.org/10.1016/j.system.2009.02.003

Leeser, M. J. (2004). Learner proficiency and focus on form during collaborative dialogue. *Language Teaching Research, 8,* 55–81. https://doi.org/10.1191/1362168804lr134oa

Mascolo, M.F. (2009). Beyond student-centered and teacher-centered pedagogy: Teaching and learning as guided participation. *Pedagogy and the Human Sciences*, *1*(1), 3–27.

McDonough, K., & De Vleeschauwerb, J. (2019). Comparing the effect of collaborative and individual prewriting on EFL learners' writing development. *Journal of Second Language Writing*, *44,* 123–130. https://doi.org/10.1016/j.jslw.2019.04.003

Neumann, H., & McDonough, K. (2014). Exploring relationships among student preferences, prewriting tasks, and text quality in an EAP context. *Journal of English for Academic Purposes*, *15,* 14–26. https://doi.org/10.1016/j.jeap.2014.05.002

Neumann, H., & McDonough, K. (2015). Exploring student interaction during collaborative prewriting discussions and its relationship to L2 writing. *Journal of Second Language Writing*, *27,* 84–104. https://doi.org/10.1016/j.jslw.2014.09.009

Pospelova, T. B. (2016). The influence of self-editing on micro skills development in academic writing in English as a second language. *Journal of Language and Education*, *2*(1), 30–38. https://doi.org/10.17323/2411-7390-2016-2-1-30-38

Rahimpour, M., Mehrang, F., & Hosseini P. (2011). Effects of task design on EFL learners' written language performance. *The Journal of Asia TEFL, 8*(2)*,* 95–117.

Simons, K. D., & Klein, J. D. (2007). The impact of scaffolding and student achievement levels in a problem-based learning environment. *Instructional Science*, *35,* 41–72. https://doi.org/10.1007/s11251-006-9002-5

Storch, N. (2005). Collaborative writing: Product, process, and students' reflections. *Journal of Second Language Writing,* 14, 153–173. http://dx.doi.org/10.1016/j.jslw.2005.05.002

Suzuki, W., & Itagaki, N. (2007). Learner metalinguistic reflections following output-oriented and reflective activities. *Language Awareness*, *16*(2), 131–146. https://doi.org/10.2167/la392.0

Swain, M. (2006). Languaging, agency and collaboration in advanced language proficiency. In H. Byrnes (Ed.), *Advanced language learning: The contribution of Halliday and Vygotsky* (pp. 95–108). Continuum.

Swain, M. (2010). 'Talking it through': Languaging as a source of learning. In R. Batstone (Ed.), *Sociocognitive perspectives on language use and language learning* (pp. 112–130). Oxford University Press.

Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *The Modern Language Journal*, *82*(3), 320–337. https://doi.org/10.1111/j.1540-4781.1998.tb01209.x

Swain, M., & Lapkin, S. (2001). Focus on form through collaborative dialogue: Exploring task effects. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks* (pp. 99–118). Longman.

TEAL Center. (2012). *Just write! Guide*. https://teal.ed.gov/resources.

Teng, L. S. & Zhang, L. J. (2019). Empowering learners in the second/foreign language classroom: Can self-regulated learning strategies-based writing instruction make a difference? *Journal of Second Language Writing,* 48, 100701. https://doi.org/10.1016/j.jslw.2019.100701

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes.* Harvard University Press.

Watanabe, Y., & Swain, M. (2007). Effects of proficiency differences and patterns of pair interaction on second language learning: Collaborative dialogue between adult ESL learners. *Language Teaching Research*, *11*(2), 121–142. http://dx.doi.org/10.1177/136216880607074599

Watanabe, Y., & Swain, M. (2008). Perception of learner proficiency: Its impact on the interaction between an ESL learner and her higher and lower proficiency partners. *Language Awareness*, *17*(2), 115–130. http://dx.doi.org/10.1080/09658410802146651

Wigglesworth, G., & Storch, N. (2009). Pair versus individual writing: effects on fluency, complexity and accuracy. *Language Testing*, *26*(3), 445–466. https://doi.org/10.1177/0265532209104670

Wigglesworth, G., & Storch, N. (2012). What role for collaboration in writing and writing feedback. *Journal of Second Language Writing*, *21*, 364–374. https://doi.org/10.1016/j.jslw.2012.09.005

Williams, J. (2001). The effectiveness of spontaneous attention to form. *System*, *29*(3), 325–340. https://doi.org/10.1016/S0346-251X(01)00022-7

Williams, J. (2008). The speaking-writing connection in second language and academic literacy development. In D. Belcher & Hirvela, A. (Eds.), *The oral-literate connection: Perspectives on L2 speaking, writing, and other media interactions* (pp. 10–25). University of Michigan Press.

Yazdi Amirkhiz, S. Y., Abu Bakar, K., Samad, A.A., Baki, R.,& Mahmoudi, L. (2013). EFL/ESL learners' language related episodes (LREs) during performing collaborative writing tasks. *Journal of Language Teaching and Research*, *4*(3), 473–479. https://doi.org/10.4304/jltr.4.3.473-479

Zambrano R., J., Kirschner, F., Sweller, J., & Kirschner, P. A. (2019). Effects of prior knowledge on collaborative and individual learning. *Learning and Instruction, 63,* 101214. https://doi.org/10.1016/j.learninstruc.2019.05.011

Appendix A

THE COLLABORATIVE DISCUSSION MODEL

Writing topic: Nowadays more and more people have to compete with young people for the same jobs. What problems does this cause? What are some possible solutions?

Understanding the task
Underline key words/ micro-key words.
Think what focus your answer should have.

Brainstorming ideas / Generating supporting details / Providing examples
Listen to members of your group. Express your ideas. Record the ideas in the table.

Problems	+-?
Solutions	+-?
Specific actions	+-?
Positive consequences	+-?
Negativeconsequences	+-?

Evaluating ideas
Mark the ideas as good (+), bad (-) or irrelevant (?).
Record the feedback your ideas get.

Organising ideas
Choose what information you will include in your essay. In which order?
Make an outline (introduction, main body, conclusion). Share it with your group.
_____
_____
Linking words / phrases: _____

Eliciting useful vocabulary
Think about what vocabulary you will need.
Synonyms: _____
Topic-related words:_____

Eliciting necessary grammar
Think about what grammar you will need.
Tenses:_____
Modal verbs:_____
Participles:_____
Conditionals: _____
Adverbial clauses: _____
Comparisons: _____
Emphatic structures: _____

Appendix B

QUESTIONNAIRE

Question 1. How helpful was it to collaborate with your peers for different components of the writing process? Put the corresponding number 1-5, where 1 = not helpful; 5 = very helpful.

Rating / **Components**	**1**	**2**	**3**	**4**	**5**	**MV**
Understanding task requirements						
Brainstorming main ideas						
Generating supporting details						
Providing examples						
Evaluating ideas						
Organising ideas						
Eliciting useful vocabulary						
Eliciting necessary grammar						

Question 2. Which prewriting activities do you personally prefer? Why?

Activities	**Reasons**
Individual	
Collaborative	

Question 3. How did you feel during the prewriting activities?

_____

Appendix C

WRITING ANALYTIC RUBRIC

	Task Achievement	Coherence and Cohesion	Lexical Resource	Grammatical Range and Accuracy
9	fully addresses all parts ofthe task presents a fully developedposition in answer to the question with relevant, fully extended and well supported ideas	uses cohesion in such a way that it attracts no attention skillfully manages paragraphing	uses a wide range of vocabulary with very natural and sophisticated control of lexical features; rare minor errors occur only as 'slips'	uses a wide range of structures with full flexibility and accuracy; rare minor errors occur only as 'slips'
8	sufficiently addresses all parts of the task presents a well-developed response to the question with relevant, extended and supported ideas	sequences information and ideas logically manages all aspects of cohesion well uses paragraphing sufficiently and appropriately	uses a wide range of vocabulary fluently and flexibly to convey precise meanings skillfully uses uncommon lexical items but there may be occasional inaccuracies in word choice and collocation produces rare errors in spelling and/or word formation	uses a wide range of structures the majority of sentences are error-free makes only very occasional errors or inappropriacies
7	addresses all parts of the task presents a clear position throughout the response presents, extends and supports main ideas, but there may be a tendency to overgeneralise and/or supporting ideas may lack focus	logically organises information and ideas; there is clear progression throughout uses a range of cohesive devices appropriately although there may be some under-/over-use presents a clear central topic within each paragraph	uses a sufficient range of vocabulary to allow some flexibility and precision uses less common lexical items with some awareness of style and collocation may produce occasional errors in word choice, spelling and/or word formation	uses a variety of complex structures produces frequent error-free sentences has good control of grammar and punctuation but may make a few errors
6	addresses all parts of the task although some parts may be more fully covered than others presents a relevant position although the conclusions may become unclear or repetitive presents relevant main ideas but some may be inadequately developed/unclear	arranges information and ideas coherently and there is a clear overall progression uses cohesive devices effectively, but cohesion within and/or between sentences may be faulty or mechanical may not always use referencing clearly or appropriately uses paragraphing, but not always logically	uses an adequate range of vocabulary for the task attempts to use less common vocabulary but with some inaccuracy makes some errors in spelling and/or word formation, but they do not impede communication	uses a mix of simple and complex sentence forms makes some errors in grammar and punctuation but they rarely reduce communication
5	addresses the task only partially; the format may be inappropriate in places expresses a position but the development is not always clear and there may be no conclusions drawn presents some main ideas but these are limited and not sufficiently developed; there may be irrelevant detail	presents information with some organisation but there may be a lack of overall progression makes inadequate, inaccurate or over use of cohesive devices may be repetitive because of lack of referencing and substitution may not write in paragraphs, or paragraphing may be inadequate	uses a limited range of vocabulary, but this is minimally adequate for the task may make noticeable errors in spelling and/or word formation that may cause some difficulty for the reader	uses only a limited range of structures attempts complex sentences but these tend to be less accurate than simple sentences may make frequent grammatical errors and punctuation may be faulty; errors can cause some difficulty for the reader
4	responds to the task only in a minimal way or the answer is tangential; the format may be inappropriate presents a position but this is unclear presents some main ideas but these are difficult to identify and may be repetitive, irrelevant or not well supported	presents information and ideas but these are not arranged coherently and there is no clear progression in the response uses some basic cohesive devices but these may be inaccurate or repetitive may not write in paragraphs or their use may be confusing	uses only basic vocabulary which may be used repetitively or which may be inappropriate for the task has limited control of word formation and/or spelling; errors may cause strain for the reader	uses only a very limited range of structures with only rare use of subordinate clauses some structures are accurate but errors predominate, and punctuation is often faulty

Note. The table does not include data for bands 3-0. Adapted from https://takeielts.britishcouncil.org/sites/default/files/ielts_task_2_writing_band_descriptors.pdf. In the public domain.

National Research University Higher School of Economics
Journal of Language & Education Volume 7, Issue 1, 2021

Segura, M., Roquet, H., & Pérez-Vidal, C. (2021). The Effects of a CLIL Programme on Linguistic Progress at Two Different Points in Time. *Journal of Language and Education, 7*(1), 171-189. https://doi.org/10.17323/jle.2021.10981

The Effects of a CLIL Programme on Linguistic Progress at Two Different Points in Time

Marta Segura[1], Helena Roquet[1], Carmen Pérez-Vidal[2]

[1] Universitat Internacional de Catalunya

[2] Universitat Pompeu Fabra

Correspondence concerning this article should be addressed to Marta Segura Molla, Josep Trueta, s/n, 08195, Sant Cugat del Vallès, Spain. E-mail: mseguram@uic.es

In an attempt to explore the effects of different kinds of English as a Foreign Language (EFL) learning contexts, content and language integrated learning (CLIL) have been at the centre of FL acquisition research over the past decade. Studies have focused on the features and gains this setting brings, whether content is learnt at the same level of success as when taught in the learners' L1, and whether that L1 is negatively affected by CLIL. However, to our knowledge, very little attention has been brought to how the seniority of the programme affects learner progress in the target language. This study aims to fill such a gap in the understanding that the programme will have developed and improved in terms of quality of exposure and interaction, and that learners' EFL performance will be higher. To do that, we measured the efficacy of a long-standing CLIL programme in Barcelona twelve years after it was launched and examined the reading, writing, and lexico-grammatical abilities of CLIL EFL learners aged 8, 11, and 14 compared with results obtained by learners measured at the onset of the programme in 2005. The results showed that the quality of the programme has increased over the last decade, guaranteeing a higher level of EFL student proficiency when raw scores are considered, but not in terms of linguistic gains, in which only improvement in older students' grammar and reading skills can be observed.

***Keywords*:** content and language integrated learning (CLIL), CLIL programme assessment, second language acquisition, English as a foreign language (EFL), language learning

Introduction

Over the last few decades, instructed second language acquisition has experienced a change towards more meaningful activities in classrooms in order to generate a more natural and authentic learning context. In light of such developments, Content and Language Integrated Learning (CLIL) has gained ground in Europe[1], and in this respect, Catalonia, in Spain, is a case in point. It is believed that through CLIL approaches, in which a curricular subject is taught through the medium of a foreign language (FL), learners are immersed in more meaningful communication than in conventional formal instruction (FI). Within CLIL, English as a FL (EFL) learners are exposed to a higher quantity and quality of input, and have more opportunities for interaction. Even if most research shows that CLIL programmes result in benefits for learners (APAC, 2006; Brinton, Wesche, & Snow, 1989; Lasagabaster, 2008; Pérez-Vidal, 2013; Roquet, 2011; Van de Craen & Pérez-Vidal, 2001), more research is needed to assess their real impact in different settings.

Therefore, the goal of the present study is to contribute to this area of research with data obtained from a CLIL programme with English as the medium of instruction in a mainstream educational institution in the city of Barcelona. The study seeks to examine the impact of the school's CLIL programme twelve years after it was launched to analyse its effects as a programme that is no longer at its pilot stage, but that has been running for over a decade. This date is important for the main objective of the study; the evaluation of the programme's efficacy in bringing along higher levels of proficiency in EFL. In this study, we examine CLIL learners' abilities from 8 to 14 years of age, with data collected in 2017, and cross-sectionally contrast it with data collected at the

[1] CLIL programmes have developed more recently in South American countries and Australia but fall outside the scope of the current study.

onset of the programme in 2005 (Roquet, 2011). This should allow us to gauge the expected beneficial impact of the programme and assess its development over the years. The present study thus represents a unique opportunity to examine in detail a CLIL programme that is no longer in its infancy (Pérez-Vidal, Escobar, & Roquet, 2013).

Thus, the present research aims at providing new and meaningful data regarding a specific CLIL programme, namely in terms of the effects of programme lengths, as seen through the students' level and gains in reading, writing, and lexico-grammatical skills. Although CLIL is a branch of EFL acquisition that has been largely researched in recent decades, many studies have stated the infancy of the CLIL programmes analysed as a research limitation. Therefore, the main aim of this paper is to fill in this gap and compare one CLIL programme at two points in time: a couple of years after it was implemented, and twelve years later, when the programme is more stablished and stable.

Literature Review

The present section is devoted to briefly presenting an overview of the existing CLIL literature to provide a context for the relevance of the current research, starting with a definition of the term CLIL, followed by a description of its main features. We then briefly sketch out the specific context of the study and present a description of the onset of CLIL programmes in Europe. We close this section by reviewing the effects of CLIL on linguistic progress and the possible reasons explaining them, with a special emphasis on how long programmes have been running.

Definition and Features of CLIL

CLIL can be defined as an approach in which 'a language different from the domestic languages is used as the medium of instruction for curricular subjects' (Pérez-Vidal & Roquet, 2015). This way the target language is used in a context of real and meaningful communication (Llinares & Morton, 2010; Lorenzo, Casal, & Moore, 2010). While being the medium of instruction, students are 'learning language through language use' (Brinton, Wesche, & Snow, 1989) by engaging in meaningful and critical interactions (Darvin, Yi Lo, & Lin, 2020), while the curricular subjects taught though CLIL are also object of attention (Pérez-Vidal, 2013). What needs to be highlighted here is that the main goal in CLIL is to teach both the FL and the curricular content in an integrated way (Lasagabaster, 2008). Therefore, a CLIL approach should include both content and language objectives.

As a result, the extant limited amount of hours of L2 exposure in the conventional educational curricula is increased and schools become a multilingual environment, where languages are to be presented as a whole (San Isidro & Lasagabaster, 2019). Therefore, through CLIL approaches, languages are handled in an integrated way in schools. Instead of being taught and used separately and independently from each other, CLIL is intended to raise students' general language awareness (Navarro-Pablo & López Gándara, 2019) and promote translanguaging (García, Aponte, & Le, 2020), that is, the alternation from one language to another (Garcia, 2009) and their integration for different pragmatic purposes, as well as pluriliteracies, as each different language is used for a specific curricular subject in parallel to developing learners' L1s (Nikula & Moore, 2015). In sum, CLIL ultimately promotes plurilingualism with a wide scope (Dalton-Puffer, 2008; Nikula & Moore, 2015).

There is a debate, which is yet not resolved, regarding the extent to which the CLIL subject adopts the L1 cultural approach and pedagogy, considering that usually both students and the teacher share the same L1. Critical to CLIL is the fact that the teachers' and students' L2 proficiency tends to be limited. CLIL programmes also show some variable characteristics that make each CLIL programme slightly different, such as timing, continuation, support for students, resources for teachers, institutional implication, social spreading, and the status of the L2 language and culture (Pérez-Vidal, 2013). Another critical variable feature of CLIL programmes is their seniority; many CLIL research has been conducted with programmes that were still at their infancy, hence, in their pilot phases (Moore, 2009). However, more research is needed to determine whether CLIL programmes that are no longer in their pilot phase can lead to higher benefits for the students in terms of their L2 learning process.

CLIL: from Europe to Spain

Since the onset of the European Union in 1949, the European Commission has shown great concern for Europe's linguistic heritage and cultural diversity, promoting citizens' mobility, linguistic diversity, and multilingualism. With the aim of guaranteeing that citizens in Europe become functionally proficient in their mother tongue and in two other European languages (Cenoz, 2015; Llinares & Morton, 2010; Pérez-Vidal, 2007), it promoted an early starting age of instruction in a second language, an intensive and transdisciplinary approach starting in primary school, the addition of a third language in secondary education taught intensively and transdisciplinarily, and a university system guaranteeing mobility through ERASMUS (Pérez-Vidal, 2013).

As a consequence, the role and relevance of 'context of learning' began to gain ground in the shape of interdisciplinary teaching approaches (Gené Gil, 2016; Llinares & Morton, 2010; Pérez-Vidal & Juan-Garau, 2012). Thus, CLIL appeared to be the most suitable approach since it provides a real, student-centred, and meaningful context for language learning. Consequently, the CLIL approach clearly gained ground at all education levels in many countries (Lorenzo, Casal, & Moore, 2010; Pérez-Vidal, 2007; Van de Craen & Pérez-Vidal, 2001).

Concerning the specific context of the present research, CLIL programmes have been launched and implemented in the majority of Spanish autonomous territories (Fernández-Sanjurjo, Fernández-Costales, & Arias Blanco, 2017; Pérez-Vidal, Lorenzo, & Trenchs, 2015). This is indeed the case in Catalonia, where the current study takes place: a bilingual community where semi-immersion programmes have gained ground by using English as FL as a medium of instruction of other curricular subjects. Semi-immersion programmes, such as CLIL, allow for more real and meaningful contact with the FL within the classroom while still learning the language in a FL context where there is little or no contact with the language outside of the educational setting. Within this background, in recent years, the number of schools implementing CLIL programmes in Catalonia has grown considerably.

Effects of CLIL Programmes

It has been stated that CLIL brings many benefits to learners and many different CLIL programmes have been analysed to see which linguistic skills benefit most from such programmes. (APAC, 2006; Brinton, Wesche, & Snow, 1989; Lasagabaster, 2008; Pérez-Vidal, 2013; Roquet, 2011; Van de Craen & Pérez-Vidal, 2001). There is however one critical aspect in any CLIL programme, the length of time a programme has been running, that has been scarcely considered when interpreting data. Moore (2009) already stressed the fact that much of the research on the impact of CLIL was being conducted with programmes that were barely in their pilot phases and that we needed to give them time to settle in schools and be refined.

In contrast with the pilot programmes that have been analysed in many studies in the last several decades, some of which will be summarized later on, a well- and fully-established CLIL programme would show more stability and complexity, which is a potential key factor for the programme's success (Navés, 2009). Developing and implementing a new CLIL programme is a complex task that requires the teachers to be trained and new materials to be created. After this pilot stage in which the programme is first set and designed, CLIL programmes that have been running for some years will have more experienced teachers who bring consistency to the programme, as well as experience when designing and refining all aspects of its actual implementation (Navés, 2009; Navés & Muñoz, 1999).

Therefore, a CLIL programme that would have been running for many years in the same school is expected to be more robust and have more experienced teachers, which should potentially grant a more stable and complex design and application of the programme. In their study, Navarro-Pablo and López Gándara (2019) suggest that the CLIL group's higher scores may be attributed to the years that the programme has been implemented, the experience of the teachers, and the time students have been part of a CLIL programme. However, as mentioned above, most CLIL research has been conducted with CLIL programmes in their pilot phase. Thus, there is a need for more research to analyse the effects of a well-established CLIL programme as compared to a pilot phase one. This is precisely what the present study seeks to study: the analysis of a long-standing CLIL programme by measuring its efficacy in terms of student's linguistic gains correlated with the programme's maturity and growth.

One study, Roquet (2011), later summarized in Pérez-Vidal and Roquet (2015) and Roquet and Pérez-Vidal (2017), drew attention to this fact when the lack of statistically significant results was attributed to the pilot stage of the newly established programme. Indeed, these authors analysed CLIL learners compared to non-CLIL learners from the same school synchronically. The longitudinal study allowed for an analysis of the degree to which the learners' level of English progressed. All learners were administered five tasks: a written composition, a reading cloze, a dictation[2], a grammar and vocabulary error correction test, and a grammaticality judgement test. Roquet's (2011) results partially confirmed the hypothesis: the group with CLIL instruction yielded significantly better results in all the tests except the dictation test as compared to their peers of the same age and grade receiving formal instruction (FI) only. It was, therefore, concluded that a FI+CLIL learning context had proved to be more beneficial than a FI learning context only. However, the lack of statistically significant results in some of the tests administered led the authors to state that an academic year may not have been sufficient and, equally importantly, that the programme was still in its infancy when data was collected, at a 'pilot phase' stage, so that it might have still been too early for it to prove to be fully efficient.

To our knowledge, no study thus far has actually tried to prove that a well-established programme should yield better results when learners' proficiency gains are tested than at its onset[3]. In the wake of such an idea, the present study seeks to explore whether a fully established CLIL programme offered at the school in 2017 yielded significantly higher results in the form of more advanced EFL abilities at two different points in time (Roquet, 2011).

As mentioned above, most research regarding CLIL programmes and their effects on linguistic skills and language development focused on newly-established programmes, comparing the results of students involved in CLIL classes with the students attending only FI EFL lessons. It has been stated that CLIL brings many benefits to learners (APAC, 2006; Brinton, Wesche, & Snow, 1989; Lasagabaster, 2008; Pérez-Vidal, 2013; Roquet, 2011; Van de Craen & Pérez-Vidal, 2001). However, there is still no consensus regarding which areas or language abilities benefit most from it. Lasagabaster (2008) found that CLIL Basque students showed significantly higher oral and written proficiency than their FI peers. Moore (2009) also proved that certain competences are more likely to be improved with CLIL: receptive skills, vocabulary, morphology, fluency, creativity, risk-taking, and emotive-affective factors. Lorenzo, Casal, and Moore (2010) found that CLIL learners outperformed their FI peers in the four main linguistic skills both productively and receptively.

On the contrary, in a longitudinal study conducted with secondary education students, Admiraal, Westhoff, and de Bot (2005) did not find any significant differences between the CLIL and the FI groups in terms of receptive vocabulary. However, the CLIL students in the study did have higher results in speaking and reading, which is contradictory with Pladevall-Ballester and Vallbona (2016), who did not find any significant differences in reading in primary school students. Surprisingly, the same study also showed that the non-CLIL group outperformed the CLIL group in listening, although significant progress was seen in both groups.

Related to oral productive skills, Rallo Fabra and Juan-Garau (2011) found that CLIL students were understood better and showed less accented pronunciation than their FI peers. Such findings are in line with Ruiz de Zarobe's (2008) previous longitudinal study in which better results were seen in the CLIL group not only in terms of pronunciation, but also in vocabulary, grammar, and fluency. Nevertheless, in a later longitudinal study with secondary education students, Rallo Fabra and Jacob (2015) did not find any statistically significant differences between the CLIL and the FI groups in fluency on pronunciation.

In terms of content learning, previous research on CLIL around Europe has reported similar results between groups, or even favourable outcomes for the CLIL cohort in both the learning of the subject content (Hughes & Madrid, 2019) and the languages used as the medium of instruction, leading therefore to state that CLIL programmes do not seem to have a detrimental effect on the learners' L1 or on their achievement when

[2] The dictation test was not administered in 2017 for technical reasons.

[3] Having compared the most relevant data bases (Scopus and JCR), no previous studies were found in which one same programme was compared at two points in time to analyze linguistic progress.

curricular content learning is assessed (Fernández-Sanjurjo, Fernández-Costales, & Arias Blanco, 2017; San Isidro & Lasagabaster, 2019).

However, in their study, Fernández-Sanjurjo, Fernández-Costales, and Arias Blanco (2017) reported opposite results, since non-CLIL students performed slightly better than their CLIL peers when assessing their science knowledge, the subject taught through CLIL. In another study by Hughes and Madrid (2019), primary and secondary CLIL and non-CLIL students' science knowledge was examined, showing slightly higher results in science content in the case of the non-CLIL group in primary education when compared to a CLIL group the same age, a trend which was however changed as the CLIL group reached secondary education. The authors claimed that this might have been the consequence of the younger primary-level learners having a lower FL level as they start following a CLIL programme, an effect which is then mitigated later on (Hughes & Madrid, 2019). Similar results were reported in a longitudinal study by De Dios Martínez Agudo (2019) in which CLIL students showed better results than their peers in all linguistic domains and this tendency kept growing until the end of secondary education.

In sum, it can be claimed that CLIL programmes have been shown to bring about benefits for the students enrolled on them when compared to their FI peers in terms of vocabulary and morphology as well as motivation and creativity (Dalton-Puffer, 2008; Lasagabaster, 2008; Lorenzo, Casal, & Moore, 2010; Pérez-Vidal, 2011). However, there are some other competences for which contradictory results have been found; these include syntax, writing, informal language, pronunciation, fluency, and pragmatics (Dalton-Puffer, 2008; Pérez-Vidal, 2011, Roquet & Pérez-Vidal, 2017). Therefore, more research is needed, namely in the form of longitudinal studies (Lasagabaster & Doiz, 2017; San Isidro & Lasagabaster, 2019) that analyse CLIL effects across all educational levels to see if said CLIL benefits are persistent until the end of secondary education (De Dios Martínez Agudo, 2019; Hughes & Madrid, 2019).

Besides the abovementioned linguistic features, some benefits regarding the students' entire educational process (Lorenzo, Casal, & Moore, 2010) and their knowledge and usage of their mother tongue have also been identified (APAC, 2006), such as greater problem-solving skills, more independence as learners and linguistic spontaneity (DeKeyser, 2000; Derakhshan & Karimi, 2015), and positive attitudes and an interest in learning (Lasagabaster & Sierra, 2009).

Such benefits probably come from the nature of the CLIL environment where students are exposed to high-quality input and opportunities for output practice and meaning negotiation are present (Roquet & Pérez-Vidal, 2017). Generally, the CLIL hours are extra hours of exposure and interaction in the FL (APAC, 2006; Dafouz Milne & Guerrini, 2009). In CLIL classrooms, learners are immersed in transdisciplinary lessons, usually including science, physical education, informatics, or arts and crafts content (Dafouz Milne & Guerrini, 2009; Pérez-Vidal & Roquet, 2015), which allows them to reduce the pressure to learn the language (Auerbach, 2006; Dalton-Puffer, 2008; Lasagabaster & Sierra, 2009; Van de Craen & Pérez-Vidal, 2001).

It is also important to note that, as pointed out by Bruton (2011) after reviewing several studies on CLIL, it seems that in many cases the students in the CLIL groups are self-selected, that is to say students who choose the CLIL option are more motivated, have a higher socio-economic family status, or have a higher linguistic level. Therefore, those CLIL benefits stated in the previous paragraphs may not only come from the communicative and more meaningful nature of such programmes, but also from the students' sociolinguistic background and initial proficiency level in the L2. However, Hüttner and Smit (2014) claimed that an approach such as CLIL, which increases student engagement in content and language learning simultaneously, attempts to diminish any discriminatory differences in learners' linguistic knowledge.

Finally, one other feature of CLIL pedagogy is that teachers' roles change from the FI conventional ones: the teacher stops being either the FL expert or the subject expert to actually become a combination of both (APAC, 2006), although sometimes the role of the content and the language teacher is still unclear (Darvin, Yi Lo, & Lin, 2020). In this respect, it has been claimed that teacher education in CLIL is a primary requirement. Information about CLIL as a multilingual approach, its challenges and advantages, the existing tools, and well-tested pedagogical practices should be subject to training. CLIL should spur collaboration between the programme's language and content teachers (Darvin, Yi Lo, & Lin, 2020; Jaén Campos, 2016; Papaja, 2013), as they will need to work together.

In sum, the previously mentioned conditions taken together, namely the meaningful context that a fully-established CLIL programme provides, can be said to be optimal for learners' language acquisition and non-linguistic positions vis-à-vis their target language in their language acquisition process (Dalton-Puffer, 2008; Kersten & Rohde, 2013). Nevertheless, many of the abovementioned studies were based on pilot programmes, which are naturally less stable since they are still developing. Therefore, further research is needed that analyses CLIL programmes that have been running for some years and that are more stable and robust. That is precisely what the present research aims at doing, analysing a fully-developed CLIL programme and contrasting how it affects students learning of the FL in comparison with the same programme in its pilot phase twelve years before.

The Present Study

The present study was conducted in a school in Barcelona, Catalonia, which had launched a well-planned CLIL programme in 2004-2005. Through the programme, learners were exposed to one extra subject, science, in English, in addition to the English FI hours. After designing the programme in the 2002-2003 school year, training the teachers, and informing the families, the programme was launched the next year with the third and fifth graders. In the following school years, the programme was extended to secondary education and now it stretches up to the first year of the Baccalauréat, the first of two years of preparation for university entry.

Parallel to the implementation of the programme, the school increased the number of EFL conventional sessions from three to four or five sessions per week, starting at the age of four. As a result, when pupils reached the third grade of primary, they had been exposed to 420 hours of English, which was considered enough to follow a science course in English. The school also made provisions to include the vocabulary needed to understand a science course in English during the English language sessions.

Our study took advantage of the fact that the impact of this same programme on learners' EFL proficiency two years after its implementation had already been analysed in 2005 (see Roquet, 2011), and full access to this data was granted. Thus, the data collected in 2005 were used and compared with the new data collected for an appraisal of the linguistic effects of the CLIL programme more than a decade later.

With that goal in mind, the following research questions were set:

RQ1. Did the 2017 CLIL learners display significantly higher linguistic abilities than the 2005 learners from the same academic grades and ages respectively, that is at the third and sixth grades of primary, and at the second grade of secondary?

We hypothesized that the 2017 cohort would yield better results because they would have benefited from an improved, more mature, and better-established CLIL programme.

RQ2. What are the differential gains in linguistic abilities (reading, writing, grammar, and vocabulary) between the 2005 cohorts and the 2017 cohort, when comparing cross-sectional gains between the third and sixth grades in primary and between the sixth grade in primary and the second grade in secondary?

We hypothesized that gains between the third and six grades in primary and between the sixth grade in primary and the second grade in secondary would be grater in the 2017 cohort than in the 2005 cohort due to the growth the programme would have experienced.

RQ3. What are the differential cross-sectional gains shown in the learners' linguistic abilities (reading, writing, grammar, and vocabulary) between the younger group (third grade in primary) and the eldest group, (second grade in secondary) when comparing the 2005 and the 2017 programmes?

We hypothesized that, as a result of the programme having evolved over the last decade, the 2017 cohort would show higher gains from the third grade of primary school to the second grade of secondary school when compared to the 2005 cohort.

Materials and Methods

In this section, the design is presented, the participants are described, and the materials and procedures for data collection and analyses are given.

Research Design

The current study cross-sectionally measured and contrasted CLIL learners following a CLIL programme consisting of a science subject taught in English, in addition to FI in that language. Students' linguistic abilities were measured in order to identify possible improvements in the programme.

As displayed in Table 1 below, learners were tested in two different years: 2005 and 2017. In both testing times, students were administered the tests in their third trimester, namely between April and May. The 2005 data is actually secondary data in this study as we are reanalysing Roquet's (2011) corpus, which has been contrasted with the new 2017 data for this study.

To that end, learners were tested at three different grades and ages coinciding with the onset, the middle, and the end of the CLIL programme. Data from the third grade in primary education serves as pre-test data, and data from the sixth grade in primary serves both as post-test data for the third graders and as pre-test data for learners in the second grade of secondary education.

Participants

The sample of participants analysed included a total of 90 students from the 2004-2005 cohort, and 90 students from the 2017 cohort. The participants in both groups were Catalan and Spanish bilingual learners from the same school and they were all following the school's CLIL programme, having contact with the English language not only in their FI class, but also in their science CLIL lessons. Thus, students tested in both testing times had had the same exposure to English in the school as their same-age peers from the other data collection cohort.

Participants were distributed in three subgroups according to testing time and educational level: the third year of primary (8-9 year olds), Group 1 in 2005 and Group 2 in 2017; the 6th year of primary (11-12 year olds), Group 3 in 2005 and Group 4 in 2017; and the second grade of secondary (13-14 year olds), Group 5 in 2005 and Group 6 in 2017. Table 1 below shows the distribution of the participants in groups. The team of English teachers and CLIL teachers involved in the programme in 2005 and in 2017 was a stable group. Furthermore, the main parameters of the programme were also kept the same, namely the CLIL programme coordinator was the same and so was the pedagogical approach, as confirmed by the school's management team, guaranteeing the programme's stability and development over the years.

Data were collected from all learners in those three grades at the two testing times, that is from 120 students per grade in primary and 90 students in the second grade of secondary. However, only a randomly selected sample of 30 students from each grade at each testing time was analysed, providing a stratified sample that included students from all the different groups in each school grade. Written consent to participate in the study was requested from all of the students' families through a form in which details regarding the research were presented. All students were tested, as the tests were also included as classroom activities of the school's English subject.

Table 1
Distribution of the participants in groups

Testing time 1 (2005)	Testing time 2 (2017)
Group 1 (*N*= 30) / 3rd primary (8 and 9 years old)	Group 2 (*N*= 30) / 3rd primary (8 and 9 years old)
Group 3 (*N*= 30) / 6th primary (11 and 12 years old)	Group 4 (*N*= 30) / 6th primary (11 and 12 years old)
Group 5 (*N*= 30) / 2nd secondary (13 and 14 years old)	Group 6 (*N*= 30) / 2nd secondary (13 and 14 years old)

Materials

A series of tests were administered to the participants at both testing times: a writing test, a grammar and vocabulary test, and a reading test. Those four main skills were chosen to be the focus of the testing since they provide an overall picture of the students' linguistic competence. However, speaking and listening were not included in the testing due to time and space constraints. The tests were kept the same for both testing times and all three grades tested to guarantee that the scores were comparable across levels and between both testing times. The instructions for all tests were written in Catalan to help understanding.

All tests were created by the researchers in 2005, following guidelines from the schoolteachers to ensure that they corresponded to the educational levels. The tests were piloted in 2005 before the first testing time and were kept the same in 2017. Thus, the tests were criterion-referenced and scored on an absolute scale.

The writing test was based on a picture that showed two policemen taking a statement from a woman and a young boy inside a flat. Students were asked to write three compositions in the space provided: one imagining a dialogue for the picture, another explaining what they thought had triggered the situation, and, finally, another guessing how the situation would end.The three questions that were given to the students were the following (they have been translated from Catalan into English):

> Write a dialogue imagining the situation involving the policemen, the boy, and his mother.
>
> Write a short paragraph describing why you think the situation you imagined in the dialogue actually took place.
>
> Write a short paragraph explaining howyou think this situation will end?

The grammar and vocabulary test consisted of two separate tasks. The first one was an error correction task. It included 30 statements grouped into three sections of 10 sentences each, with an increasing degree of difficulty, which needed to be corrected in a multiple-choice answer format by choosing either the correct option to fill in the blank in the sentence or selecting the correct sentence out of the three options given. Some examples of the questions included in the error correction test are given below:

_____? No, I'm sorry, I haven't got any change.

A. Do you can change this 5 € note?
B. Have you change for this 5 € note?
C. Have you got change for this 5 € note?

I visited London last year.

A. How long have you stayed there?
B. How long did you stay there?
C. How long you stay there?

How _____ people are there in class?

A. much
B. more
C. many

Why are you putting on your coat?

A. Because I took the dog for a walk.
B. Because I'll take the dog for a walk.
C. Because I'm going to take the dog for a walk.

The second grammar and vocabulary test, a grammaticality judgement test, consisted of 20 sentences of increasing difficulty for which the students had to decide whether they were correct or incorrect. Some examples of the questions included in the grammaticality judgement test are presented below:

She can't to sky.

☐ Correct
☐ Incorrect

I am taller than my brother.
☐ Correct
☐ Incorrect

The cat is sleep under the chair.
☐ Correct
☐ Incorrect

I am going to write a novel some day.
☐ Correct
☐ Incorrect

The reading test consisted of a cloze activity with a multiple-choice answer format. The learners were administered a text dealing with a topic studied in the science course with 20 missing words and four options for each of the blanks. The second paragraph (of three in total) of the reading test, with its corresponding possible answers, is presented below:

> Tsunami is a Japanese word that (6) 'harbour wave'. But why do tsunamis (7)? Tsunamis are usually caused by earthquakes at the bottom of the sea. At first, the (8) in the sea is quite small, but it moves very (9). When the wave gets close to the coast, the ocean floor makes it grow enormously. By the time it reaches the (10) it has become huge. Some tsunamis can be 30 meters (11). These giant waves can hit Japan, Indonesia, Central (12) and South America.

(6)	mean	signifies	means	tells
(7)	happens	happen	exists	go
(8)	wave	water	animals	one
(9)	quick	speedy	rapid	quickly
(10)	sea	cost	costal	coast
(11)	high	height	tall	of high
(12)	america	America	American	USA

Procedure

Data collection procedures were kept basically the same in 2017 as in 2005. In 2005, the participants were tested in their classrooms over one two-hour session. During the first hour, three tests were administered: the error correction test (15 minutes), the grammaticality judgement test (10 minutes), and the writing test (20 minutes). In the second hour, the reading test (15 minutes)[4] was completed. All tests were pen and paper, and teachers were given previous instructions such as to strictly control the time and to guarantee an exam-like environment.

In the second testing time, in 2017, the tests were administered over two separate sessions. The first one took place in the classrooms and included the writing test (20 minutes), which was a pen and paper test, as in 2005. The second session took place in the computer room and included the rest of the tests: the error correction test (15 minutes), the grammaticality judgement test (10 minutes), and the reading test (15 minutes), which were digitalized and administered through three different Google Forms questionnaires. Teachers were also given precise instructions for test administration. The change introduced in this data collection was a reflection of the well-established use of computers both in the school and in the learners' individual work outside school hours, which made 'pen-and-paper only' tests obsolete to a certain extent, and perhaps less valid as a measure of proficiency.

In addition to the abovementioned tests, students were administered a socio-linguistic background questionnaire, tapping onto the participants' linguistic background such as their L1 and their knowledge of other FLs. The results obtained from the sociolinguistic background questionnaire allowed for controlling external factors such as the students' mother tongue and their participation in English extracurricular lessons. Thus, students with English as their mother tongue or enrolled in English extracurricular activities were

[4] The reading test was not administered to the third grade of primary school group in 2005.

excluded from the sample of participants. However, even though those two main factors were controlled, it is important to underline that in 2017, as compared to 2005, there was more extramural contact in English, namely students had more access to resources in English in their daily lives.

Analyses

The following procedure was followed for the analysis. To correct the grammar and vocabulary tests, and the reading test, correct answers added points and incorrect answers did not. As for the writing test, learners' compositions were evaluated on the basis of Friedl and Auer's (2007) wholistic rating model, which includes four main areas (task fulfilment, organization, grammar, and vocabulary) rated along a scale where 0 is the lowest grade and 5 the highest. Interrater reliability was calculated on the basis of 10% of the compositions, which were corrected by two different ratters.

With the raw results obtained from the grading procedure explained above, statistical analyses were conducted, namely t-tests. All data was statistically analysed with Version 16 of IBM's SPSS Statistics package and the p value was set at (p=0.05).

Results

The first research question enquired into the comparative results obtained from the participants in 2005 and 2017 at the three educational levels analysed. It was hypothesized that the 2017 cohort would obtain significantly higher results than the 2005 cohort. When comparing the results in all tests, as shown in Table 2 and Figure 1 below, participants in the third grade of primary obtained higher grades in 2017 than in 2005. Regarding the writing test, the mean score obtained in 2005 was 1.05 out of 10 points; while in 2017 it was 1.92, the difference being statistically significant (p=0.000). In the error correction test, the mean in 2005 was 3.28; while in 2017, it was 4.60, with a statistically significant difference (p=0.001). In the grammaticality judgement test, the mean in 2005 was 3.90; while in 2017 the mean was 5.76, such difference being statistically significant (p=0.000). Concerning the reading test, no comparison was possible as this test was not administered in 2005 to this grade.

Table 2
Group statistics for third graders in 2005 (Group 1) and 2017 (Group 2)

Group		N	Mean	Std. Deviation	Std. Error	P-value
Writing test	1	30	1.050	.6345	.1158	.000
	2	30	1.917	1.0178	.1858	.000
Error correction test	1	30	3.283	1.3110	.2394	.001
	2	30	4.600	1.6785	.3064	.001
Grammaticality judgement test	1	30	3.900	1.9538	.3567	.000
	2	30	5.767	1.3309	.2430	.000
Reading test	1	30	.000	.0000	.0000	.000
	2	30	3.850	1.2536	.2289	.000

Regarding the results obtained by the sixth primary grade learners, as presented in Table 3 and Figure 2 below, higher grades were reported in 2017 for the writing and on the grammaticality judgement test. As for the writing test, the mean score in 2005 was 4.22; while in 2017, it was 4.63. In the grammaticality judgement test, the mean in 2005 was 5.34; while it was 6.04 in 2017. Such a difference was not statistically significant in the writing test (p=0.128) whereas it was significant for the grammaticality judgement test (p=0.026).

On the contrary, lower results were reported in 2017 than in 2005 for error correction and the reading tests. Regarding the error correction test, the mean score in 2005 was 6.82; while in 2017, it was 6.70, but this was not statistically significant (p=0.729). On the reading test, the mean in 2005 was 6.62; while in 2017, it was 5.82, the difference not being statistically significant either (p=0.106).

Figure 1
Test results for the third year of primary at both testing times

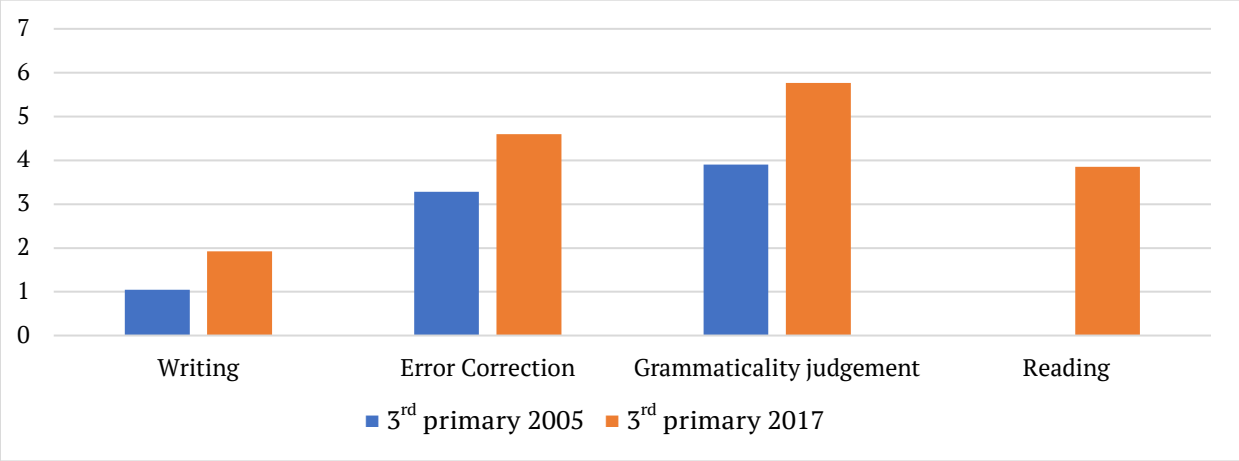
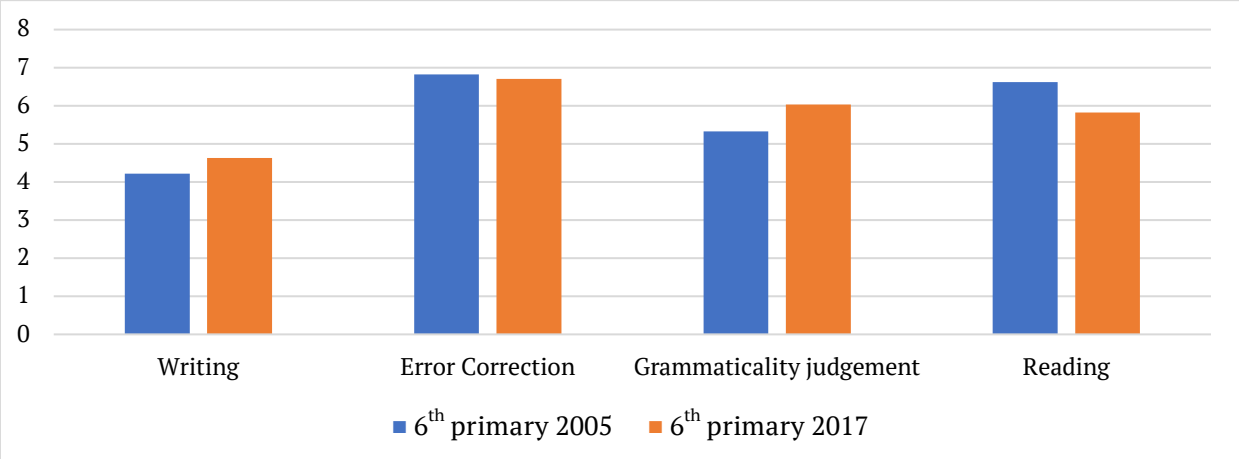


Table 3
Group statistics for the sixth graders in 2005 (Group 3) and 2017 (Group 4)

Group		N	Mean	Std. Deviation	Std. Error	P-value
Writing test	3	30	4.217	1.0722	.1958	.128
	4	30	4.633	1.0165	.1856	.128
Error correction test	3	30	6.823	1.4689	.2682	.729
	4	30	6.700	1.2630	.2306	.729
Grammaticality judgement test	3	30	5.333	1.0450	.1908	.026
	4	30	6.033	1.3126	.2397	.026
Reading test	3	30	6.617	1.7003	.3104	.106
	4	30	5.817	2.0615	.3764	.107

Figure 2
Test results for the sixth year of primary at both testing times



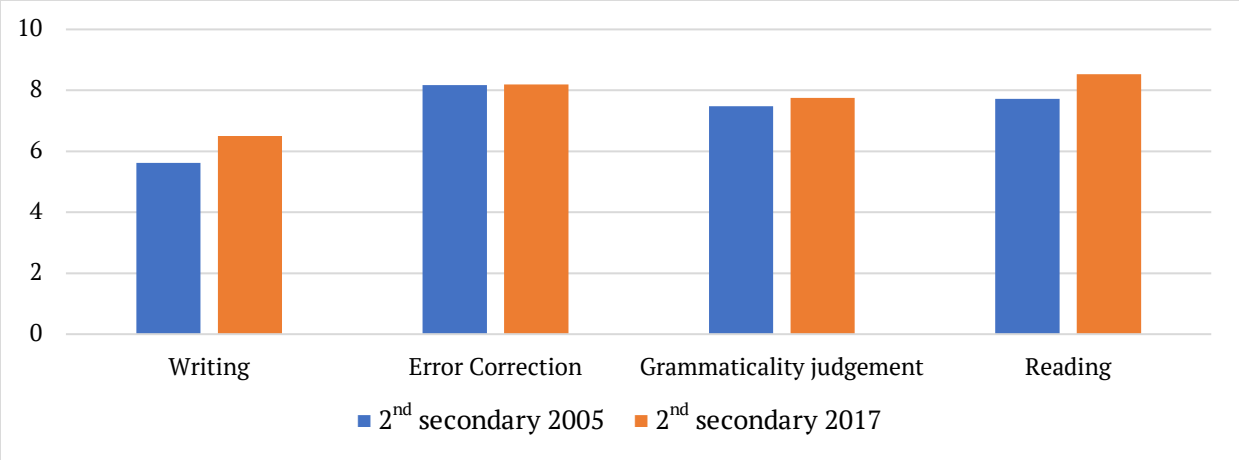
When analysing the results of participants in the second grade of secondary, as displayed in Table 4 and Figure 3 below, higher grades were reported in all four tests in 2017, although in the case of the error correction test, the difference was very small: the mean in 2005 was 8.17, while in 2017 it was of 8.19, thus being slightly higher in 2017, however not statistically significant (p=0.909). On the writing test, the mean score in 2005 was 5.62; while in 2017 it was 6.50, a difference that was statistically significant (p=0.032). Concerning the grammaticality judgement test, the mean reported in 2005 was 7.48; while in 2017 it was 7.75, the difference not being

statistically significant (p=0.353). Finally, for the reading test, the mean reported in 2005 was 7.72; while in 2017 it was 8.53, with a statistically significant difference (p=0.013).

Table 4
Group statistics for the second year of secondary in 2005 (Group 5) and 2017 (Group 6)

Group		N	Mean	Std. Deviation	Std. Error	P-value
Writing test	5	30	5.617	1.5125	.2761	.032
	6	30	6.500	1.5920	.2907	.032
Error correction test	5	30	8.170	.7479	.1366	.909
	6	30	8.193	.8166	.1491	.909
Grammaticality judgement test	5	30	7.483	1.2490	.2280	.353
	6	30	7.750	.9354	.1708	.353
Reading test	5	30	7.717	1.5011	.2741	.013
	6	30	8.533	.8996	.1642	.014

Figure 3
Test results for the second year of secondary at both testing times



In sum, the previous results prove that the first hypothesis can only be partially confirmed. The 2017 cohort showed higher results than the 2005 cohort on most tests. Two exceptions must be noted: the sixth graders in primary school whose error correction test and reading test in 2005 yielded higher results; and the second graders in secondary whose error correction test scores were almost the same at both data collection times.

The second research question enquired into the gains shown cross-sectionally between the third and sixth graders in primary; and those same sixth graders and second graders in secondary. It was hypothesized that gains in the 2017 cohort would be higher than those reported for the 2005 cohort.

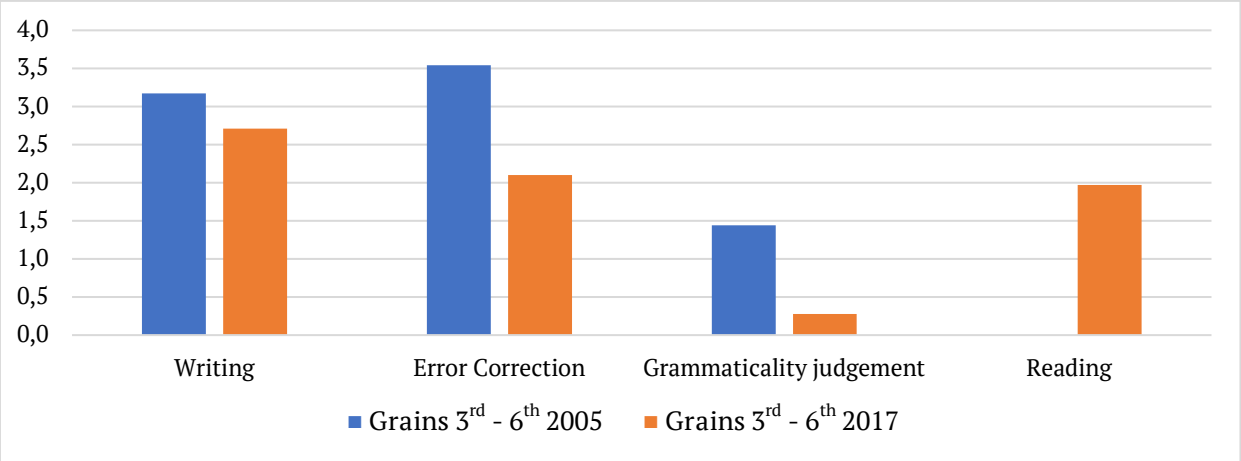
When comparing third and sixth graders in primary, significantly better results were obtained by the 2005 cohort in the case of the error correction test and the grammaticality judgement test, and a statistical tendency showed in writing[5]. Indeed, the results for the writing test, as displayed in Table 5 and Figure 4 below, reveal that the gains obtained in 2005 were 3.17 out of 10; while in 2017, they were 2.71, but the difference was not statistically significant (p=0.208). On the error correction test, the gains in 2005 were 3.54 points; while they were 2.10 in 2017, such a difference being statistically significant (p=0.014). On the grammaticality judgement test, the gains in 2005 were 1.44 points; while in 2017, they were 0.28, with a statistically significant difference (p=0.027).

[5] As mentioned above, no reading test was administered to the third graders in 2005, for technical reasons, so no comparison is possible in this respect.

Table 5
Mean gains in all tests between the third graders and the sixth graders in primary

Variables	Group	Gains	P-value
Writing test	2005 2017	+3.17 +2.71	.208
Error correction test	2005 2017	+3.54 +2.10	.014
Grammaticality judgement test	2005 2017	+1.44 +0.28	.027
Reading test	2005 2017	 +1.97	

Figure 4
Gain scores between third grade and sixth grade in primary, for both cohorts (2005 and 2017)



We find the reverse pattern when contrasting those same sixth graders and second graders in secondary. Table 6 and Figure 5 below show that between sixth grade in primary and second grade in secondary, a tendency towards greater gains for the 2017 testing time accrued, reaching significance only in the case of the reading test. On the writing test, the gains were 1.40 points in 2005; while they were 1.87 in 2017, a difference that was not statistically significant (p=0.301). On the error correction test, the gains in 2005 were 1.35; while they were 1.49 in 2017, such a difference not being statistically significant (p=0.712). On the grammaticality judgement test, the gains were 2.14 in 2005; while in 2017 they were 1.71, without statistical significance (p= 0.277). Finally, on the reading test, the gains in 2005 were 1.09; while they were of 2.71 in 2017, a difference that was statistically significant (p=0.005).

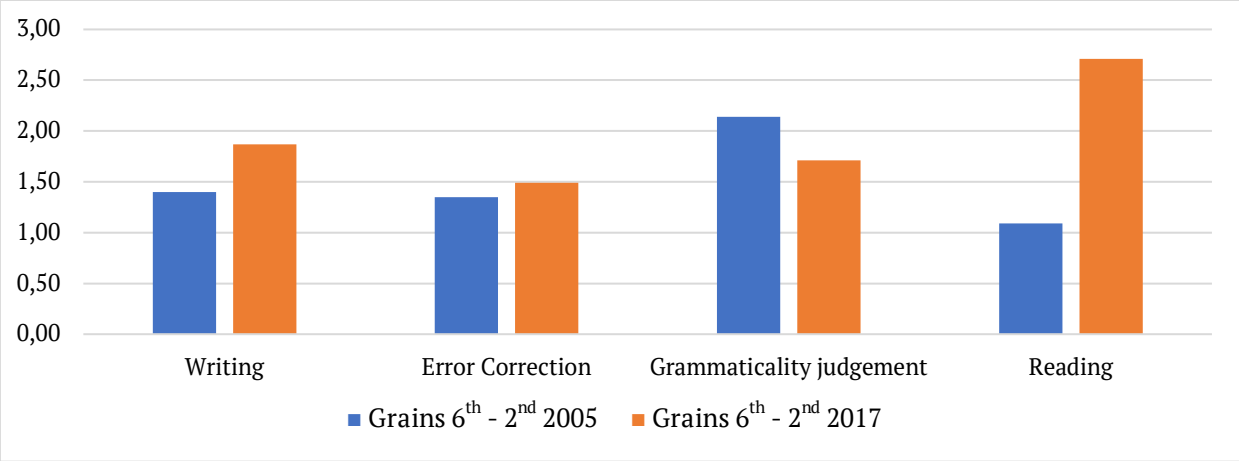
Table 6
Mean gains in all tests between the sixth graders in primary and the second graders in secondary

Variables	Group	Gains	P-value
Writing test	2005 2017	+1.40 +1.87	.301
Error correction test	2005 2017	+1.35 +1.49	.712
Grammaticality judgement test	2005 2017	+2.14 +1.71	.277
Reading test	2005 2017	+1.09 +2.71	.005

By examining the previous results, it can be stated that the second hypothesis cannot be confirmed. Significantly larger gains were obtained from the third to sixth grade of primary by the 2005 cohort on the

error correction and the grammaticality judgement tests, and a statistical tendency towards significance was shown in writing (reading was not measured). In contrast, the pattern seemed to change when we compared the sixth graders in primary and the second graders in secondary: significantly larger gains were found for the 2017 cohort in the case of reading and a tendency in the case of the writing, the error correction. Nevertheless, it has to be noted that in 2017, the sixth graders in primary showed lower results on the error correction and reading tests than those reported in 2005 (see Table 3). This can explain why, in the case of reading, they showed higher gains in this comparison, they had more room for improvement given their lower initial level.

Figure 5
Gain scores between the sixth graders in primary and the second graders in secondary for both cohorts (2005 and 2017)



The third and last research question enquired into the possible gains between the initial testing time (third grade in primary) and the last testing time (second grade in secondary). It was hypothesized that the 2017 cohort would show larger gains than the 2005 cohort.

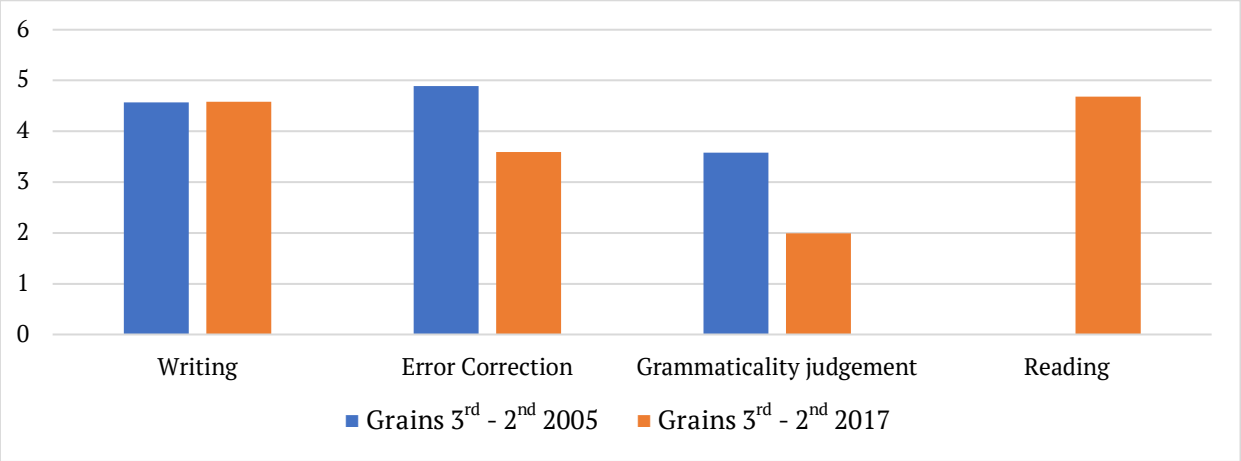
Table 7 and Figure 6 below display the results, which do not confirm our hypothesis. They were similar at both testing times for writing: in 2005 gains reached 4.57; while in 2017, they were 4.58, and the difference was not statistically significant (p=0.972). In the case of the error correction test, higher gains were obtained in 2005, 4.89 points; than in 2017, when they were 3.59, such difference being statistically significant (p=0.003). The same was true in the case of the grammaticality judgement test, for which the gains obtained in 2005 reached 3.58, also a superior figure than in 2017, 1.99, a difference that was also statistically significant (p=0.006). As for the reading test, in 2005, third graders did not take this test.

Table 7
Mean gains in all tests between the third graders in primary and the second graders in secondary

Variables	Group	Gains	P-value
Writing test	2005	+4.57	.972
	2017	+4.58	
Error correction test	2005	+4.89	.003
	2017	+3.59	
Grammaticality judgement test	2005	+3.58	.006
	2017	+1.99	
Reading test	2005		
	2017	+4.68	

In conclusion, the third hypothesis cannot be confirmed either, since for both the error correction test and the grammaticality judgment test, significantly larger gains were obtained in 2005 than in 2017, with a statistical tendency shown in the case of the writing.

Figure 6
Gain scores between initial testing time and the last testing time for both cohorts (2005 and 2017)



Discussion

The current study has aimed to contrast the benefits of a FI+CLIL EFL programme at two different points in time. It sought to examine whether learners' linguistic proficiency proved to be higher at the time when the programme was no longer at its onset, that is twelve years after it was launched. It was assumed that more than a decade should have resulted in a well-established learning context no longer at the pilot stage, and, as a result, was potentially more efficient.

Most studies analysing the benefits of CLIL have been conducted with programmes that are in their pilot stages (Moore, 2009). It is therefore of utmost interest to analyse whether the infancy of the programme may negatively impact its outcomes by comparing the same programme years after its launching. A well-stablished and robust programme may be expected to provide better learning conditions by guaranteeing greater stability and complexity, leading to higher results in the students' linguistic proficiency (Moore, 2009; Navarro-Pablo & López Gándara, 2019).

Therefore, the first research question in the present study enquired into the EFL linguistic benefits of the programme by comparing learner proficiency at three different learner ages with a cross-sectional population. It was hypothesized that learners enrolled in the 2017 programme would show higher proficiency than their peers in the 2005 programme on the four tests administered, tapping into writing, grammar and vocabulary, and reading. This first hypothesis has been only partially confirmed, since participants in the 2017 cohort showed significantly higher grades on 8 out of 11 tests: all tests in the third grade in primary; writing and the grammaticality judgement test in the sixth grade of primary; and writing, the grammaticality judgement test, and reading in the second grade in secondary. On the contrary, on three occasions higher results were found in 2005 than in 2017: primary sixth graders on the error correction and reading tests, and secondary second graders on the error correction test.

Previous CLIL research has also reported contradictory results regarding the linguistic benefits of the programmes scrutinized with respect to syntax and writing (Dalton-Puffer, 2008; Pérez-Vidal, 2011). In the present study, the 2017 cohort yielded higher results in writing than the 2005 for the three age groups tested with a statistically significant difference for two of them. Such results may be due to the longevity and robustness of the programme, which may allow students to benefit from a more complex programme and more experienced teachers. As for syntax, out of the six grammar tests that were administered in total (error correction and grammaticality judgement in all three grades), higher results were reported in the 2017 cohort in four, with the exception of the error correction tests for sixth graders in primary and second graders in secondary. Finally, vocabulary has been found to be a skill benefitting from CLIL (Dalton-Puffer, 2008; Pérez-Vidal & Roquet, 2015). As for reading, higher results were found in the present study in older students (2nd secondary), while younger learners (6th primary) had higher results in the 2005 programme.

Against the previous backdrop, it seems evident that although most of the previous research in the field has focused on comparing CLIL and FI learners' language development, most of the skills that were previously found to benefit from CLIL have also showed higher results in the well-stablished programme analysed in the current study. Therefore, such results may mean that a more robust and stable CLIL programme could potentially lead to greater benefits in said skills.

Thus, these findings allow us to state that when groups from the early and the well-established programmes are compared at the ages of 8-9, 11-12, 13-14 and the same conditions are kept in terms of number of hours of instruction and pedagogical approach, learners' progress was higher in 2017 than in 2005 in most of the tests administered (8 out of the 11 tests). Indeed, the programme seems to have had a higher positive impact on learners' linguistic progress once fully established, twelve years after it was launched for most tests when raw scores are analysed, with the exception of the error correction and reading tests mentioned above. These results are in line with a previous study (Navarro-Pablo & López Gándara, 2019), in which higher results in the CLIL group were attributed to the seniority of the programme and teachers' expertise.

The second and third research questions approached the comparison of the impact of the CLIL programme when it first was launched and twelve years later, by measuring the gains in linguistic abilities between the three different grades tested. When comparing learners' gains between the third and sixth grades in primary at both testing times (2005 and 2017), much higher gains were seen in the 2005 cohort as well as significant gains on the error correction and grammaticality judgement tests. Hence the programme's length of time in operation does not seem to have yielded extra benefits in terms of linguistic gains. The opposite results were obtained when analysing the gains between the sixth grade in primary and the second grade in secondary, namely higher gains in 2017 for all tests except for the grammaticality judgement test, with a significant difference in reading and a statistical tendency towards significance in writing. Considering these findings, it cannot be claimed that learners will reap higher benefits from a well-established CLIL programme in terms of linguistic gains between the beginning, middle, and end points of the programme. However, there is a tendency towards greater improvement in the older students, who showed higher gains in 2017 than in the areas of writing, grammar, and reading.

Finally, when comparing the gains between the younger and the older groups, almost equal gains were reported on the writing test and higher gains were found in 2005 on the error correction and grammaticality judgement tests. As a result, the second and third hypotheses cannot be confirmed.

To our knowledge, no previous studies on the influence of programme length have analysed linguistic gains; however, one study (Navarro-Pablo & López Gándara, 2019) focused on analysing the effects of programme seniority using raw scores in English proficiency. Therefore, further research is needed in this domain to further examine the effects of programme length.

Conclusion

Overall, it can be stated that the quality of the programme has indeed improved since learners in the 2017 cohort had better results on most of the tests at all three educational levels when raw results are compared. Such improvement in the programme can be attributed to the programme's seniority, since the programme has been kept within the same parameters in terms of the number of hours, the CLIL coordinator, and the pedagogical approach followed. After 15 years of running, the programme in 2017 was more robust and stable, the teachers were more experienced, and its design and application more complex. This allowed students to benefit from a much more developed programme and attain higher levels in English as a FL.

In contrast, the analysis of gains does not seem to prove an effect of the programme's seniority, neither when contrasting the development between the three educational levels, nor between the initial and final testing times. Thus, it cannot be confirmed that a fully-stablished CLIL programme will lead to higher results in terms of the linguistic gains in the students enrolled in the programme, as compared to the programme in its pilot phase. Additionally, the absence of higher gains in 2017 may, however, be attributed to a ceiling effect: initial scores in 2017 were slightly higher on most tests both in the third and sixth grades of primary school, which were used as pre-test scores to analyse gains; therefore, starting levels of English were higher

in 2017. Thus, higher initial language levels may have left less room for improvement and, as a result, explain the lower gains.

However, this study has some limitations that may partially explain its results. The first and possibly the largest limitation is that the data used was not longitudinal, but cross-sectional. Although students' exposure and contact with English inside and outside the school was controlled though the sociolinguistic background questionnaire at both testing times, the cross-sectional nature of the study poses a limitation, since the raw scores and gains compared were from different samples of participants. Second, individual differences are considered to have an impact on FL learning; however, we did not take them into account due to time and space constraints. Only the mother tongue of the subjects and their contact with the English language outside of school was considered to make sure that none of the learners had English as their native language or had more hours of contact with the language beyond the school sessions. In addition, it must be noted that the amount of extramural exposure to English in 2017 was much higher than in 2005 as access to English resources accessible from home has increased in the last decade. Therefore, we cannot be completely sure that some of the higher results reported in 2017 were not also enhanced due to the fact that children nowadays have more access to the English language.

This study has sought to contribute with new, meaningful, and updated data to our understanding of the effects of a CLIL programme in mainstream education. In a nutshell, our results reveal that the programme's quality has indeed changed over the years, guaranteeing a higher level of EFL proficiency when raw scores are contemplated, but not when gains are analysed. This calls for further studies being conducted that might add longitudinal data to the cross-sectional data used in this study.

Acknowledgments

This research was supported by grants PGC2018-098815-B-I00 and 2017 SGR 1028. Marta Segura is a Universitat Internacional de Catalunya PhD grant student at the Institute for Multilingualism. We would like to thank the editors for their insightful comments.

Conflict of interests

The authors declare that they have no conflict of interest.

References

Admiraal, W., Westhoff, G., & de Bot, K. (2005). Evaluation of bilingual secondary education in the Netherlands: Students' language proficiency in English 1. *Educational Research and Evaluation 12*(1), 75-93. https://doi.org/10.1080/13803610500392160

APAC. (Ed.). (2006). *CLIL in Catalonia: From theory to practice*. Associació de Professors d'Anglès de Catalunya.

Auerbach, E. R. (2006). Reexamining English only in the ESL Classroom. *TESOL Quarterly*, 27(1), 9-32.https://doi.org/10.2307/3586949

Brinton, D., Wesche, M., & Snow, A. (Eds.). (1989). *Content-based second language instruction*. Neubury House Publishers.

Bruton, A. (2011). Is CLIL so beneficial, or just selective? Re-evaluating some of the research. *System, 39,* 523 – 532. https://doi.org/10.1016/j.system.2011.08.002

Cenoz, J. (2015). Content-based instruction and content and language integrated learning: The same or different? *Language, Culture and Curriculum*, *28*(1), 8–24. https://doi.org/10.1080/07908318.2014.1000922

Dafouz Milne, E., & Guerrini, M. C. (Eds.). (2009). *CLIL across educational levels: Experiences from primary, secondary and tertiary contexts*. Richmond Publishing – Santillana Educación S.L.

Dalton-Puffer, C. (2008). Outcomes and processes in Content and Language Integrated Learning (CLIL): Current research from Europe. In W. Delanoy, & L.Volkmann (Eds.), *Future perspectives for English language teaching* (pp. 139-157). Carl Winter.

Darvin, R., Yi Lon, Y., & Lin, A. M. Y. (2020). Examining CLIL through a critical lens. *English Teaching and Learning, 44*(2)*,* 103-108. https://doi.org/10.1007/s42321-020-00062-2

DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, 22(4), 499-533. https://doi.org/10.1017/S0272263100004022

Derakhshan, A., & Karimi, E. (2015). The Interference of first language and second language acquisition. *Theory and Practice in Language Studies, 5*(10), 2112-2117 http://dx.doi.org/10.17507/tpls.0510.19

De Dios Martínexz Agudo, J. (2019). The impact of CLIL on English language competence in monolingual context: A longitudinal perspective. *The Language Learning Journal, 48*(1), 36-47. https://doi.org/10.1080/09571736.2019.1610030

Fernández-Sanjurjo, J., Fernández-Costales, A., & Arias Blanco, J. M. (2017). Analysing students' content learning in science in CLIL vs. non-CLIL programmes: Empirical evidence from Spain. *International Journal of Bilingual Education and Bilingualism*, *22*, 661-674. https://doi.org/10.1080/13670050.2017.1294142

García, O. 2009. *Bilingual education in the 21st century: A global perspective*. Wiley-Blackwell.

García, O., Aponte, G.Y., & Le, K., (2020). Primary bilingual classrooms: Translations and translanguaging. In S. Laviosa & M. González Davies, (Eds.). *The Routledge handbook of translation and education* (pp. 81-94). Taylor and Francis. https://doi.org/10.4324/9780367854850

Gené Gil, M. (2016). *Efecto del AICLE sobre la producción escrita en estudiantes de secundaria de inglés como lengua extranjera* [CLIL effects on written production in secondary school students of English as a foreign language]. Universitat de les Illes Balears

Hughes, S. P., & Madrid, D. (2019. The effects of CLIL on content knowledge in monolingual contexts. *The Language Learning Journal*, *48*(1), 48-59. https://doi.org/10.1080/09571736.2019.1671483

Hüttner, J., & Smit, U. (2014). CLIL (Content and Language Integrated Learning): The bigger picture. A response to: A. Burton. (2013) CLIL: Some of the reasons why… and why not. System, 41, 587-597. *System, 44,* 160-167. https://doi.org/10.1016/j.system.2014.03.001

Jaén Campos, M. (2016). *Redefining the roles of the EFL teacher within CLIL secondary settings*. Universidad Complutense de Madrid.

Kersten, K., & Rohde, A. (2013). Teaching English to young learners. In A. Flyman Mattsson & C. Norrby (Eds.), *Language acquisition and use in multilingual contexts* (pp. 107–118). Lund University.

Lasagabaster, D. (2008). Foreign Language Competence in Content and Language Integrated Courses. *The Open Applied Linguistics Journal*, *1*, 31–42. https://doi.org/10.2174/1874913500801010030

Lasagabaster, D., & Doiz, A. (2017). A longitudinal study on the impact of CLIL on affective factors. *Applied Linguistics*, *38*, 688–712. https://doi.org/10.1093/applin/amv059

Lasagabaster, D., & Sierra, J. M. (2009). Language attitudes in CLIL and international EFL classes. *International CLIL Research Journal*, *2*(2), 4–17.

Llinares, A., & Morton, T. (2010). Historical explanations as situated practice in content and language integrated learning. *Classroom Discourse*, *1*(1), 46–65. https://doi.org/10.1080/19463011003750681

Lorenzo, F., Casal, S., & Moore, P. (2010). The effects of content and language integrated learning in European education: Key findings from the Andalusian bilingual sections evaluation project. *Applied Linguistics*, *31*(3), 418–442. https://doi.org/10.1093/applin/amp041

Moore, P. 2009. *On the emergence of L2 oracy in bilingual education: A comparative analysis of CLIL and mainstream learner talk* [Unpublished doctoral dissertation]. Universidad de Sevilla. https://doi.org/10.1093/applin/amp041

Navarro-Pablo, M., & López Gándara, Y. (2019). The effects of CLIL on L1 competence development in monolingual contexts. *The Language Learning Journal, 48*(1), 18–35. https://doi.org/10.1080/09571736.2019.1656764

Navés, T. (2009). Effective Content and Language Integrated (CLIL) programmes. In Y. Ruiz de Zarobe & R. M. Jiménez Catalán (Eds.), *Content and Language Integrated Learning. Evidence from research in Europe* (pp. 22-40). Multilingual Matters.

Navés, T., & Muñoz, C. (1999). CLIL experiences in Spain. In D. Marsh & G. Langé (Eds.), *Implementing Content and Language Integrated Learning*. Jyväskyla.

Nikula, T., & Moore, P. (2015). Exploring translanguaging in CLIL. *International Journal of Bilingual Education and Bilingualism, 22,* 237-249. https://doi.org/10.1080/13670050.2016.1254151

Papaja, K. (2013). The role of a teacher in a CLIL classroom. *Glottodidactica An International Journal of Applied Linguistics*, *40*(1) 147-154. https://doi.org/10.14746/gl.2013.40.1.11

Pérez-Vidal, C. (2007). The need for focus on form (FoF) in content and language integrated approaches: An exploratory study. *Revista Española de Lingüística Aplicada*, *1*, 39-54.

Pérez-Vidal, C. (2011). Language acquisition in three different contexts of learning: Formal instruction, study abroad and semi-immersion (CLIL). In Y. Ruiz de Zarobe, J.M. Sierra, & F. Gallardo del Puerto (Eds.), *Content*

and foreign language integrated learning: Contributions to multilingualism in European contexts (pp. 25-35). Peter Lang. https://doi.org/10.3726/978-3-0351-0171-3

Pérez-Vidal, C. (2013). Perspectives and lessons from the challenge of CLIL experiences. In C. Abello-Contesse, P. M. Chandler, M. D. López-Jiménez, & R. Chacón-Beltrán (Eds.), *Bilingual and multilingual education in the 21st century building on experience* (pp. 59-82). Multilingual Matters. https://doi.org/10.21832/9781783090716

Pérez-Vidal, C., Escobar, C., & Roquet, H. (2013). Perspectives sobre l'aprenentatge de llengües i multilingüisme [Perspectives on language learning and multilingualism]. *Butlletí Informatiu del Col·legi de Doctors i Llicenciats en Filosofia i Lletres i en Ciències de Catalunya, 14*, 4-5.

Pérez-Vidal, C., & Juan-Garau, M. (2012). To CLIL or not to CLIL? From bilingualism to multilingualism in Catalan/Spanish communities in Spain. In Y. Ruiz de Zarobe, & D. Lasagabaster (Eds.), *CLIL in Spain: Implementation, results and teacher training* (pp. 115–138). Cambridge Scholars Publishing.

Pérez-Vidal, C., Lorenzo, N., & Trenchs, M. (2015). Les llengües en l'educació: El plurilingüisme i la internacionalització [Languages in education: plurilingualism and internationalization]. In J. M. Vilalta (Ed.), *Reptes de l'educació a Catalunya. Anuari 2015* [Education challenges in Catalonia. Annual report 2015] (pp. 139-197). Fundació Jaume Bofill.

Pérez-Vidal, C., & Roquet, H. (2015). The linguistic impact of a CLIL Science programme: An analysis measuring relative gains. *System, 54,* 80-89. https://doi.org/10.1016/j.system.2015.05.004

Pladevall-Ballester, E., & Vallbona, A. (2016). CLIL in minimal input contexts: A longitudinal study of primary school learners' receptive skills. *System, 58*, 37-48. https://doi.org/10.1016/j.system.2016.02.009

Rallo Fabra, L., & Jacob, K. (2015). Does CLIL Enhance Oral Skills? Fluency and Pronunciation Errors by Spanish-Catalan Learners of English. In M. Juan-Garau & J. Salazar-Noguera (Eds.), *Content-based language learning in multilingual educational environments (Edition: Educational Linguistics)* (pp. 163–177). Springer. https://doi.org/10.1007/978-3-319-11496-5_10

Rallo Fabra, L., & Juan-Garau, M. (2011). Assessing FL pronunciation in a semi-immersion setting: The effects of CLIL instruction on Spanish-Catalan learners' perceived comprehensibility and accentedness. *Poznan Studies in Contemporary Linguistics*, *49*(1), 96–108. https://doi.org/ 10.2478/psicl-2011-0008

Roquet, H. (2011). *A study of the acquisition of English as a foreign language integrating content and language in mainstream education in Barcelona*. Universitat Pompeu Fabra.

Roquet, H., & Pérez-Vidal, C. (2017). Do productive skills improve in content and language integrated learning contexts? The case of writing. *Applied Linguistics*, *15*(3), 489-511. https://doi.org/10.1093/applin/amv050

Ruiz de Zarobe, Y. (2008). CLIL and foreign language learning: A longitudinal study in the Basque country. *International CLIL Research Journal, 1*(1), 60-73

San Isidro, X., & Lasagabaster, D. (2019). The impact of CLIL on pluriliteracy development and content learning in a rural multilingual setting: A longitudinal study. *Language Teaching Research,* 23(5), 584-602. https://doi.org/10.1177/1362168817754103

Van de Craen, P., & Pérez-Vidal, C. (2001). *The multilingual challenge, Le Défi Multilingue* [The multilingual challenge]. Printulibro Intergrup, S.A.

National Research University Higher School of Economics
Journal of Language & Education Volume 7, Issue 1, 2021

Subon, F., & Unin, N. (2021). The Efficacies of an Explicit Vocabulary Instruction Model on ESL Learners' Vocabulary Size and Writing Skills. *Journal of Language and Education, 7*(1), 190-205. https://doi.org/10.17323/jle.2021.10209

The Efficacies of an Explicit Vocabulary Instruction Model on ESL Learners' Vocabulary Size and Writing Skills

Frankie Subon[1], Norseha Unin[2]

[1]Management and Science University

[2]Universiti Teknologi MARA Cawangan Sarawak

Correspondence concerning this article should be addressed to Dr. Frankie Subon, School of Education and Social Sciences, Management and Science University, 40100 Shah Alam, Malaysia.
E-mail: frankie_subon@msu.edu.my

Vocabulary size is believed to have positive correlations with learners' language skills, reading comprehension, knowledge acquisition, and academic achievement. In Malaysia, the lack of vocabulary has been identified as one of the causes for students' inability to acquire a second language. This quasi-experimental study examined the efficacies of employing the Contextualized Word Family (CWF) Model for direct vocabulary instruction (DVI) on vocabulary size and writing skills among secondary school students based on three proficiency levels. Through a purposive sampling method, 143 Grade 8 students were classified into three proficiency groups and were administered a similar treatment of contextualized word families. The Productive Vocabulary Level Test (PVLT) was used to measure the participants' vocabulary size and guided essay writing tests were employed to obtain their test scores in the pre-test and post-test. The findings revealed the positive effects of using the CWF Model for enhancing students' vocabulary size and writing ability. After a ten-week intervention, the majority of the students from low, intermediate, and high proficiency groups were able to increase their vocabulary size from below 1000 words to 1000-2000 words. There was also an increase in the number of students who were able to reach the 2000-word level of vocabulary size for all three groups. The findings from this study ascertained the positive effects of using the CWF Model to boost not only the vocabulary size, but also the writing skills of the high and intermediate proficiency students. On the writing tests, the high proficiency students showed a significant increase while the intermediate proficiency students experienced a slight increase in their writing test scores, but it was insignificant. Surprisingly, the low proficiency students experienced a significant decrease in their writing test scores after the treatment despite experiencing an increase in their vocabulary size. Future studies are recommended to utilize a true experimental design with a longer treatment period especially for examining the efficacies of the model on students' writing skills.

Keywords: vocabulary acquisition, vocabulary knowledge, contextualized word families, vocabulary size, explicit vocabulary instruction

Introduction

Vocabulary knowledge plays an essential role in literacy development and academic success. It supports the mastering of reading, speaking, listening, and writing skills (Al-Khasawneh, 2019; Silsüpür, 2017). Scholars have acknowledged the importance of vocabulary in second language (L2) learning and language use. In fact, past studies (e.g., August, Carlo, Dressler, & Snow, 2005; National Reading Panel, 2000; Tosuncuoğlu, 2015) claim that vocabulary acquisition is vital for language learning because vocabulary encompasses all the lexical items language learners need to know in order to access their schemata, express ideas, communicate effectively, and learn about new concepts (Sedita, 2005). Moreover, there is a strong relationship between learners' vocabulary knowledge and their reading comprehension and future reading success (Cunningham & Stanovich, 1997; Senechal, Oulette, & Rodney, 2006; Xie et. al. 2019). Therefore, teachers are encouraged to train students to model an explicit vocabulary learning strategy to boost vocabulary acquisition. However, previous studies on vocabulary learning and acquisition had not come to a definite finding for an effective strategy or a structured,

direct form of vocabulary instruction. Hence, this quasi-experimental research aims to examine the efficacies of administering direct vocabulary instruction (DVI) by employing the Contextualized Word Family (CWF) Model (Subon, 2016) on the learners' vocabulary size and writing ability.

This study utilizes the CWF Model (Subon, 2016) as a strategy for explicit vocabulary instruction. There are two main rationales for employing this model. First, it consists of eight pertinent principles for effective DVI: giving definitions of the new words, teaching vocabulary in context, teaching tier-two words, teaching vocabulary in word families, using non-linguistic representation/symbols/pictures, providing multiple exposures, using sentences that construct a story, and integrating vocabulary learning and the acquisition of word families with contexts and a variety of language learning activities (Subon, 2018). Unlike other vocabulary instruction models in the past, e.g., the Frayer Model (Frayer, Frederick, & Kalausmeier, 1969), Grave's Visionary Model (Grave, 2000), and the STAR Model (Blachowicz, 2005) that utilize individual word in their vocabulary learning, the CWF Model (Subon, 2016) emphasizes learning words by groupings, or what is commonly known as word families.Secondly, the CWF Model exemplifies rich, systematic, and structured vocabulary instruction that integrates vocabulary learning with contexts. This is to enable learners to understand the meaning of a word more effectively by embedding it within a rich context of supportive and indicative information (Stahl & Fairbanks, 1986). The word family members of this model are adapted from Bauer & Nation's (1993) level 2 (inflectional suffixes) words. They are formed based on the different parts of speech, namely the base or plural verbs (teach), singular verbs (teaches), the past tense (taught), past participles (taught), nouns (teacher), adjectives (teachable), and present participles (teaching). The CWF Model consists of seven steps for teaching and learning a word family, as shown in Table 1 and Figure 1.

For this study, the three research questions are: 1) Does direct vocabulary instruction using the CWF Model increase learners' vocabulary size? 2) Does direct vocabulary instruction using the CWF Model result in significant difference in learners' writing ability? and 3) Is there any positive correlation between learners' vocabulary size and writing scores after direct vocabulary instruction using the CWF Model? This study also seeks to test one hypothesis (Hypothesis 1): There is a positive correlation between learners' vocabulary size and writing scores after the direct vocabulary instruction using the CWF Model.

Problem Statement

The lack of vocabulary has been identified as one of the factors contributing to learners' inability to acquire L2 (Kaur & Kabilan, 2007; Kaur, Othman, & Abdullah, 2012; Zakaria, 2005). In English deficit environments such as Malaysian schools, the issue of low language proficiency is still prevalent because of an underlying factor associated with vocabulary size. Although the apparent lack of proficiency among L2 leaners has been corroborated with a paucity of systematic, intentional vocabulary teaching and learning for ESL learners (e.g., Biemiller, 2005; Durkin, 1979; Dutro & Moran, 2003; Gersten & Baker, 2000; Moore, 2010; Scarcella, 1996; Scott & Nagy, 1997), there have been relatively few studies on explicit vocabulary instruction using word families (Nadarajan, 2009; Schmitt, 2008). Although some studies have been conducted on DVI, these past studies usually focused on learning vocabulary as individual words or teaching words in isolation. Furthermore, limited studies on DVI have been conducted among early middle school students between the age of 12-14 years old (Mast, 2011), as well as among young learners at the elementary level. Hence, this study was conducted to explore new and effective strategies for vocabulary intervention in L2 classrooms.

Literature Review

Acquiring an ample vocabulary size is important for learners to be competent in English as a second (ESL) or foreign language (EFL). Although there exists a dire need to increase learners' vocabulary size for English proficiency, limited research has been carried out on ways to increase learners' vocabulary size in Malaysia (Haris & Yunus, 2018). More empirical data are needed on learners' vocabulary size for educators and practitioners to address limitations in second language acquisition (SLA). Studies by Asgari and Mustapha (2011; 2012) and Haris and Yunus (2018) suggest that a limited vocabulary size affects the acquisition of English as an L2. When learners are actively acquiring new words through various learning experiences, e.g., engaging with reading texts in the teaching and learning activities, they will be able to increase their vocabulary size. Hence, L2 classrooms should include explicit vocabulary learning for students to receive the comprehensible

input of new vocabulary words. Arguably, the gradual increase in vocabulary size can enhance the breadth and depth of lexical knowledge, as well as proficiency in ESL.

A positive relationship between vocabulary size and language skills has also been supported by past studies (e.g., Albrechtsen, Haastrup, & Henriksen, 2008; Alderson, 2005; Laufer, 1992; Laufer & Goldstein, 2004; Laufer & Sim, 1985).According to Laufer and Sim (1985) and Qian (2002), vocabulary size is an important prerequisite for learners to function as an efficient reader and achieve success in the academic field (as cited in Tschirner, 2004). While some scholars (Albrechtsen, Haastrup, & Henriksen, 2008; Laufer, 1992; Laufer & Goldstein, 2004) corroborated a close relationship between vocabulary size and reading, Alderson (2005) ascertained a strong relationship with other language skills such as reading, writing, listening, and grammar. Hence, 'language ability is, to quite a large extent, a function of vocabulary size' (ibid. p. 88). Additionally, learners need to possess an ample vocabulary in terms of word families to perform well in all language skills (Schmitt, 2008; Waring, 2002). A 'word family' is a group of words that share the same basic meaning. For example, 'succeed' is in the same word family as success, succeeds, succeeding, succeeded, successful, and successfully. For reading and writing, learners will need 3,000-word families for reading authentic texts (Schmitt, 2007) and a minimum of 5,000-word families is required for the purpose of deeper processing (Thornbury, 2002). Spoken discourse requires a lower threshold of vocabulary i.e., 2,000-word families to cover about 95% of a typical speech (Adolphs & Schmitt, 2003). On that note, L2 learners need a minimal vocabulary size of 2,000 high-frequency words to understand about 80% of a running text (Nemati, 2010). Therefore, an ample vocabulary size is vital for ESL learners to be proficient speakers or users of the L2.

However, there are limited studies on the effects of vocabulary instruction on students' writing skills. Some of these past studies (e.g., Duin & Graves, 1987; Hestad, 2014; Solati-Dehkordi & Salehi, 2016; Yonek, 2008) revealed some positive findings. Duin and Graves (1987) examined the effects of teaching vocabulary during pre-writing on students' writing quality. They found a significant increase in the number of target words used in writing, the overall quality of the writing, and on measures of vocabulary knowledge. They concluded that teachers could help students improve the quality of their writing by administering rich vocabulary instruction. Yonek's (2008) study supports the finding by Duin and Graves (1987) that rich instruction is more effective for increasing students' word knowledge. As a matter of fact, explicit vocabulary instruction allows learners to utilize newly learned words in complex literacy skills such as writing.

Additionally, Hestad (2014) conducted a study to evaluate the effects of targeted vocabulary instruction on writing fluency. He found that vigorous vocabulary instruction can improve writing fluency. By using a reading comprehension activity and explicit vocabulary instruction, Solati-Dehkordi and Salehi (2016) examined the effects of vocabulary instruction on L2 learners' writing skills and vocabulary retention. They concluded that explicit vocabulary instruction followed by an immediate writing task enabled the newly learned words to become productive vocabulary and allowed for enhanced vocabulary retention. Considering the importance of possessing an ample vocabulary size, this study offers a systematic, structured, and rich vocabulary instruction method for L2 learners in the Malaysian context, which employs the CWF Model (Subon, 2016) of DVI to examine the efficacies on students' vocabulary size and writing skills.

Materials and Methods

Research Design

This study employed a quasi-experimental design without any control group because the research was conducted under normal school conditions whereby the students remained in their own classrooms during the treatment period. Furthermore, the use of this design was to avoid disrupting the school's on-going programs such as extra-curricular activities. Since it was conducted under normal conditions in a natural setting, this design was more likely to have external validity (Seliger & Shohamy, 2008).

Participants

The participants were selected from a rural secondary school in Sarawak, Malaysia. A purposive sampling method was used to select 143 Grade 8 students from four classes.They were classified into three proficiency

groups (High, Intermediate, and Low) based on the grades they obtained in the English subject on the school examination that tested reading and writing skills. This examination paper that is usually prepared following the format for lower secondary public examination determines their proficiency levels. Based on this examination, students who scored 80-100% were classified as the High Proficiency Group (42 students), those who obtained 60-79% formed the Intermediate Proficiency Group (58 students), and the Low Proficiency Group comprised of those who scored 35-59% (43 students). Since the same group was used for both the pre-test and the post-test, it was not necessary to match one group to another (Seliger & Shohamy, 2008).

Data Collection Methods

A quantitative method was used for the data collection. First, Laufer and Nation's (1999) Productive Vocabulary Levels (PVLT) Test[1] was adopted and administered in the pre-test and post-test to measure the students' vocabulary size. This test has been proven to be a reliable measure of vocabulary levels (Zimmerman, 2005). In this study, the vocabulary tests measured the 2000-word level for both the pre-test and the post-test. This word level contains the high-frequency words that all learners need to acquire in order to function effectively in English (Nation, 1990). The PVLT or a Check list test are ways to estimate vocabulary size that require learners to mark the words on a list that they believe they know (Hunt & Beglar, 1998). For example: Teenagers often adm ___ and worship pop singers. Each of the PVLTs consisted of 18 items. A score of 83% shows that a student has reached the 2000-word level for the vocabulary tests as specified by Laufer and Nation (1999). The test scores were rated in percentages and computed using SPSS version 22.0 for the data analysis.

The second assessment measured the effects of explicit instruction using the CWF Model on students' writing skills. Students were required to write a guided essay of not less than 120 words before (pre-test) and after (post-test) the treatment. The tests followed the format of guided essay writing for Grade 8 students, with short notes provided for the given topics. Some of the words included in the notes were the words that they had learnt using the CWF Model during the intervention. They were required to expand the notes into an essay of about 120 words. The generic analytic scale i.e., the ESL Composition Profile (Jacobs et al., 1981) was used to rate the essays. The scale consists of five ESL writing sub-constructs and each sub-scale carries a different weight: content is given the highest weight (30%), language use, organisation, and vocabulary have moderate weights (25%, 20%, and 20% respectively), while mechanics was the lowest (5%). The students' final scores on the pre-test and post-test were determined by averaging the scores awarded by the two raters. Finally, these average scores were computed using SPSS version 22.0 for the data analysis.

Data Collection Procedures

The English teachers who taught the four classes accepted the informed consent on the confidentiality of their involvement in the research. The four teachers have more than ten years of teaching experience and they agreed to participate in the research by administering the explicit vocabulary instruction.In order to ensure a smooth implementation of the vocabulary intervention, they were briefed on the seven steps for conducting the direct vocabulary instruction using the CWF Model during their English lessons by adhering to the fixed schedule. They were provided with a pocket file that contained the necessary instructional materials such as the research study schedule, glossary of tier-two target words and definitions, list of word families, samples of entries of the contextualized word families, and a scripted lesson plan (the steps for conducting the direct vocabulary instruction using the Contextualized Word Family Model). Informed consent was also obtained from the students' parents of the four classes to allow their children to participate in the research and they were informed about the purpose of the research. The students involved were also briefed about the purpose of the research, their roles as respondents, and the confidentiality of their involvement.

Two experienced English teachers were also appointed as raters for the guided essay writing tests. They have more than 20 years of teaching experience in teaching the English language in schools. They rated the guided essay writing scripts using the rubrics provided i.e., the ESL Composition Profile (Jacobs et al., 1981). One rater has the qualification of a Bachelor of Education in Teaching English as a Second Language (TESL) and the other has a Bachelor of Human Resource Management, with a subject option for teaching English. Usually, two raters are employed to evaluate the same samples and if they failed to rate consistently, a third rater would evaluate the

[1] From "Productive Vocabulary Levels Test (PVLT)" by B. Laufer & P. Nation (1999). http://www.lextutor.ca/tests/levels/productive/

samples again (Hamp-Lyons, 1990). To measure the inter-rater reliability, Pearson product moment correlation was applied. The average values of the two raters' scores for the pre-test and post-test were used for the data analysis.

The vocabulary instruction was conducted over ten weeks. Four English teachers and four classes were involved in the study. All three proficiency groups remained in their own classrooms. Every group was given the same treatment using the CWF Model immediately after the pilot study. First, the pre-test using the PVLT 2000-word level (Test A) and the guided essay writing (Test 1) were administered a day before the treatment started. This was followed by explicit vocabulary instruction or the treatment using the CWF Model for ten weeks. The seven steps (Table 1) for the teaching and learning of word families using the CWF Model were followed very closely for the explicit instruction of contextualized word families in this quasi-experimental study.

Table 1

The seven steps of the Contextualized Word Family Model

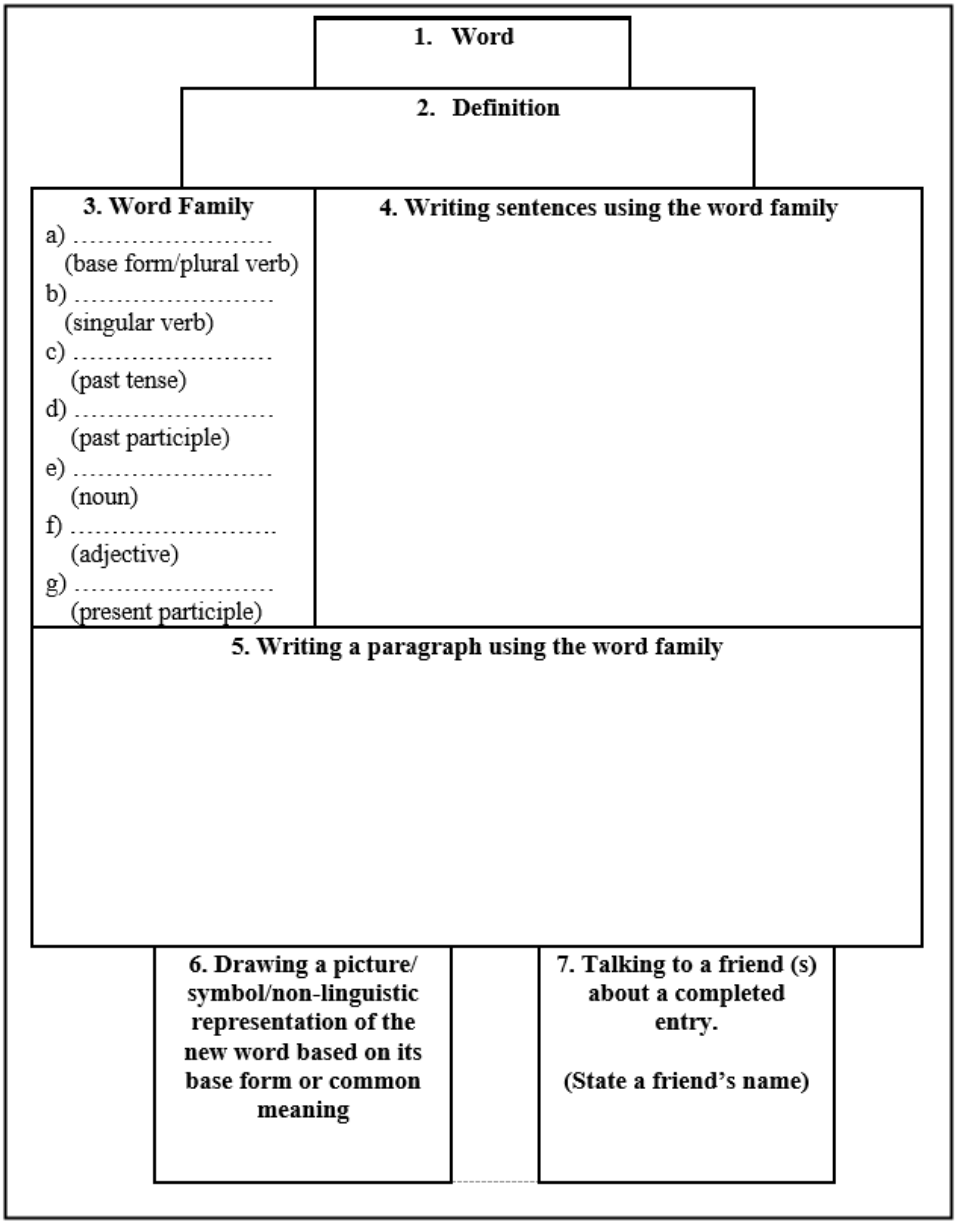
1.	Identifying a new tier-two word in a reading text
2.	Giving a definition or meaning of the word
3.	Listing the members of a word family
4.	Constructing meaningful sentences using the members of the word family to show contexts
5.	Writing a short paragraph using the members of the word family
6.	Drawing a picture or symbol (non-linguistic representation) of the word family based on its common meaning or base form
7.	Talking to friends about a completed entry of a word family

Note. Adapted from "The Contextualized Word Family Model of Direct Vocabulary Instruction" by F. Subon, 2018, IOSR Journal of Humanities and Social Science, 23, 57-70. Copyright 2016 by Frankie Subon.

During the treatment period, 30 sets of word families were taught to all three groups of learners, and the teaching and learning process took 30-35 minutes per lesson. The word families were from 30 tier-two words that were selected from reading texts used in the Grade 8 English textbook. Tier two consists of high-frequency words that are used across a variety of domains and are more characteristic of written language. They can be found across a variety of contexts and topics and understanding the meanings of these words promotes everyday reading and listening comprehension (Beck & McKeown, 1985). Besides that, these words carry multiple meanings and are often used in mature language discourse such as in adult conversations and literature. Tier-two words are the most significant words for direct instruction (Hutton, 2008) because they are good measures of a student's progress through school. Beck and her colleagues suggest teachers give the direct instruction of tier-two words to add productively to learner's language ability. Examples of tier-two words include obvious, fortunate, complex, measure, industrious, and establish. In English, there are about 7,000-word families (or 700 per year) in tier two (Hutton, 2008).

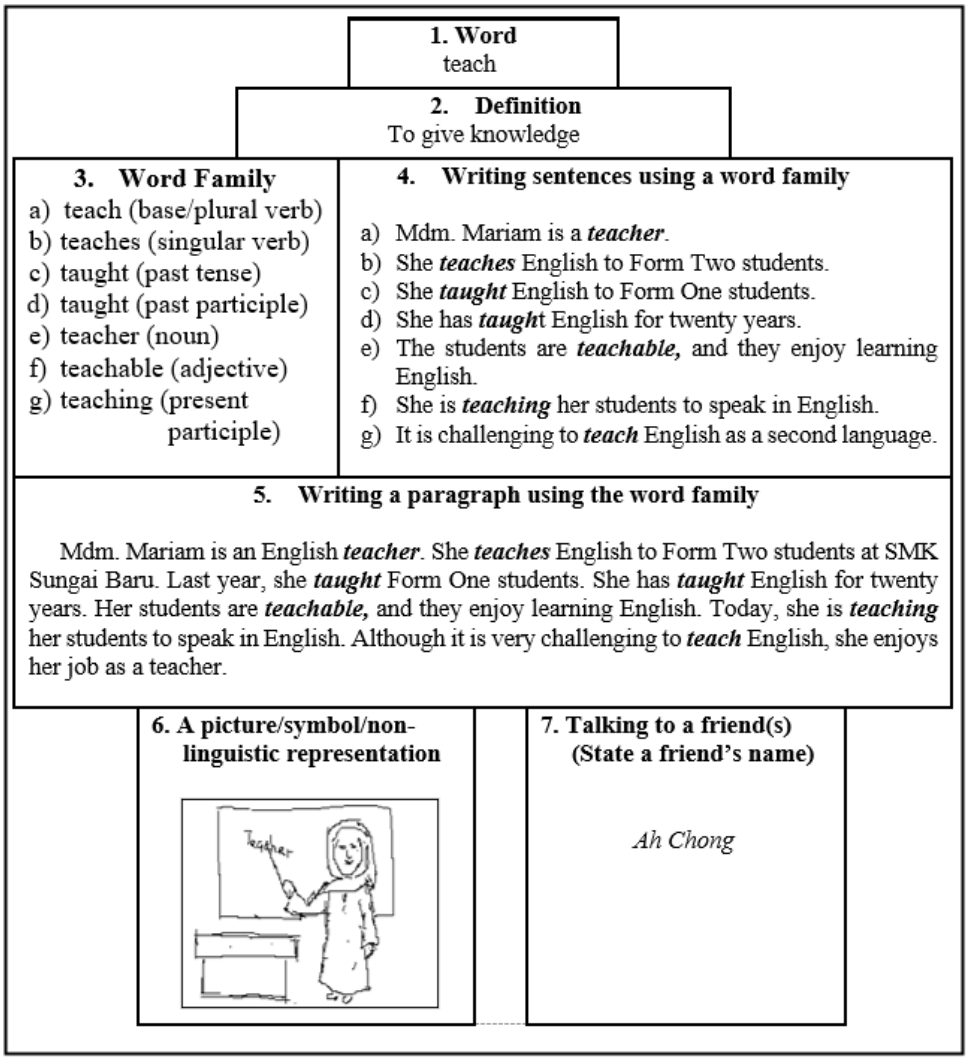
For this study, the tier-two words used in the intervention included conserve, guide, participate, develop, investigate, educate, etc. Students were given a pocket file equipped with a research study schedule, a checklist of tier-two target words, a sample of a completed entry of the contextualized word family (Figure 2), and 30 blank graphic organizers of the CWF Model on A4 paper. For each lesson, students were required to complete one graphic organizer (Figure 1) of the CWF Model (which consists of seven segments based on the seven steps of the model) and they were encouraged to use a dictionary when necessary. After the last session of the treatment, the post-test was administered using both instruments; PVLT 2000-word Level (Test B) and a guided essay writing test (Test 2).

Figure 1
The Contextualized Word Family Model of Direct Vocabulary Instruction.



Note. From "The Contextualized Word Family Model of Direct Vocabulary Instruction" by F. Subon, 2018, I*OSR Journal of Humanities and Social Science (IOSR-JHSS)*, 23, 57-70. Copyright 2016 Frankie Subon

Figure 2
An example of a completed entry of the Contextualized Word Family



Note. From "The Contextualized Word Family Model of Direct Vocabulary Instruction" by F. Subon, 2018, *IOSR Journal of Humanities and Social Science*, 23, 57-70. Copyright 2016 Frankie Subon

Data Analysis

The data obtained from the pre-test and post-test scores were coded for statistical analysis to respond to the research questions. For Research Question 1, descriptive statistics for frequencies were applied to obtain patterns for Productive Vocabulary Levels Test (PVLT) scores (both pre-test and post-test) to measure the vocabulary size based on proficiency levels. For Questions 2, inferential statistics (paired samples t-tests) were computed to determine whether there were significant differences in the writing test scores before and after direct instruction using the CWF Model. The paired samples t-test is used based on Chua's (2013) explanation that this test is used when two sets of data are obtained from the same subject group (one sample). Pearson correlation coefficients were run to answer Research Question 3 on whether there was a positive correlation between vocabulary size and writing scores after the treatment. The data obtained from the statistical analyses were tabulated and significant changes were analysed to generate the results of the study.

Results

Vocabulary Size based on Proficiency Levels

A minimum score to master a 2000-word vocabulary level is 15 out of 18 correct words or 83.3%, which is equivalent to 1670 words (Laufer & Nation, 1999). The data obtained from the pre-test and post-test were analysed using descriptive statistics i.e., frequency to show the students' vocabulary size according to their proficiency levels.

Vocabulary Size of the HIGH Proficiency Group

There were 42 students in the High Proficiency Group. Table 2 below shows the frequencies of the vocabulary scores on the pre-test and post-test.

Table 2

Frequencies of the High Proficiency Group's PVLT 2000-word level pre-test & post-test scores/vocabulary size

Test Score (%)/Vocabulary Size	Frequency (Pre-Test)	Percent (Pre-Test)	Frequency (Post-Test)	Percent (Post-Test)
1 (0-<25% / 0-<500)	1	2.4	0	0
2 (25-<50% / 500-<1000)	10	23.8	2	4.8
3 (50-<75 %/ 1000-<1500)	15	35.7	16	38.1
4 (75-<83.3% / 1500-<2000)	8	19.0	3	7.1
5 (83.3-100% /Mastered 2000)	8	19.0	21	50.0
Total	**42**	**100.0**	**42**	**100.0**

Half of the students in the High Proficiency Group were able to gain mastery of the PVLT 2000-word level after the treatment. The students had experienced positive vocabulary growth and many had mastered the PVLT 2000-word level after the direct instruction. The findings showed that the High Proficiency Group had achieved a marked increase in vocabulary size after the direct instruction as 21 out of 42 (50%) students had mastered the PVLT 2000-word level.

Vocabulary Size of the INTERMEDIATE Proficiency Group

The Intermediate Proficiency Group was made up of 58 students. Table 3 below shows the descriptive statistics of the PVLT pre-test and post-test scores of the Intermediate Proficiency Group.

Table 3

Frequencies of the Intermediate Proficiency Group's PVLT 2000-word level pre-test & post-test scores/vocabulary size

Test Score (%)/ Vocabulary Size	Frequency (Pre-Test)	Percent (Pre-Test)	Frequency (Post-Test)	Percent (Post-Test)
1 (0-<25% / 0-<500)	5	8.6	2	3.4
2 (25-<50%/ 500-<1000)	29	50.0	10	17.2
3 (50-<75% / 1000-<1500)	23	39.7	32	55.2
4 (75-<83.3% / 1500-<2000)	1	1.7	6	10.3
5 (83.3-100% /Mastered 2000)	0	0	8	13.8
Total	**58**	**50.0**	**58**	**100.0**

Table 3 shows that more students from the Intermediate Proficiency Group were able to obtain a vocabulary of between 1000 and 2000 words after the explicit vocabulary instruction and eight students were able to master the 2000-word level.

The Vocabulary Size of the LOW Proficiency Group

There were 43 students in the Low Proficiency Group. The descriptive statistics (Table 4) show the frequencies of the Low Proficiency Group's vocabulary scores on the pre-test and post-test.

Table 4

Frequencies of the Low Proficiency Group's PVLT 2000-word level pre-test & post-test scores/vocabulary size

Test Score (%)/Vocabulary Size	Frequency (Pre-Test)	Percent (Pre-Test)	Frequency (Post-Test)	Percent (Post-Test)
1 (0-<25% / 0-<500)	17	39.5	11	25.6
2 (25-<50% / 500-<1000)	17	39.5	12	27.9
3 (50-<75% / 1000-<1500)	7	16.3	14	32.6
4 (75-<83.3% / 1500-<2000)	1	2.3	3	7.0
5 (83.3-100% / (Mastered 2000)	1	2.3	3	7.0
Total	**43**	**100.0**	**43**	**100.0**

Table 4 above shows the low proficiency group of learners who had successfully increased their vocabulary size and the three students who had mastered the 2000-word level after direct instruction with the CWF Model.

The Writing Test Scores based on Proficiency Levels

In order to measure the inter-rater reliability, Pearson correlation coefficients were run for both the pre-test and post-test guided essay writing scores of the two raters. The inter-rater reliability coefficient was .83 for the pre-test scores and .86 for the post-test scores, implying a good level of reliability. Next, paired samples t-tests were computed to identify whether there were any significant differences in the writing skill test scores after the treatment of direct instruction using the CWF Model. The results of the data analyses for the three proficiency groups shown below.

The Writing Test Scores of the HIGH Proficiency Group

The paired-samples t-test for the High Proficiency Group's guided essay writing test scores is presented in Table 5.

Table 5

Paired samples t-test of the High Proficiency Group's guided essay writing test scores

Paired Samples Statistics

		M	N	SD	Std. Error M
Pair 1	"Guided Essay Writing(Post-Test)"	73.49	42	10.87	1.68
	"Guided Essay Writing(Pre-Test)"	64.02	42	12.43	1.92

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		M	SD	Std. Error M	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Guided Essay Writing (Post-Test) - Guided Essay Writing (Pre-Test)	9.46	11.97	1.85	5.73	13.19	5.12	41	.001

Students from the High Proficiency Group were able to increase their test scores for guided essay writing on the post-test. After the treatment through direct instruction using the CWF Model, the mean and standard deviation (*M*= 73.49, *SD*= 10.87) were better than before the treatment on the pre-test (*M* =64.02, *SD*= 12.43). There was a statistically significant mean increase of 9.47, 95% CI [5.73, 13.19], t (41) = 5.12, p = .001, d = .80. There was also a statistically significant difference in the mean of the guided essay writing test scores after the direct instruction using the CWF Model with the High Proficiency Group. Moreover, the difference had a high practical significance as it had a large effect size of d = .80.

The Writing Test Scores of the INTERMEDIATE Proficiency Group

The paired–samples t-test of Table 6 shows the mean difference between the pre-test and post-test of the guided essay writing test scores of the Intermediate Proficiency Group after direct instruction using the CWF Model.

Table 6

Paired samples t-test of the Intermediate Proficiency Group's essay writing test scores

Paired Samples Statistics

		M	N	SD	Std. Error M
Pair 1	"Guided Essay Writing(Post-Test)"	55.22	58	9.65	1.27
	"Guided Essay Writing(Pre-Test)"	52.97	58	9.53	1.25

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		M	SD	Std. Error M	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Guided Essay Writing (Post-Test) - Guided Essay Writing (Pre-Test)	2.25	11.93	1.57	-.88	5.39	1.44	57	.156

The intermediate proficiency students were able to increase their guided essay writing test scores on the post-test. The post-test score after the treatment (*M*= 55.22, *SD*= 9.65) was slightly better than before the treatment on the pre-test (*M* =52.97, *SD*= 9.53). There was a slight mean increase but not significant at 2.25, 95% CI [-.88, 5.39], *t* (57) = 1.44, *p* = .156, *d* = .19. This shows that there was no significant difference in the mean of the guided essay writing test scores after direct instruction using the CWF Model with the Intermediate Proficiency Group. However, the difference had low practical significance as it had an effect of d = .19, which did not meet the standard of even a small effect size (Cohen, 1988).

The Writing Test Scores of the LOW Proficiency Group

Table 7 below shows the paired samples t-test for the Low Proficiency Group after direct instruction using the CWF Model.

Table 7

Paired samples t-test of the Low Proficiency Group's essay writing test scores

Paired Samples Statistics

		M	N	SD	Std. Error M
Pair 1	"Guided Essay Writing(Post-Test)"	44.85	43	8.39	1.28
	"Guided Essay Writing(Pre-Test)"	47.93	43	7.79	1.19

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		M	SD	Std. Error M	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Guided Essay Writing (Post-Test) - Guided Essay Writing (Pre-Test)	-3.08	9.13	1.39	-5.89	-.27	-2.21	42	.032

Students from the Low Proficiency Group showed a decrease in their guided essay test scores on the post-test after the treatment of direct instruction using the contextualized word families (*M* = 44.85, *SD* = 8.39) than before the treatment on the pre-test (*M* = 47.93, *SD* = 7.79), a statistically significant mean decrease of -3.08, 95% CI [-5.89, -.27], *t* (42) =-2.21, *p* = .032, *d* = .34. This shows that there was a statistically significant difference in the mean of the guided essay test scores after direct instruction using the CWF Model with the Low Proficiency Group. However, the difference had low practical significance because it had a small effect size of d = .34.

The Correlation between Learners' Vocabulary Size and Writing Scores

The Pearson correlation coefficients were run to determine whether there was a positive correlation between learners' vocabulary size and writing scores after the direct instruction using the CWF Model. The result of the data analysis showed that there was a moderate positive correlation between the two variables with $r = .53$, $p = .001$. This reveals that when the vocabulary size increased the writing scores also increased. Therefore, we rejected the null hypothesis and accepted the alternative hypothesis (Hypothesis 1) that there was a positive correlation between learners' vocabulary size and writing scores after the direct vocabulary instruction using the CWF Model

Discussion

From the findings, it becomes clear that explicit vocabulary instruction using the CWF Model (Subon, 2016) had some positive effects on the vocabulary size of learners from all three proficiency groups. The descriptive analysis showed that they were able to demonstrate an increase in their vocabulary size ranging from 1000 to 2000 words after a ten-week intervention on the contextualized word families. There was also an increase in the number of participants who were able to master the 2000-word level after the intervention. The High proficiency group experienced a satisfactory increase in the mastery of the 2000-word level from only eight out of 42 students on the pre-test to 21 students on the post-test while the Intermediate and the Low proficiency groups experienced a slight increase from none to eight out of 58 students and one to three out of 43 students respectively. Hence, this data confirms the findings from previous studies (e.g., Ghapanchi, Eskandari, & Tabasi, 2012; Nadarajan, 2009; Naeimi & Foo, 2015; Paribakht & Wesche, 1997; Raasch, 2018; Schmitt, 2008; Soureshjani, 2011) that explicit vocabulary learning strategies have a significant effect on vocabulary acquisition of language learners. The findings are also consistent with past studies that showed teaching vocabulary in word families can lead to a greater growth in learners' vocabulary size (e.g., Beck & Mckeown, 1991; Nadarajan, 2009; Schmitt, 2008; Schmitt & Zimmerman, 2002; Soureshjani, 2011; The National Reading Panel, 2000). It has become more evident that the use of explicit vocabulary instruction enables learners to boost vocabulary growth as well as cultivate a productive habit for exploring word derivatives (Schmitt, 2000). Additionally, by employing word families for vocabulary instruction, this study sheds light on the potential of the CWF model for addressing the lack of structured vocabulary instruction, especially for increasing vocabulary size and improving writing skills among middle schoolers.

The findings also demonstrated some positive efficacies on learners' writing skills. A moderate positive correlation between learners' vocabulary size and writing scores was evident after the intervention with the CWF Model. The increase in the learners' vocabulary size after the treatment contributed to some positive effects on the learners' writing ability, especially among the high proficiency group. They showed an increase in the mean score after the treatment. The intermediate group also experienced a slight increase in the mean of their guided essay writing post-test scores after the treatment, although it was statistically insignificant. The increase in the marks for vocabulary, language use, and organization resulted in the increase in writing scores. This is consistent with the finding by Yonek (2008), who found that rich vocabulary instruction was more effective in the application of newly learned words to their expository writing than traditional instruction. This finding also ascertained that rich vocabulary instruction utilized (1) both definitional and contextual information for breadth of knowledge; (2) multiple exposures; and (3) active or deep processing of words (Beck, McKeown, & Kucan, 2002; Stahl & Fairbanks, 1986) was found to positively affect comprehension. After the intervention, students' writing skills also improved to a certain extent, as ascertained by an increase in their writing post-test scores. Thus, explicit vocabulary instruction using this model has shown to benefit learners' writing skills, especially among the high proficiency learners.

Consequently, this study revealed some fundamental implications. First, there is a pressing need for ESL learners to increase their vocabulary size, and this can be implemented by utilizing the CWF Model (Subon, 2016). Despite the positive increase in vocabulary size after the CWF intervention, the participants were still lacking the vocabulary needed to function well in the second language. Half (21 out of 42 students) of the High proficiency group and a big percentage (86.2% or 50 out of 58 students) of the Intermediate group, as well as 93% (40 out of 43 students) of the Low proficiency groups still had a vocabulary size below the 2000-word level after the ten weeks of treatment. Although the intermediate and low proficiency groups experienced a

statistically significant mean increase in their vocabulary size, only a small percentage from these two groups were able to gain a mastery of the 2000-word level after the treatment. In view of this prevailing weakness in students' vocabulary sizes, there is a need to implement explicit vocabulary instruction that gives special attention to the low and intermediate proficiency learners. This finding resonates with the literature that recommends the explicit teaching of word meanings for struggling and average readers (Coyne et al., 2010; Stahl & Fairbanks, 1986). Past studies also suggest that if students were taught in a way that positively supported their vocabulary growth, they would be able to clarify and enrich their understanding of the meanings of known words (İlter, 2017, 2018). Through this vocabulary intervention program, learners can be purposefully trained to acquire at least the minimum significant level of vocabulary that they should possess i.e., 2,000-word families for typical speech (Adolphs & Schmitt, 2003), 3,000-word families for reading authentic texts (Schmitt, 2007), and 5,000-word families for the purpose of deeper processing (Thornbury, 2002).

Despite its success, the efficacy of this CWF method for enhancing students' guided essay writing requires further investigation. Among the three proficiency groups, only the high proficiency group experienced a statistically significant difference, with an increase in the writing test scores mean after the intervention. There was no significant difference in the test scores after the treatment among the intermediate group, as there was only a slight increase in the mean test scores. On the other hand, the low proficiency group showed a statistically significant difference in their guided essay writing test scores after the intervention, but the results showed a decrease in the mean score. This implies that many of the learners, especially among the intermediate and low proficiency groups, were still unable to write well on the guided essay writing test. Most of them, especially among the low proficiency group, showed a decrease in scores for almost all aspects, including vocabulary, language use, organisation, and content. Hence, their lack of vocabulary was one of the reasons for their inability to obtain good scores for their guided essay writing on the post-test. Despite encouraging growth in the learners' vocabularies after the intervention, the findings also showed that the majority of the students still had a vocabulary size that was below the 2000-word level, i.e., a minimum number of words students should possess to be competent in the target language. According to Nation and Waring (1997), learners need at least 2000 words to be able to obtain 80% text coverage for written discourse. This is consistent with the notion that the lack of vocabulary is often quoted as one of the main reasons for the lack of proficiency in English. This finding provides an important implication that secondary school teachers should help ESL learners develop their L2 productive skills. As suggested by Danilović Jeremić (2015), more emphasis should be paid to the development of productive lexical knowledge at the primary and secondary level of education, which is crucial not only for their future success in English degree courses but also to communicate well in English.

Another important implication is for ESL learners to be given multiple exposures to tier-two or high-frequency words. The CWF Model (Subon, 2016) provides up to seven exposures to the same word and the learning activities involve all four language skills of reading, writing, listening, and speaking. This notion of multiple exposures is supported by Stone and Urguhart (2008) who reiterate that learners should have ample opportunities of at least six encounters with words that are beneficial to learning a new content. Nation (1990) also recommends between five and 16 repetitions to learn a word. Seven or more repetitions are supported by Mckeown, Beck, Omanson, and Pople (1985), who reiterate that even a rich vocabulary program of instruction requires more encounters with a word.Teaching a new word only once is a big mistake because there is a high probability that the word will be forgotten (Schmitt, 2008). Additionally, tier-two words or academic words should be the major focus of vocabulary instruction (Mckeown & Beck, 2011) for writing. It is essential to teach tier-two words (Mckeown & Beck, 2011) or academic vocabulary (Nation, 2001) for the explicit vocabulary instruction of word families because they are used in various types of texts and the meaning they bring to a text (Nation, 2001). Welcome, measure, discover, advertisement, and threatens are some examples of the tier-two words used in this study. The rationale for selecting these words is that they are not only general words but are also high-utility words for literate language users.

This study also has its limitations. Firstly, the treatment of this study was only conducted for ten weeks, which was quite a short period of time for students to practise and improve their writing skills using this method. A longer period of time was not permissible, as students had school activities and holidays. Secondly, giving direct vocabulary instruction of tier-two words could have posed difficulties to the low proficiency group as they might not have acquired many basic vocabulary or tier-one words. The categories of words under tier one are sight words, nouns, verbs, adjectives, and early reading words. There are about 8,000-word families in

English included in tier one. Furthermore, tier-two words are suitable for more advanced learners (Chung, 2012). We take for granted that native English speakers know most of the tier-one words, but this is not the case for L2 learners. Many tier-one words may be unknown to L2 learners and these words are essential for the comprehension of a passage and necessary for writing. Therefore, before giving direct instruction of tier-two or high-frequency words, L2 learners, especially the intermediate and low proficiency groups, need to be given direct instruction of tier-one or basic vocabulary, or both tiers of vocabulary.Giving them better exposure to not only new but the right vocabulary will allow for more impactful effects on their writing. Future research needs to take this factor into consideration when low proficiency students are among the study participants. Other limitations such as teachers' and learners' change in mood and interests during the treatment period were beyond the control of the researcher and beyond the scope of the present study. The instructors' teaching abilities might have led to different impacts on the students' performance due to different expectations, motivation, and attitudes.

Conclusion

This study has attempted to make fundamental contributions to vocabulary acquisition and second language acquisition research. Employing the CWF Model for explicit vocabulary instruction has revealed that teaching and learning vocabulary in word families can have a positive impact on learners' vocabulary size and writing skills. This seven-step explicit vocabulary instruction model, which involves a lot of repetition in learning word families, can help to reinforce learners' vocabulary depth and breadth. Therefore, teachers can adopt this model for enhancing learners' vocabulary acquisition and writing skills in ESL classrooms. Nevertheless, this quasi-experimental study is insufficient for determining the effects of this model fully on learners' vocabulary growth and writing skills due to its limitations. Further empirical research is essential to support and validate the efficacies of this model for enhancing ESL learners' vocabulary size and writing skills. Future studies may ascertain its effectiveness with different participants or participants from different contexts. For instance, studies can also be conducted among elementary school students or pre-university students in higher institutions by employing suitable tier words for their vocabulary learning and acquisition. In the coming months, we will conduct a fully experimental research study to verify and validate the efficacies of the CWF Model for enhancing ESL leaners' vocabulary growth, syntactical structures, and writing skills by making comparisons with other vocabulary learning strategies. This will enable us to employ participants for both control and the experimental groups. By comparing the performances between these two groups, we will be able to generate more compelling and reliable data to ascertain the findings of this future study.

References

Adolps, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics*, 24*(4),* 425-438. https://doi.org/10.1093/applin/24.4.425

Albrechtsen, D., Haastrup, K., & Henriksen, B. (2008). *Vocabulary and writing in a first and second language: Processes and development.* Palgrave Macmillan.

Alderson, J. C. (2005). *Diagnosing foreign language proficiency*. Continuum.

Al-Khasawneh, F. (2019). The impact of vocabulary knowledge on the reading comprehension of Saudi EFL learners. *Journal of Language and Education, 5*(3), 24-34. https://doi.org/10.17323/jle.2019.8822.

Asgari A., & Mustapha, G. (2012). Vocabulary learning strategies of Malaysian ESL students. *Pertanika Journal of Social Science & Humanity,* 20*(3)*, 751-764.

Asgari, A., & Mustapha, G. (2011). The type of vocabulary learning strategies used by ESL students in University Putra Malaysia. *English Language Teaching*, 4*(2)*, 84-90. https://doi.org/10.5539/elt.v4n2p84

August, D., Carlo, M., Dressler, C., & Snow, C. (2005). The critical role of vocabulary development for English Language learners. *Learning Disabilities Research & Practice*, 20*(1),* 50-57. https://doi.org/10.1111/j.1540-5826.2005.00120.x

Bauer, L., & Nation, I . S. P. (1993). Word families. *International Journal of Lexicography*, *6,* 1-27.

Beck, I. L., & McKeown, M. G. (1985). Teaching vocabulary: Making the instruction fit the goal. *Educational Perspectives*, 23*(1),* 11-15.

Beck, I. L., McKeown, M.G., & Kucan, L. (2002). *Bringing words to life; Robust vocabulary instruction.* Guilford Press.

Biemiller, A. (2005). Size and sequence in vocabulary development: Implications for choosing words for primary grade vocabulary instruction. In A. Hiebert & M. Kamil (Ed.)., *Teaching and learning vocabulary: Bringing research to practice* (pp. 223-242). Erlbaum.

Blachowicz, C. L. Z. (2005). *Vocabulary essentials: From research to practice for improved instruction*. Scott Foresman and Co.

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

Chung, S. F. (2012). Research-based vocabulary instruction for English language learners. *The Reading Matrix*, *12*(2), 105-120.

Cohen, A. D. (1988). *Strategies in learning and using a second language.* Longman.

Coyne, M., McCoach, B., Loftus, D., Zipoli, S., Ruby, R., Crevecoerut, Y., & Kapp, S. (2010). Direct and extended vocabulary instruction in kindergarten: Investigating transfer effects. *Journal of Research on Educational Effectiveness*, *3*, 93-120. https://doi.org/10.1080/19345741003592410

Cunningham, A. E., & Stanovich, K. E. (1997) Early reading acquisition and its relation to reading experience and ability ten years later. *Developmental Psychology, 33*, 934-45. https://doi.org/10.1037//0012-1649.33.6.934

Danilović Jeremić, J. (2015). Lexical knowledge of Serbian L1 English L2 learners: Reception vs. production. *Nastava i vaspitanje*, *64*(1), 87-100. https://doi.org/ 10.5937/nasvas1501087D

Duin, A.H., & Graves, M. F. (1987). Intensive vocabulary instruction as a prewriting technique. *Reading Research Quarterly, 22*(3), 311-330. https://doi.org/10.2307/747971

Durkin, D. (1979). What classroom observation reveal about reading comprehension instruction. *Reading Research Quarterly, 14*, 481-538.

Dutro, S., & Moran, C. (2003). Rethinking English language instruction: An architectural approach. In G. Garcia (Ed.), *English learners: Reaching the highest level of English literacy* (pp. 227-258). International Reading Association.

Frayer, D. A., Frederick, W. C., & Kalausmeier, H. J. (1969). *A schema for testing the level of concept mastery.* Technical Report No. 16. University of Winconsin Research and Development Centre for Cognitive Learning.

Gersten, R., & Baker, S. (2000). Effective instruction for English-language learners: What we know about effective instructional practices for English language learners. *Exceptional Children, 66*(4), 454-470. https://doi.org/10.1177/001440290006600402

Ghapanchi, Z., Eskandari, Z., & Tabasi, E. (2012). The effect of text-based direct vocabulary instruction on vocabulary acquisition. *World Journal of English Language*, 2*(1)*, 74-80. https://doi.org/10.5430/wjel.v2n1p74

Graves, M. F. (2000). A vocabulary programme to compliment and bolster a middle grade comprehension programme. In B. M. Taylor, M. F. Graves, & P. van den Broek (Eds.), *Reading for meaning: Fostering comprehension in the middle grades* (pp.116-135). International Reading Association.

Hamp-Lyons, L. (1990). Second language writing: Assessment issues. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 69-87). Cambridge University Press.

Haris, M., & Yunus, M. M. (2018). Vocabulary size of ESL IGCSE students. *International Journal of Advance Research*, *6*(8), 263-267. https://doi.org/10.21474/IJAR01/7518

Hestad, L. C. (2014). *The impact of vocabulary instruction on writing fluency* [Unpublished doctoral dissertation]. Retrieved from ProQuest Dissertations and Theses. (UMI No. 3633490)

Hunt, A., & Beglar, D. (1998). Current Research and Practice in Teaching Vocabulary. *Japan Association for Language Teaching, 22*(1), 7-11.

Hutton, T. L. (2008). *Three tiers of vocabulary and education*. https://www.superduperinc.com/handouts/pdf/182_v

İlter, I. (2017). Teaching word meanings to students at different reading ability: A controlled assessment of the contextual-based vocabulary instruction on reading comprehension. *Education and Science*, *42*(190), 437-463. https://doi.org/10.15390/EB.2017.6482

İlter, İ. (2018). Effects of the instruction in inferring meanings from context on the comprehension of middle school students at frustration reading level. *Journal of Education*, *198*(3), 225-239. https://doi.org/10.1177/0022057418818818

Jacobs, H., Zingraf, S., Wormuth, D., Hartfiel, V.F., & Hughey, J. (1981). *Testing ESL composition: Apractical approach.* Newbury House Publishers.

Kaur, N., & Kabilan, M. K. (2007). *Autonomy in ESL: To what extent?* Paper presented at Literary Conference (LITCON) 2007, Penang, Malaysia.

Kaur, N. Othman, N. H., & Abdullah, M. K. K. (2012). Lexical competence among tertiary students: Teacher-student perspectives. *The English Teacher,* 37, 90-104.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. J. L. Arnaud & H. e joint (Ed.). *Vocabulary and applied linguistics* (pp. 126-132). Macmillan.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*(3), 399-436. https://doi.org/10.1111/j.0023-8333.2004.00260.x

Laufer, B., & Sim, D. D. (1985). Measuring and explaining the reading threshold needed for English for academic purposes texts. *Foreign Language Annals*, *18*(5), 405-411. https://doi.org/10.1111/j.1944-9720.1985.tb00973.x

Mast, D. W. (2011) Using semantic maps and word families in the beginning-level middle school foreign language classroom. *A Journal for K – 16+ Foreign Language Educators, 68*(1), 63-70.

Mckeown, M. G., & Beck, I. L. (2011). Making vocabulary interventions engaging and effective. In R. E. O'Connor, & P. F. Vadasy, (Ed.), *Handbook of reading interventions* (pp. 138-168). New York: Guilford Press.

McKeown, M. G., Beck, I. L., Omanson, R. C., & Pople, M. T. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. *Reading research Quarterly, 20*, 522-535. https://doi.org/10.2307/747940

Moore, D. W. (2010). *Direct Instruction: Targeted strategies for student success*. http://www.insideng.com/profdev/guides/Moore_Instruction.pdf

Nadarajan, S. (2009). The effect of instruction and context on L2 learners' vocabulary development. *Electronic Journal of Foreign Language Teaching, 6*(2), 177-189.

Naeimi, M., & Foo, T. C. V. (2015). Vocabulary Acquisition through Direct and Indirect Learning Strategies. *English Language Teaching*, *8*(10), 142-151. https://doi.org/10.5539/elt.v8n10p142

Nation, I. S. P. (1990). *Teaching and learning vocabulary.* Heinle & Heinle Publishers.

Nation, I. S. P. (2001). *Learning Vocabulary in another Language*. Cambridge University Press.

Nation, P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In Schmitt, N., & McCarthy, M. (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 6-9). Cambridge University Press.

National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific literature on reading and its implications for reading instruction*. National Institute of Child Health and Human Development.

Nemati, A. (2010). Proficiency and size of receptive vocabulary: Comparing EFL and ESL environments. *International Journal of Educational Research and Technology,* 1(1), 46-53.

Paribakht, T. S., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary development. In J. Coady & T. Huckin (Ed.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp.174-200). Cambridge University Press.

Qian, D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, *52*, 513-536. https://doi.org/10.1111/1467-9922.00193

Raasch, K. (2018). *The Effects of Purposeful Vocabulary Instruction*. https://sophia.stkate.edu/maed/274

Scarcella, R. C. (1996). Secondary education in California and second language research: Instructing ESL students in the 1990's. *The CATESOL Journal, 9*, 129-152.

Schmitt, N. (2008). *Teaching vocabulary. White Ppains.* Pearson Education.

Schmitt, N. (2007). Currentperspectivesonvocabulary teaching and learning. In J. Cummins, & C. Davison (Eds.), *International Handbooks of English Language Teaching* (pp. 827-839). Springer.

Schmitt, N. (2000). *Vocabulary in language teaching.* Cambridge University Press.

Schmitt, N., & Zimmerman, C. B. (2002). Derivative word forms: What do learners know? *TESOL Quarterly*, *36(*2), 145-171. https://doi.org/10.2307/3588328

Scott, J. A., & Nagy, W. E. (1997). Understanding the definitions of unfamiliar words. *Reading Research Quarterly, 32*, 184-200.

Sedita, J. (2005). Effective Vocabulary Instruction. *Insights on Learning Disabilities*, *2*(1) 33-45.

Seliger, H. W., & Shohamy, E. (2008). *Second language research methods.* Oxford University Press.

Senechel, M., Oulette, G., & Rodney, D. (2006). The misunderstood giant: On the predictive role of early vocabulary to future reading. In D. K. Dickinson, & S. B. Neuman (Eds.), *Handbook of early literacy research* (pp. 173-182). Guilford

Shostak, J. (2002) The value of direct and systematic vocabulary instruction. *Professional development series, 7*, 1-11.

Silsüpür, B. (2017). Does using language games affect vocabulary learning in EFL Classes? *Journal of Foreign Language Education and Technology, 2*(1), 83-104.

Solati-Dehkordi, S. A., & Salehi, H. (2016). Impact of explicit vocabulary instruction on writing achievement of upper-intermediate EFL learners. *International Education Studies*, *9*(4), 141-154. https://doi.org/10.5539/ies.v9n4p141

Soureshjani, K. H. (2011). The effects of contextualizing and decontextualizing techniques on lexical-oriented knowledge of Persian EFL language learners. *Theory and Practice in Language Studies*, *1*(5), 547-552. https://doi.org/10.4304/tpls.1.5.547-552

Stahl, S. A., & Fairbanks, M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research, 56*(1), 72-110. https://doi.org/10.3102/00346543056001072

Stone, B., & Urquhart, V. (2008). *Remove limits to learning with systematic vocabulary instruction*. https://files.eric.ed.gov/fulltext/ED544614.pdf

Subon, F. (2016). Direct vocabulary instruction: The effects of contextualised word families on learners' vocabulary acquisition. *Procedia - Social and Behavioral Sciences, 224,* 284-291. https://doi.org/10.1016/j.sbspro.2016.05.461

Subon, F. (2018). The contextualized word family model of direct vocabulary instruction. *IOSR Journal of Humanities and Social Science, 23,* 57-70. https://doi.org/10.9790/0837-2303035770

Thornbury, S. (2002). *How to teach vocabulary.* Pearson Education Limited.

Tosuncuoğlu, İ. (2015). Teaching vocabulary through Sentences. *Journal of History Culture and Art Research, 4*(4), 1-11. https://doi.org/10.7596/taksad.v4i4.501.

Tschirner, E. (2004). Breadth of vocabulary and advanced English study: An empirical investigation. *Electronic Journal of Foreign Language Teaching, 1*(1), 27-39.

Waring, R. (2002). *Why should we build up a start-up vocabulary quickly*? http://www1.harenet.ne.jp/~waring/vocab/principles/early.htm

Xie, R.,Zhang, J.,Wu, X., & Nguyen, T. P. (2019). The relationship between morphological awareness and reading comprehension among Chinese children. *Frontiers Psychology*, *10*, 1-13. https//doi.org/10.3389/fpsyg.2019.00054

Yonek, L. M. (2008). *The effects of rich vocabulary instruction on students' expository writing* [Unpublished doctoral dissertation]. University of Pittsburgh. https://d-scholarship.pitt.edu/9995/1/yoneklisa122008.pdf

Zakaria, Z. (2005). *Dictionary as tool in vocabulary acquisition for rural students* [Unpublished master's dissertation], Universiti Sains Malaysia.

Zimmerman, K. (2005). Newly placed versus continuing students: Comparing vocabulary size. *ESL Reporter, 38*(1), 52-60.

National Research University Higher School of Economics
Journal of Language & Education Volume 7, Issue 1, 2021

Tuan, V. V. (2021). The Impact of Social Networking Sites on Study Habits and Interpersonal Relationships among Vietnamese Students. *Journal of Language and Education, 7*(1), 206-218. https://doi.org/10.17323/jle.2021.9818

The Impact of Social Networking Sites on Study Habits and Interpersonal Relationships among Vietnamese Students

Tuan Van Vu
Hanoi Law University

Correspondence concerning this article should be addressed to Tuan Van Vu, Hanoi Law University, 87 Nguyen Chi Thanh street, Dong Da district, Ha Noi city, Vietnam. E-mail: tuanvv@hlu.edu.vn

Social media has a profound influence on every aspect of human beings nowadays. This study investigated the impact of social networking sites on study habits and interpersonal relationships at the tertiary level. A total of 125 college students from different universities in Hanoi were chosen through a convenience sampling technique. Quantitative methodology was employed for the research instrument and a descriptive survey design was adopted for this study. The researchers designed questionnaires with Cronbach's alpha reliability coefficients of at least 0.84 to collect data for the study. Analysis of the data was carried out using frequencies, percentages, means, *t*-tests, and Pearson correlation statistics at the 0.05 alpha level. The findings revealed that students' level of using social networking sites had a negative influence on their study habits and their interpersonal relationships. Based on the findings, it was recommended that regular orientations should be given to students on how and when to use social media to enhance their study habits or to spend time improving their interpersonal relationships with their families, friends, and teachers.

Keywords: interpersonal relationship, social networking sites, learning styles, study habits, ubiquitous

Introduction

Technological innovation has led to the development of many social networking sites (SNSs), which have drawn much attention from netizens. Thomas (2005) writes that "Never before in the history of the planet have so many people - on their own - had the ability to find so much information about so many things and about so many other people" (p. 159). He also emphasizes the impact of the contribution of search engine companies, remarkably Google. Additionally, people are becoming their own editors, researchers, and selectors of entertainment in their homes. Thomas (2005) also states that "personal digital devices can be digitized and therefore shaped, manipulated and transmitted; virtual – these processes can be done at high speed with total ease; mobile – can be done anywhere, anytime by anyone; and personal – can be done by you" (p. 161). Thus, the influence of digital, mobile, personal, and virtual gadgets might impove connectivity, which serves as a basis for the development of the fourth industrial revolution or Industry 4.0 for short, on which Schwab (2016) presented at the World Economic Forum in Geneva, saying that Industry 4.0 strengthens the continuous breakthroughs in fields ranging from gene sequencing to nanotechnology and from renewables to quantum computing. Industry 4.0 cooperates tightly these technologies and their interaction across the physical, digital, and biological domains.

Thanks to SNSs, virtual interpersonal relationships are strengthened to develop a long-term association between them. Nowadays, students use many types of SNSs for different purposes such as updating information, confessions, sharing resources, and entertainment. There are questions about whether SNSs influence the study habits and interpersonal relationships of students in Vietnam in general and those in Hanoi Law University, Vietnam in particular. Although SNSs provide many benefits such as ease and comfort for students, it has been observed that students tend to focus on non-educational, unethical, and inappropriate activities such as useless chatting and using SNSs on their mobile phones, laptops, and other devices. Currently, students

seem addicted to using their mobile gadgets. They even use mobile devices to get online while the teaching is holding a class. Outside the classroom, students waste a great deal of time on social networking activities; they neglect their studies, which results in poor academic performance (Badri et al., 2017; Langat, 2015; Gok, 2016; Oginni et al., 2016; Perpetua & Kelechi, 2018). Until now, the relationship between the frequent use of SNSs and the effect on interpersonal exchanges has been investigated in many studies to demonstrate whether SNSs help students improve their involvement with other people. The findings in the studies such as Arjun and Juna (2015) and Shakiratul (2013) have shown that students receive negative results from an over-dependence on SNSs. On the other hand, these provisional virtual lives provide chances for those who do not have the confidence to speak face-to-face with other people. It can be more convenient and they feel freer to interact in their virtual societies. Some researches, namely Amin et al. (2016), Bernard and Dzandza (2018), Thomas et al. (2016), and Yamakanith and Gurusamy (2014) support the fact that SNSs contribute to the enhancement of communication. When students log on to SNSs, they feel like they are in heaven and they do not realize that addiction to SNSs can harm their inner self-confidence forever.

Literature Review

Definitions of Social Networking Sites

There are many definitions relating to SNSs. According to the APA Dictionary of Psychology, SNSs are defined as online social networks that consists of a set of interconnected individuals who communicate with each other via computer-based technologies rather than interacting face-to-face. In addition, CALD (2013)[1] defines SNSs as websites designed to help people communicate and share information, photographs, etc. with a group. OALD (2015)[2] defines an SNS as an application through which users can communicate with each other by adding information, messages, images, etc. The Collins Dictionary[3] calls an SNS "a website that allows subscribers to interact, typically by requesting that others add them to their visible list of contacts, by forming or joining sub-groups based around shared interests, or publishing content so that a specified group of subscribers can access it." Together with the development and innovations in internet-based connections, SNSs have recently mushroomed. Although some popular SNSs have appeared and disappeared since the internet became popular in the 1990s, SNSs such as Twitter, Facebook, Weibo, and YouTube have drawn a great deal of attention from both businesses and subscribers. Actually, there isn't much difference between the terms 'social network sites' and 'social networking sites'. Therefore, these two phrases are used interchangeably without much misunderstanding. Boyd and Ellison (2007) differentiate between them by arguing that 'networking' refers to the commencement of the relationship the strangers while 'network' covers a broader meaning including not only meeting strangers but also articulating and making their social networks visible. While Kaplan and Haenlein (2010) propose an equivalent term called 'social media', which is "a group of internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content" (p. 61). However, they focus only on the technical features that allow subscribers to exploit the facilities of the websites.

Influence of Social Networking Sites on Students' School Performance

It is necessary for students to engage themselves in school performance to keep updated about their schooling activities. Maslowski (2001) claims that school performance involves "the effectiveness and efficiency of the schooling process" (p. 13). The difference between effectiveness and efficiency is that the accomplishment of the school's objectives reflects effectiveness while efficiency refers to the conditions through which the objectives were completed in a timely and cost-effective manner. In a broader view, the schools' objectives are to supply learners with effective and efficient learning approaches and environments.

[1] CALD (2013). Social networking site: 1 the use of websites and other internet services to communicate with other people and make friends. *Cambridge Advanced Learner's Dictionary - 4th Edition*. https://dictionary.cambridge.org/dictionary/english/social-networking-site

[2] OALD (2015). Social network noun 1 a social networking site or application through which users can communicate with each other by adding information, messages, images, etc. *Oxford Advanced Learner's Dictionary 9th edition © Oxford University Press, 2015*

[3] COBUILD 9th Ed. (2018). *Collins COBUILD Advanced Learner's Dictionary: The Source of Authentic English*. Glasgow: HarperCollins UK; Ninth Edition, Ninth edition.

When examining whether SNSs help or harmfully the students' school performance, Alwagait et al. (2015) carried out research on 108 students in Saudi Arabia and concluded that the usage of SNSs by students did not have a negative effect on their GPA scores. This matches the findings of Sutherland et al. (2018), who investigated whether SNSs facilitated engagement in offline environments with peers within university communities. The study claimed that social media profiles improve students' communication within their academic community. Similarly, an empirical study conducted by Erick (2015) concluded that SNSs have a positive effect on students' academic performance. Badri et al. (2017) also suggested that students should choose suitable SNSs because their SNSs help them relieve the stress in their lives. In addition, Sandra and Ismail (2016) examined the impact of the use of social media on students' academic performance in Malaysian tertiary institutions. Their findings revealed that, under the supervision and thorough guidance of the universities and colleges, their students exploited the resources from SNSs, which eventually resulted in a positive impact on their academic performance. Thus, SNSs can increase convenience and be useful resources for students.

Doubts might arise whether SNSs have a positive effect on students' performance. Rabbani et al. (2015) investigated the effect of SNSs on student performance. Their findings revealed that students use SNSs for connecting with other people or for enjoyment; they do not use SNSs for academic purposes. In another research study on the relationship between the use of SNSs and communication skills, Shakiratul (2013) claimed that students suffer from poor communication skills because they are highly addicted to SNSs and disengaging from reality. He also recommended that students cut down on the time spent on SNSs and spend more time communicating with real people in the real world, face-to-face, not just via virtual settings. Bernard and Dzandza (2018) investigated the level of engagement of students at the University of Ghana on SNSs and determined the effect of their use on their academic work. They claimed that students' academic lives were negatively influenced due to it distracting their attention. SNSs have a negative effect on student achievement according to a study by Elizabeth (2012). Similarly, Wong et al. (2017) found that being addicted to SNSs led students to acquire less new knowledge and be lazier regarding outdoor activities.

Influence of Social Networking Sites on Study Habits

This section discusses the influence of SNSs on study habits with the positive aspects first. Thanks to the availability and convenience of the internet, SNSs arm students with a powerful tool to acquire knowledge from a variety of internet resources via search engines such as Google. SNSs have become an educational tool thanks to students' ability to use search engines. SNSs can also serve as a 'meetup' for students to exchange ideas with their friends. The more students have good study habits and various social skills, the more valuable and useful the social networking sites can be (Arjun & Juna, 2015). The positive influence of SNSs on study habits is also reflected in the findings by Salman, Mohannad, and Joannes (2016) on social media utilization and the study habits of nursing students at Ha'il University. They claimed that the more the nursing students exploited SNSs, the better they improved their study habits. Thus, the utilization of SNSs plays an important role in predicting students' study habits. Furthermore, Thomas, et al. (2016) also presented their findings with the conclusion that SNSs may improve the study habits of nursing students in the College of Nursing.

On the other hand, many other studies have proved that SNSs have negatively influenced students' study habits. The frequent use of SNSs by students at Alvan Ikoku Federal College of Education distracted them from the lecture notes and students either failed to submit or carry out their assignments (Perpetua & Kelechi, 2018). These findings are somewhat similar to those of other previous studies (Ariel & Raymund, 2015; Langat, 2015) whose results revealed that SNSs could negatively influence students' study behavior and academic progress. Tolga (2016) conducted research at a vocational school of higher education with 220 students. The author pointed out that SNSs can have a negative impact on students' studying and habits. Simultaneously, Oginni et al. (2016) examined the influence of students' engagement with social networks on their study habits and cognitive skills in science classrooms. Their findings indicated that SNSs can have a bad effect on study habits because of continuous students' engagement via online social networks. In short, it has not been widely concluded whether SNSs are good or bad for students' study habits.

Influence of Social Networking Sites on Interpersonal Relationships

Industry 4.0 speeds up more activities happening on the internet. New terms appear overnight, such as tele-working, tele-sales, and e-learning. In a thesis by Mwangi (2013) on the impact of SNSs among

teenagers, the results revealed that teenagers develop positive relationships thanks to SNSs; they think that the world would be boring without them. Yamakanith and Gurusamy (2014) also share the same views when pointing out that SNSs have improved communication among college students and their family members, friends, and teachers. Although students use different SNS applications, their overall social connectedness has greatly improved (Jesse, 2016). Aishwarya and Vinod (2017) also claimed that using SNSs improves communication for users who prefer participating in SNSs more than face-to-face communication.

On the contrary, many researches blame SNSs for hindering the closeness of interpersonal relationships. Drussell (2012) investigated the influence of SNSs on interpersonal communication and conflict resolution skills among college freshmen and the findings pointed to the importance of face-to-face interactions between individuals for resolving conflicts. Similarly, Hamza (2013) accused Facebook of reducing the desire for interpersonal communication among students, which resembled the findings of Pritta's research published in 2014. The findings strongly emphasized that face-to-face interactions are required for serious and personal conversations to avoid the misundestanding that people can have when communicating via SNSs. The findings also confirmed that while the influence of social media is acknowledged, it cannot completely replace the interactions of human beings. Due to the fast pace of life, people now depend heavily on SNSs; they seem to avoid face-to-face communication by switching over to mobile phones (Subramanian, 2017). The use of SNSs is highly influential as using SNSs might ease or create family tension if family members who live far away use SNSs for connectedness to avoid becoming sad or lonely, but they cannot usually communicate with each other over SNSs (Anika et al., 2015; Saida, 2017). Spencer (2018) shared the same ideas by accusing SNSs of decreasing the quality of family relationships. Alona, Arnon, and Rebecca (2015) concluded that there was a positive association in the teacher-student relationships on SNSs, especially Facebook. Arnon and Alona (2017) also carried out research on the significant associations between student-teacher relationships and student-teacher Facebook-mediated communication. Similar to their previous study in 2013, they claimed that students feel it is comfortable and convenient to befriend teachers via Facebook. Another research study conducted by Arnon (2018) revealed that one-to-one computing programs positively affect student-teacher relationships. In conclusion, the effects of SNS use on students' interpersonal relationships is still controversial.

This study investigates the hypotheses that:

1. There is no influence from SNSs on students' study habits and the interpersonal relationships with their families, friends, and teachers.
2. There is no difference in the use of SNSs between male and female students.

Materials and Methods

Participants

Through a convenience sampling technique, 125 respondents were chosen, namely 40 freshmen, 50 sophomores, 22 juniors, and 13 seniors. Students walking past the gates to the campuses in Hanoi were randomly invited to be the research participants. Over 500 students were invited to answer the questionnaires, but approximately 135 passers-by declined to participate. After two days of collecting data, the screening process was implemented to ascertain that all participants had taken it seriously. The 365 answered survey questionnaires were screened and 240 answer sheets were rejected as the passers-by had not correctly filled them out in spite of the thorough guided instructions. The final sample consisted of 61 male students (48.8%) and 64 female students (51.2%). Of these respondents, 25.6% were aged 18, 29.6% were 19 years old, 26.4% were 20, and 18.4% were 21 years old or older. The participants were almost all single (118 single, 7 married). They resided in rural areas (50.4%), mountainous areas (29.6%), and urban areas (20%).

Measures

The survey questionnaire consisted of demographic data, the use of devices, the places where they engaged with SNSs, the frequency of visits, and the length of engagement times.

The following limits of description were also used.

	Limits of Description	Qualitative Interpretation
5	4.2 – 5.0	very high
4	3.4 – 4.19	high
3	2.6 – 3.39	moderate
2	1.8 – 2.59	low
1	1.0 – 1.79	very low

Procedure

The researcher-made survey questionnaire was presented to five experts on educational assessment and accreditation for content validation. The questionnaire was then fine-tuned with a group of 70 students who participated in a pilot study to validate the strengths and weaknesses of the instrument. Cronbach's alpha for the respondent's reasons for use of SNSs was .952 for 20 items. A 20-item scale questioning the participants on social networking and their study habits yielded a Cronbach's alpha of .910. Similarly, three 10-item scales investigating interpersonal relationships with the respondents' family, friends, and teachers returned Cronbach's alphas of .837, .834, and .842, respectively. Finally, the five experts were again consulted for the finalization of the questionnaire before the researcher personally administered it.

The author assigned a score to each of the 125 participants based on their self-reports of SNS use. Those who reported "never" were assigned a score of 1. Those who reported at least once per month earned a 2, at least once per week earned a score of 3, at least once each day earned a 4, and those who reported more than five times per day were assigned a score of 5. There were 20 items for "Study Habits." The author also reversed scored items 1, 2, 3, and 18 (5 became 1 & vice versa, 4 becomes 2 & vice versa, & 3 remained 3). Then he added the scores for each of the 20 items under "Study Habits" such that a score of 5 indicated total agreement and a score of 1 indicated total disagreement. Thus, each participant could earn a score between 20 and 100. He then ran correlation coefficients between SNS use and total Study Habit score.

There were 10 items each in the Families, Friends, and Teachers categories. All of these items were scored in a positive direction. Thus, total scores on Families were correlated with self-reports of SNS use. The same procedure was used for the Friends and Teachers categories. Furthermore, a *t*-test was used to test the significant difference of respondent's interpersonal relationships in terms of the profile variable gender.

The questionnaires were treated with utmost confidentiality. All collected data were organized, summarized, analyzed, and interpreted. IBM SPSS software version 25 was used for the treatment of the collected data.

Results and Discussion

In terms of the communication tools the participants used to get connected to SNSs, mobile phones were used popularly with 52 users, which made up 41.6%, followed by 43 laptop users (34.4%), the lowest number of respondents used their desktops for participating in SNSs – 30 students (24%). According to Vietnam Internet statistics 2019[4], the number of the internet users who accessed via mobile devices was 61.73 million, which accounted for 96% of the 64 million internet users in Vietnam. Regarding the venues where participants joined SNSs, school was the most popular with 64 respondents logging on to SNSs (51.2%). The second place was their homes for 41 students (32.8%). The bottom rank belonged to the internet cafes with 20 participants, accounting for 16%. Wifi hotpots are ubiquitous in public places in Vietnam, which was ranked 75 over 200 countries in the worldwide broadband speed league 2019 (Cable organization, 2019)[5]. However, students do not frequently connect to SNSs at internet cafes.

The Table 1 presents the frequency of usage on SNSs by the participants. Zalo, YouTube, and Facebook were the most frequently used, with mean scores of 4.03, 4.02, and 3.49, respectively. Zalo is one of the most common

[4] Vietnam Internet statistics 2019. https://vnetwork.vn/en/news/cac-so-lieu-thong-ke-internet-viet-nam-2019
[5] Worldwide broadband speed league 2019. https://www.cable.co.uk/broadband/speed/worldwide-speed-league/

social networking applications in Vietnam. Respondents also used Instagram for sharing their pictures to a moderate degree. WhatsApp, Viber, Twitter, and LinkedIn were on the lower end according to the participants' preferences. Lastly, Google Plus and other similar sites did not attract the attention of the respondents, they were at very low levels. Overall, the respondents mainly used the SNSs for the purpose of communication and entertainment.

Table 1
The frequency of usage on social networking sites

	N	Minimum	Maximum	Mean	Std. Deviation
Facebook	125	1	5	3.49	1.175
Twitter	125	1	5	2.15	1.129
Zalo	125	1	5	4.03	1.177
LinkedIn	125	1	5	2.10	1.054
Instagram	125	1	5	2.83	1.120
Viber	125	1	4	2.35	1.072
WhatsApp	125	1	4	2.47	.921
YouTube	125	3	5	4.02	.615
Google Plus	125	1	4	1.59	.834
Other sites	125	1	2	1.42	.495
Valid N (listwise)	125				

When asked about the frequency that the participants visited a social networking site, the results showed that 64 respondents visited at least one time each day, which accounted for 51.2%, the highest rate. It's quite astonishing to know that 34 participants (27.2%) logged on a social networking site more than five times each day. The third rank was 15 participants (12%) who surfed it at least once each week. The smallest figure was 12 respondents (9.6%); they confessed that they only joined a social networking site at least once each month. Generally, students spent a lot of time on SNSs and did not have any difficulties connecting to the internet.

Regarding the number of hours the respondents spent on SNSs per day, 49 students (39.2%) stated that they spent five hours or more. For the length of 3-4 hours and 4-5 hours, the respondents shared the same numbers 30 students each, meaning 24% each. Only 16 participants or 12.8% answered that they spent one or two hours per day visiting SNSs. We can infer from these numbers that students have a chance to connect to the internet at a high level. They do not have any difficulties connecting to SNSs.

Considering the respondents' reasons for using SNSs, the results showed that the overall weighed means were high (3.54). When eliciting the participants' viewpoints, students confessed that sharing and posting content (3.22), and chatting and talking online (3.03) were at moderate levels. Other items were at high levels, namely reading content (3.45), viewing photos and videos (3.52), commenting and replying to others (3.42), expressing one's thoughts and ideas (3.52), uploading photos and videos (3.66), making new contacts (4.00), playing online games (3.45), advocating for a cause (3.52), watching virtual people (3.45), making appointments (3.95), find information for their studies (3.45), getting information about contacts (3.52), teaching cooking recipes, decorating, etc. (3.66), sharing interests (3.58), keeping in touch (3.90), editing pictures with designs and layouts (3.45), gathering advice and opinions from others (3.52), and for relaxation (3.46). The evolution of the internet is so fast that people find it hard to imagine the internet disappearing. Connecting the world has supplied people with new opportunities, new challenges, and new partners, but also new dangers (Thomas, 2005). Similarly, participants in this study were eager to make new contacts using SNSs. They also found that making appointments though SNSs was useful, convenient, and easy for keeping in touch with other people. However, the respondents were reluctant to share or post content on their SNSs. From this, we can infer that they are afraid of harassment or bullying on SNSs. The results also made it known that the respondents did not want to waste time chatting and talking online.

Table 2 presents the problems that the respondents experienced when using SNSs. Similarly, the figures pointed out that respondents encountered a high overall weighted mean. SNSs allow the respondents to post

photos, show their offline behaviors, or send threatening messages to others. Practically, personal privacy is violated when the law on cybersecurity is not strictly complied with. For example, the respondents claimed that they did not want their private instant messages, text messages, or e-mails to be disclosed illegally. It can be inferred that the participants do not want to make some of their personal matters public, they want to keep them reserved for the sake of avoiding problems with other people. Other problems such as harmful conversations, emotional distress, harassment, cyberbullying or sexual advances have been shown to be common. Actually, although many SNSs have listed some rules or warnings on their sites, no strict policies are in place to make sure that the participants comply with all these regulations, which are the sources of the problems. Spending a lot of time on SNSs can also cause some health problems. Stomachaches, sleeping problems, or headaches are the symptoms of less movement and habit changes. It is clear that when normal habits are disrupted, this can harm people's health. The findings in Table 2 also show that the participants suffered from health problems. Emotional affection was high. Real-world conflicts, danger regarding pedophiles, laziness for finishing homework assignments on time, and narcissistic tendencies or too much love for self were confessed by the participants. They spent too much time getting online, so the virtual world separated them from their real lives and the participants were negatively influenced by spending too much time on cyberspace. It is easy to understand that the relationships with their family and friends or school performance were impacted. In order to increase the number of interactions and attract more participants, most SNSs are linked to cyber games so that users do not feel bored and the virtual world which poisons the communities of SNSs is more attractive.

Table 2
The related social networking problems of respondents

		Weighted Mean	Descriptive Interpretation
1	I am permitted to be exposed to harassment, bullying, or sexual advances because of social networking.	3.45	high
2	I am permitted to publicize private instant messages, text messages or e-mails because of social networking.	2.50	low
3	I am permitted to post threatening messages because of social networking.	3.45	high
4	I am permitted to post photos that will cause embarrassment and spreading of rumors because of social networking.	3.52	high
5	I experience emotional distress because of social networking.	3.45	high
6	I conflict with others because of social networking.	3.52	high
7	I am in greater danger regarding pedophiles because of social networking.	3.66	high
8	I experience harmful conversations because of social networking.	4.00	high
9	I share too much information about name, date of birth, and location causing others to copy my identity.	3.45	high
10	I am permitted to express online expressions of offline behaviors such as bullying, clique-forming, and even sexual experimentation because of social networking.	3.52	high
11	I am prone to internet addiction because of social networking.	3.45	high
12	I find difficulty in sleeping and forgetting to eat because of social networking.	3.95	high
13	I have gap with my parents in terms of knowledge and technical skills because of social networking.	3.45	high
14	My focus on my studies is being blocked because of social networking.	3.52	high
15	I get lazy about doing work because of the over-usage of social networking.	3.66	high
16	I am distracted during instruction time and experience a negative impact on my learning environment because of social networking.	3.58	high
17	I experience stomachaches, sleeping problems, and headaches because of regular social networking use.	3.94	high
18	I show narcissistic tendencies or too much love for self because of too much posting of my photos.	3.45	high
19	I experience harassment or cyber bullying because of social networking.	3.52	high
20	I am addicted to cyber games because of social networking.	3.46	high
	Overall weighted mean	**3.52**	**high**

Table 3 presents the effect of SNSs on the study habits. From the findings, it can be seen that using SNSs impacted the understanding of the textbooks and professors. Students paid too much attention to the websites and apps, or they might think about how to make their virtual world better, such as their characters, their farm house, or their virtual families. When reading books or listening to their professors, they were obsessed by their virtual sites, which led to the poor school achievement (Bernard & Dzandza, 2018; Elizabeth, 2012). Using SNSs to discuss lessons with peers was not preferred by the respondents. This fact seems quite surprising because SNSs are so popular and easy to connect to. Students use SNSs for connecting with other people and for enjoyment, not for the sake of academic achievement (Rabbani et al., 2015). SNSs did not contribute positively to the acquisition of English skills, written expressions, or creativity according to the remarks of the students.

English is acquired as a foreign language (EFL) in Vietnam, so students can benefit from using SNSs for their purpose of learning English through making friends with people from around the world, or registering in various forums on the internet. SNSs, however, only had a minor influence on Vietnamese students in terms of acquiring English. It is interesting to question students about whether they prioritized academic-related activities. The finding revealed that no preference for academic purposes was shown in this case. Similarly, their reading and writing skills did not receive any improvement when using SNSs. In general, students do not use SNSs for the purpose of improving their study habits (Oginni et al., 2016). Using SNSs to arrange group discussions with their classmates and finding time to study were rated at moderate levels. The respondents seemed indifferent to their studies; they were very busy getting online or taking care of their personal matters, which put them in danger as this deterred them from improving in their studies (Alwagait et al., 2015; Sutherland et al., 2018). Due to not focusing on their studies, they often failed to hand in their assignments because they spent much time on SNSs. There should be strict policies in schools for those who do not obey the teachers' instructions. Students' academic weaknesses due to SNSs were also found in other studies (Langat, 2015; Perpetua & Kelechi, 2018). Thus, their study habits were negatively influenced by using SNSs to a certain extent.

Table 3
The extent of the effect of SNSs on study habits

		Weighted Mean	Descriptive Interpretation
1	I am late handing in assignments because I spent time on a social networking site instead of doing homework	3.78	high
2	I spend much of my time on a social networking site, thus it interferes with my ability to study or complete assignments.	3.50	high
3	I still keep up with the required readings in school even when doing social networking.	3.52	high
4	I use social networking for collaborative learning.	3.66	high
5	I still do well on my exams even when doing social networking.	3.55	high
6	I am provided with academic activities and establish an interaction with peers and teachers in the school through social networking	3.90	high
7	I am able to develop research skills because of social networking.	3.45	high
8	I can arrange group discussions with my classmates through the use of social networking and this improves my academic performance.	3.34	moderate
9	I still find time to study even when doing social networking.	3.14	moderate
10	I learn with social networking, it helps me participate in class discussions.	4.07	high
11	I am able to improve my interactions with classmates and lecturers via social networking.	3.45	high
12	I still get good grades even when I do social networking.	3.52	high
13	I am able to improve my communication skills because of social networking.	3.66	high
14	I feel that my opinions have been taken into account in the class because of social networking.	3.41	high
15	I understand my textbooks and my professors because of social networking.	2.20	low
16	I have the opportunity to discuss lessons with peers because of social networking.	2.02	low
17	I am able to reinforce my skills in English, written expression, and creativity because of social networking.	2.14	low
18	I tend to focus on networking and that blocks my focus on my studies.	3.52	high

		Weighted Mean	Descriptive Interpretation
19	I prioritize academic-related activities over social networking.	2.02	low
20	I am able to improve my reading and writing skills because of social networking.	2.11	low
	Overall weighted mean	**3.19**	**moderate**

When asked about the interpersonal relationships of respondents with their family, the participants had more negative responses to the questions about desiring to interact more with family (2.02), prioritizing having quality time with family (2.11), keeping in touch with family (2.02), better communication with family (2.11), doing the chores and helping their families (2.02), and enjoying having time with their families (2.11). These findings revealed that students tended to be less open with their family. They did not feel involved in their daily lives. Sadly, they wanted to live within their own world by participating in SNSs. The greater use of SNSs is highly influential to the breakdown of family relationships (Anika et al., 2015; Saida, 2017; Spencer, 2018). Surprisingly, the participants were more positive about some of the questions, particularly expressing themselves more with their families (3.45), balancing their time with their family (3.52), relating more and better with their family (4.07), and obeying the rules of their parents (3.45). After taking a serious look at these findings, some concerns surfaced because of the vague lines between these items. Some items were designed with nearly the same contents but they received different responses. The respondents were not really sure about their answers. Nevertheless, the overall weighted mean was 2.69, meaning to a moderate extent.

In terms of the students' connectedness with their friends through SNSs, the respondents showed low levels for the main ideas of questions such as interacting more with their friends (2.20), relating more with friends (2.02), being on good terms with friends (2.14), communicating better with friends (2.02), and enjoying spending time with friends (2.20). SNSs hindered the closeness of relationships with friends. Students became reserved and they did not become the focus of the attention with their friends (Drussell, 2012; Hamza, 2013; Subramanian, 2017; Wong et al., 2017). For other questions, they reported higher levels of interest, namely keeping in touch with friends (3.52), spending more time with friends, like hanging out with them (3.45), expressing themselves more with friends (3.52), being open to sharing their experiences with friends (3.66), and being ready to help friends when needed (3.58). These items were supported by the findings in other studies (Aishwarya & Vinod, 2017; Jesse, 2016; Mwangi, 2013; Yamakanith & Gurusamy, 2014). In general, the overall weighted mean was at a moderate level (2.83). In the same situation as relationships with their family, the participants were not rigid in their answers, some of the nearly same meanings produced different results.

With regards to the interpersonal relationships of respondents with their teachers, the findings revealed that the respondents reported lower weighted means for interacting more with teachers (2.12), focusing on the lesson being discussed by teachers (2.02), communicating with teachers easily (2.12), being helped much better in terms of asking teachers regarding school work (2.00), easily talking to or approaching teachers (2.12), and feeling valuable to teachers (2.00). From these findings, it can be inferred that students felt unwilling to communicate with their teachers, even when they had problems regarding school work. The reasons did not come from the difficulties talking to or approaching teachers. They might feel shy to communicate because they thought that they were not important to their teachers. The more students communicated with their teachers, the more successful they would be (Alona et al., 2015; Arnon & Alona, 2017). They also showed high levels of interest in expressing themselves more to teachers (3.42), easily opening up about their concerns/problems with teachers (3.46), prioritizing tasks given by teachers (4.06), and interacting with teachers (3.42). These findings highlighted the fact that face-to-face communication might not be of interest to students. Thanks to SNSs, students felt more comfortable discussing personal matters with their teachers. SNSs were useful for some students as an available means of communication with their teachers. In addition, students could benefit from using SNSs for doing their tasks given by their teachers. They could search the internet or consult forums or sites (Arnon, 2018; Arnon & Alona, 2017). On the whole, the overall weighted mean for this part was also at a moderate level (2.67).

Table 4 presents the chi-square test on the significant relationship between the respondents' use of SNSs and their connectedness with their study habits, families, friends, and teachers. The findings revealed that there was a significant relationship, as the four values were under 0.05. That means SNSs affect students' study habits as well as their relationships with their families, friends, and teachers.

Table 4
Relationship between the respondents' use of SNSs and their surroundings

	X^{2-} value	df	Asymptotic Significance (2-sided)	Decision at .05 level
Their study habits	2150.089	1716	.000	significant
Their families	1388.239	924	.000	significant
Their friends	1362.854	1012	.000	significant
Their teachers	1091.881	968	.003	significant

Table 5 presents the *t*-test analysis on the significant difference of the respondents' reasons for use of SNSs in relation to gender. As shown in Table 5, the Levene's test had *sig*. 0.150, which is higher than the preset value 0.05. Therefore, the results in the row "Equal variances assumed" are taken into account. Thus, the finding shows that there was no difference in the reason for use of SNSs when grouped according to gender. Moreover, *sig*. (2-tailed) 0.755 confirmed that this figure was higher than the 95% confidence level so that the conclusion was the same with the Levene's test. The mean in the group statistics also discloses the fact that the difference between the group was not high. The mean in the male row was 3.516 compared with that of 3.5570 for the female students. Therefore, the reasons for using SNSs between male and female students were not different.

Table 5
Difference between respondents' reasons for use of SNSs when grouped according to gender

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Reason	Male	61	3.5156	.80278	.10279
	Female	64	3.5570	.67799	.08475

Reasons for use of SNSs and gender

		Levene's Test for Equality of Variances		t-test for Equality of Means						
								95% Confidence Interval of the Difference		
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Reasons for use of SNSs	Equal variances assumed	2.093	.150	-.312	123	.755	-.04146	.13268	-.30409	.22118
	Equal variances not assumed			-.311	117.563	.756	-.04146	.13322	-.30528	.22236

Table 6 deals with the relationship between social networking influence with gender. As seen from the group statistics, the mean presented little difference in the figures 3.5090 for male students and 3.5398 for female students. To put it simply, the *sig*. 0.197 in the Levene's test for equality of variances and sig. 0.824 (2-tailed) in t-test for equality of means were higher than the preset value 0.05. Therefore, the conclusion was that social networking influence affected the genders somewhat equally.

Table 6
The relationship between social networking influence with gender

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Social Networking Influence	Male	61	3.5090	.83469	.10687
	Female	64	3.5398	.71232	.08904

Reasons for use of SNSs and gender

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference		
								Std. Error Difference	Lower	Upper
Social Networking influence	Equal variances assumed	1.685	.197	-.222	123	.824	-.03083	.13857	-.30513	.24347
	Equal variances not assumed			-.222	118.040	.825	-.03083	.13910	-.30629	.24463

Conclusion

With the ubiquity of communication devices, together with the fast development and innovation in terms of information and communications technology, SNSs have a far-reaching influence on all aspects of life. Regarding this study, which only focuses on the impact of SNSs on students' study habits and the interpersonal relationships with their families, friends, and teachers, the outcomes prove that that students' communication skills could be negatively affected by their addiction to SNSs. By being highly addicted, students have a tendency to disengage themselves from reality, which results from wasting all their time keeping updated on the internet. As a result, the overuse of the internet leads to their poor communication skills with their surroundings. Their socialization with their teachers, friends, and families is, to some extent, badly affected due to the lack of the exchange of information in real life. Obviously, frequent face-to-face interactions help students build a better rapport with other people in the real world. Apart from the poor interchange with other people, as discussed above, this study has shown that negative academic performance is a result of students being distracted from their studies because they spend too much time on SNSs. Thus, the overuse of SNSs could have serious consequences on the academic lives of students, which might lead to harmful career prospects for students when they cannot meet the demands of the real world after graduation.

Acknowledgments

The author would like to express his deepest gratitude and sincere appreciation to the blind reviewers and the editors who encouraged, assisted, and gave him constructive feedback to shape up this paper.

Funding information

This article is financially supported by Hanoi Law University.

Conflict of Interests

The author declares no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Alona, F., Arnon, H., & Rebecca, P. A. (2015). Teacher-student relationship and SNS-mediated communication: Perceptions of both role-players. *Interdisciplinary Journal of e-skills and Life Long Learning*, *11*, 273-289.

Alwagait, E., Shahzad, B., & Alim, S. (2015). Impact of social media usage on students' academic performance in Saudi Arabia. *Computers in Human Behavior, 51*, 1092-1097. https://doi.org/10.1016/ j.chb.2014.09.028

Arjun, M., & Juna, B. (2015) A study on the impact of social networking sites among higher secondary students. *International Journal of Social Sciences and Management, 2*(4), 339-342. https://doi.org/10.3126/ ijssm.v2i4.13681

Amin, Z., Mansoor, A., Hussain, S. R., & Hashmat, F. (2016). Impact of social media of student's academic performance. *International Journal of Business and Management Invention, 5*(4)*,* 22-29. https://www.ijbmi.org/papers/Vol(5)4/version-2/D050402022029.pdf

Ariel, R. J., Raymund, B. G. (2015). Influence of social networking on the study habits and performance of students in a state university. *Journal of Social Sciences & Humanities Research, 1*(2), 1-12.

Aishwarya, M. S., & Vinod, L. (2017). Impact of social media on interpersonal communication. *Imperial Journal of Interdisciplinary Research, 3*(7), 137-140.

Anika, A., Ishrat, A., Tanzima, H., & Jalal, M. (2015). Impact of imbalance usage of social networking sites on families. *Human-computer Interaction, 1*(1), 1-12. https://arxiv.org/abs/1510.07382

Arnon, H. (2018). The student-teacher relationship in the one-to-one computing classroom. *Páginas de Educación*, *11*(1), 37-65. https://doi.org/10.22235/ pe.v11i1.1553

Arnon, H., & Alona, F. (2017). Teacher-student relationship and Facebook-mediated communication: Student perceptions. *Critical citizenship and social empowerment in the emerging cybersociety, 53,* 91-100.

Badri, M., Nuaimi, A. A., Guang, Y., & Rashedi, A. A. (2017). School performance, social networking effects, and learning of school children: Evidence of reciprocal relationships in Abu Dhabi. *Telematics and Informatics*, *34*(8)*,* 1433-1444. https://doi.org/10.1016/j.tele.2017.06.006

Bernard, K. J., & Dzandza, P. E. (2018). Effect of social media on academic performance of students in Ghanaian universities: A case study of University of Ghana, Legon. *Library Philosophy and Practice*. https://digitalcommons.unl.edu/libphilprac/1637

Boyd, D. M., & Ellison, N. B. (2008). Social network sites: Definition, history, and scholarship. *Journal of Computer-mediated Communication*, *13*, 210-230. https://doi.org/10.1111/j.1083-6101.2007.00393.x

Dörnyei, Z., & Taguchi, T. (2010). *Questionnaires in second language research: Construction, administration, and processing* (2nd ed.). Taylor & Francis.

Drussell, J. (2012). *Social networking and interpersonal communication and conflict resolution skills among college freshmen*. https://sophia.stkate.edu/msw_papers/21

Erick, M. M. (2015). *An investigation on the impact of online social networking on academic performance among high school students in urban areas: A case study of Westlands sub-country, Nairobi*. https://journalism.uonbi.ac.ke/sites/default/files/chss/journalism/journalism/FINAL%20PROJECT.pdf

Elizabeth, P. (2012). *Online social networking sites and students achievement* [Unpublished master thesis]. Memorial University of Newfoundland.

Jesse, J. L. (2016). The impact of social networking sites (SNSS) on students' social interaction. *European Journal of Education Studies, 2*(7), 135-147. http://doi.org/10.5281/zenodo.163490

Gok, T. (2016). The effects of social networking sites on students' studying and habits. *International Journal of Research in Education and Science*, *2*, 85-93. https://www.ijres.net/index.php/ijres/article/view/89/53

Hamza, K. A. (2013). Impact of social networks on interpersonal communication of the students University College Irbid Girls: Facebook as a model. *Cross-Cultural Communication, 9*(5), 17-22. http://dx.doi.org/10.3968/j.ccc.1923670020130905.2776.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, *53*, 59-68. https://doi.org/10.1016/j.bushor.2009.09.003

Langat, A. C. (2015). Influence of social media on study habits of undergraduate students in Kenyan universities. *International Journal of Novel Research in Humanity and Social Sciences, 2*(4), 42-55.

Maslowski, R. (2001). *School culture and school performance: An explorative study into the organizational culture of secondary schools and their effects*. Twente University Press

Mwangi, A. W. (2013). *Impact of social networking sites on interpersonal relationships among teenagers. A case study of Murang'a east district*. http://erepository.uonbi.ac.ke/handle/11295/59984

Oginni, A. M., Saibu, S. O., Awobodu V. Y., & Alaka M. O. (2016). Influence of students' engagement with social networks on their study habits and cognitive skills in science classrooms. *International Journal of Digital Society*, *7*(3)*,* 1201-1205. https://doi.org/10.20533/ijds.2040.2570.2016.0147

Perpetua, O. E., & Kelechi, E. E. (2018). Effect of social media on the study habits of students of Alvan Ikoku federal college of education, Owerri. *International Scholarly and Scientific Research & Innovation, 12*(1), 220-224.

Piotrowski, C. (2015). Social media: Major topics in dissertation research. *Education, 135*(3), 299-302.

Pritta, C. (2014). *Social Networking Sites Impacts on Interpersonal Communication Skills and Relationships* [Unpublished master thesis]. National Institute of Development Administration. http://libdcms.nida.ac.th/thesis6/2014/b185644.pdf

Rabbani, M., Javeed, A., Asif, M., & Ibrahim, M. (2015). Impact of Social Networking Websites on Students learning. *New Media and Mass Communication, 44*, 45-50.

Saida, S. (2017). Social networking sites (SNS) and family relationship: A study on youths of Dhaka city. *IOSR Journal of Humanities and Social Science*, *22*(4), 46-52. https://doi.org/10.9790/0837-2204024652

Sandra, O. M., & Ismail, N. (2016). The impact of social media on students' academic performance - A case of Malaysia tertiary institution. *International Journal of Education, Learning and Training, 1*(1)*,* 14-21. https://doi.org/10.24924/ijelt/2016.11/v1.iss1/1.13

Salman, H. A., Mohannad, J. A., & Joannes, P. T. H. (2016). Influence of utilization of social networking sites (SNS) on study habits among Saudi nursing students in Ha'il university. *Journal of Biology, Agriculture and Healthcare, 6*(24), 45-56.

Schwab, K. (2016). *The fourth industrial revolution*. World Economic Forum.

Spencer, P. C. (2018). *Social media use and its impact on relationships and emotions* [Unpublished master's dissertations]. Brigham Young University. https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=7927&context=etd

Shakiratul, H. A. R. (2013). Can't live without my FB, LoL: The influence of social networking sites on the communication skills of TESL students. *Procedia - Social and Behavioral Sciences, 134,* 213 – 219. https://doi.org/10.1016/j.sbspro.2014.04.241

Subramanian, K. R. (2017). Influence of social media in interpersonal communication. *International Journal of Scientific Progress and Research*, *109*(38), 70-75.

Sutherland, K., Davis, C., Terton, U., & Visser, I. (2018). University student social media use and its influence on offline engagement in higher educational communities. *Student Success, 9*(2), 13-24). https://doi.org/10.5204/ssj.v9i2.400.

Thomas, L. F. (2005). *The world is flat. A brief history of the twenty-first century* (1st ed). Farrar, Straus and Giroux.

Thomas, O. A., Omotoke, O. O., & Ademola, O. R. (2016). Assessment of social media utilization and study habit of students of tertiary institutions in Katsina state. *Journal of Education and Practice, 7*(3), 178-188.

Tolga, G. (2016). The effects of social networking sites on students' studying and habits. *International Journal of Research in Education and Science, 2*(1), 85-93.

Wong, D. H., Phang, C., Maarop, N., Samya, G. N., Ibrahima, R., Yusoff, R. C. M., Magalingama, P., & Azmia, N. F. M. (2017). Effect of social media on human interpersonal communication: A Review. *Open International Journal of Informatics, 5*(2), 1-6.

Yamakanith, P., & Gurusamy, P. (2014). A study on the influence of social networking sites on the interpersonal relationships of college students in Chennai and Coimbatore cities of India. *International Journal of World Research*, *I*(XI), 48-56.

National Research University Higher School of Economics
Journal of Language & Education Volume 7, Issue 1, 2021

Rind, I. A. (2021). Becoming an ESL Researcher: A Personal Monologue. *Journal of Language and Education, 7*(1), 219-228. https://doi.org/10.17323/jle.2021.10298

Becoming a Qualitative ESL Researcher: A personal monologue

Irfan Ahmed Rind
Sukkur IBA University

Correspondence concerning this article should be addressed to Irfan Ahmed Rind, Airport Road, Delhi Muslim Housing Society, Sukkur, Sindh, Pakistan. E-mail: irfan.rind@iba-suk.edu.pk

This reflective paper narrates my research journey from a naïve researcher to a critic and from a behaviorist to a post-structuralist. It highlights the different philosophical, methodological, and theoretical dilemmas I faced in conceptualizing students' experiences in an English as a Second Language program in higher education during my doctoral studies. This journey is divided into three phases: construction, deconstruction, and reconstruction. In the construction phase, I conceptualized students' experiences from my own established knowledge, which was grounded in my presumptions about teaching and learning. During the deconstruction phase, I questioned my understanding of knowledge and social realities. In the reconstruction phase, I interacted with Phenomenography, Activity Theory, Symbolic Interactionism, Communities of Practice, and Bourdieusian Structuralism. This paper narrates these interactions, focusing mainly on the dilemmas I faced as a researcher. These reflections could be highly beneficial for new researchers who may face the same situations at different stages of their research careers.

Keywords: reflexivity, self-conversation, qualitative research, philosophical stance, social stance, novice researchers

Introduction

This paper presents a reflective journey to becoming a researcher while examining students' learning experiences in an English as a Second Language (ESL) program in a higher education institute as a topic for my Ph.D. (see the summary of the original research in Rind, 2014; and the full research study in Ahmed, 2012). The Ph.D. research was based on the assumption that learning a second language is a complex process in which several individuals and social factors play vital roles in shaping students' ESL learning experiences. In the process of examining these individual and social factors, I reflected on the way I developed a theoretical understanding of students' ESL learning, as well as of myself as a researcher. This paper narrates these reflections during different developmental stages of becoming a researcher. The purpose of these reflections can be effectively summarized in the words of Glesne (2016):

> Learning to reflect on your behavior and thoughts, as well as on the phenomenon under study, creates a means for continuously becoming a better researcher. Becoming a better researcher captures the dynamic nature of the process. Conducting research, like teaching and other complex acts, can be improved; it cannot be mastered (p. xiii).

I have been inspired by researchers such as Cohen and Crabtree (2006), Finlay (2002), Glesne (2016), Malterud (2001), and many others who highlighted the importance of reflection for new researchers. As a new researcher, I experienced strong anxiety at the outset of my Ph.D. research. Although the literature on qualitative research was a source of some comfort, I found that I was the primary instrument of data collection and analysis. This process led me to realize that reflexivity is essential for conducting a compelling research study on my topic. Cohen and Crabtree (2006) defined reflexivity as an attitude for systematically attending to the context of knowledge construction, especially to the effect of the researcher, at every step of the research process. Malterud (2001) argued that "a researcher's background and position will affect what they choose to investigate, the angle of investigation, the methods judged most adequate for this purpose, the findings considered most appropriate, and the framing and communication of conclusions" (p.483-484). The perspective or position of the researcher shapes all research whether qualitative, quantitative, or mix-method. Although I found this

argument very convincing on a theoretical level at the beginning of the research process, I had little idea what this meant in concrete terms.

However, this began to change in the first year of my doctoral studies, when my supervisor asked me to revise my initial research proposal and reconsider my research questions. During this time, I began to reflect on my research by recording self-conversations. These personal narratives helped me identify why I had chosen the topic of the study, how my assumptions about teaching and learning had informed my initial research proposal, and how these assumptions guided my methodology.

This first stage of reflection allowed me to realize that I had undertaken my doctoral studies with numerous assumptions about teaching and learning. Like many novice researchers, I assumed that I was familiar with all the problems faced by ESL students in the university where I wanted to conduct my research. This was mainly due to my positionality as an insider, as I had both studied for four years and taught ESL for a year and a half at that university. Before starting my doctoral studies, I had assumed that students faced problems in the ESL program due to certain behaviors; my interest, therefore, lay in introducing an intervention that would change these behaviors and improve their learning outcomes.

After reflecting on these assumptions, I realized that I was viewing students' learning experiences from a behaviorist perspective. That is, the learning process is a way to construct or modify behaviors. These initial reflections also served as a memory prompt that encouraged the second stage of reflection after I had carried out my research. Reflecting on further self-conversations, I made connections between the methodology literature, decisions taken during the research, the process of reflexivity, and my evolving understanding of the complexities inherent in qualitative research. By analyzing these self-conversations, I became more aware of how I had managed each research stage, the issues and tensions that had arisen, and how I had dealt with these issues as a researcher.

A retrospective examination of my research, therefore, permitted me to make meaningful connections between theory and practice. Moreover, this process of reflection provoked in-depth learning, which may not have been achieved through other methodological means. The experience of reconsidering my initial research proposal for my doctoral studies led me to realize that reflection is an essential mediator of the research process. Most importantly, the reflective self-conversations allowed me to meaningfully construct a personal sense of what it means to become a researcher. This paper attempts to narrate these self-conversations and ways in which I, as a researcher, developed in phases, i.e., construction, deconstruction, and reconstruction.

Construction Phase (Occupied with Assumptions)

Unsurprisingly, the process of research did not seem to be as deep and complicated as I initially assumed. This was mainly because of my established assumptions related to the topic of my choice, i.e., analyzing students' learning experiences in an ESL program in higher education. My positionality as an insider concerning this research (Bryman, 2016), where I lived in the context, experienced the ESL program as a student, and taught the language as an instructor, made me believe that I knew everything about the ESL program. I firmly believed that I could write a convincing Ph.D. thesis without even collecting data from my research participants. Indeed, the very choice of the topic for my doctoral study was based on my experiences, which shaped my assumptions about the weaknesses and strengths of programs of this kind as well as the challenges and opportunities it offered to the students and teachers. Therefore, I firmly felt that I was the best person to conduct this research as I knew almost everything in this domain.

In my initial proposal, I assumed that students faced problems in the ESL program because of certain 'bad behaviors', i.e., not paying attention to lectures, not revising the lessons at home, choosing to sit in the backbenches of the large classroom to avoid participation, and so on. Assuming that I already knew the problems, my focus should have been to introduce some interventions to change these behaviors for a better outcome.

> […] I know what is their [university students'] problem [in ESL] and I know how to fix it. I faced it [the problems] myself [as a student]. Remember the time when I struggled in concentrating on the lectures. Z [anonymized friend's name] struggled because he liked to sit at the

> backbenches. X and Y [anonymized friends' names] hardly studied at home. See, these are their problems, and my students must be doing the same. […] So, let's write the proposal accordingly. Let's propose a new method or something that makes these students change these habits. Think in this direction.
>
> (Reflective note, 12th December, 2008)

However, to reinforce what I already knew, I developed a set of closed-ended questions and conducted a survey. The items in the questionnaire were developed based on my understanding of the challenges and opportunities of the ESL program. Students were required to respond with *yes* or *no* to some items and select the relevant percentage category in response to others. The questionnaire contained items like 'Do you face difficulties in listening skills?', 'How much (in percentage) do you understand the lectures in English?', 'How much (in percentage) do you participate in the class?', and so on. In brief, the items were quite straightforward and did produce the expected results. However, they were far from conducting an in-depth investigation into students' ESL learning experiences. Indeed, the survey was just a way to reassure me of the validity of my assumptions. Even more, I had planned an intervention to remedy these assumed problems and bring changes in students' behaviors before the survey results were even generated.

> The results [from the questionnaire] look promising. […] almost all my assumptions turned out to be true. So, the next task is to (1) present it to X [anonymized Ph.D. supervisor's name], (2) develop an intervention. X is back in his office by 14 Jan, so I should work on developing an intervention. Let's say, on the 14th, I am going to present both findings [of the questionnaires] and intervention to X, and also develop a timeline to implement the intervention.
>
> (Reflective note, 5th January, 2009)

Soon, I realized the simplicity of my methodology through the shallowness of my findings. Of course, my supervisor played a vital role in making me rethink my overall research approach. He advised me to stop thinking about the students' learning experiences and instead reflect upon my understanding of knowledge and social reality. He posed questions like 'what is knowledge for you?', 'how can you acquire this knowledge?', 'how do you see society and social actors?', and 'how do you see the social interaction of social actors?'

> Okay, this was unexpected. […] But let's sort it out one by one. Let's park the questionnaire findings aside, as well as the intervention plan. Let's first listen to X's conversations again [my supervisor allowed me to tape-record our conversations]. So, first, listen and transcribe X's conversation for better understanding. Then highlight all the issues he raised and let's plan to address these issues one by one.
>
> (Reflective note, 14th January, 2009)

Deconstruction Phase (Self-Exploration)

I, like many others, became a researcher from a "background where the practice is seen as unproblematic and expectations of research are directed towards positivist assumptions about generating evidence" (Dunne, Pryor, & Yates, 2005, p.4). However, based on the constant feedback from my supervisor, peers, and my readings, I realized that my research would be judged against criteria involving the questioning of common-sense notions of research practice, knowledge, and social reality. I, therefore, had to first explore the logic of my research practices and explore my understanding of knowledge and social reality, and the tensions inherent in these. This raised some important questions (i.e., how I see social settings and social actors) that I had to reflect upon before I researched students' ESL learning experiences.

How do I See Social Settings and Social Actors?

Initially, I conceptualized students' ESL learning experiences in isolation. In other words, I thought that students' ESL learning experiences were shaped by their own choices and actions, which were independent of the context in which these experiences took place. Soon I realized, as Scott (1999, p.2) suggested, that a researcher must work between "structures/mechanisms of society and the practical consciousness of social actors". A researcher must realize that research in social settings is often recursive in nature (Dunne et al., 2005), which means that if I wanted to understand students' ESL learning experiences, I needed to take into consideration contexts wider than the immediate social action in which I was a research-designer and the

students, as research-objects, were participating. In other words, I needed to take a stance and decide if I considered social actors (students) as independent in making their choices and actions, and solely responsible for the consequences with little influence from the social setting. Or if it was the classroom norms, institutional policies and practices, and/or the wider field of higher education in which this institute operates (social setting) was responsible for students' actions and choices.

> [...] this is interesting, so students' choices are not independent of the context, or are they? Okay, let's think about it. Let's say I am a student, and when I make a choice, it is made without keeping the context in mind, or otherwise? Okay, I make my choices based on my wishes and interests, but I also cannot make any choice that is not acceptable in my social context. So, what does this show? Does this show that my all choices are influenced by the social structure? [...] this is slightly confusing, so let's read a little more on it before deciding.
>
> (Reflective note, 3rd March, 2009)

To make this decision, I needed to understand the "complexity of social settings and the ensuing difficulty this creates for separating the effects of different variables" (Scott, 1999, p.2). My readings into this matter raised an interesting question: is the complexity of any social setting inherent in it, or do I, as a researcher, read the complexity into social action? For example, if I see a student not participating in an ESL classroom, then the potential for an analytic description is almost infinite (Dunne et al., 2005). There are several disciplinary lenses that I can use to describe this phenomenon from micro to macro levels. For example, I may choose a neurological perspective, cognitive perspective, or one of the many sociological perspectives. This means that if I choose a cognitive perspective, my construction of the students and their experiences would be radically different from that of a social perspective. This is mainly because each perspective uses a different lens to see the same individual and phenomenon and draws different conclusions(Dunne et al., 2005).

So which perspective should I choose to get a valid conclusion? Dunne et al. (2005, p.13) suggest that

> Researchers sometimes work with unexamined and vernacular models of the person. In order to make sense of social action, researchers often make the assumption, and it is worth examining, that people act reasonably within a cause-effect model of social action. For psychologists and phenomenologists the answer might be at the level of the individual, while for structuralists of all sorts, from functionalists to Marxists, the individual is largely determined by the wider realities of social forces. For some postmodernists and Foucauldians, the subject may be understood as an effect rather than a cause and the subjective experience of self, seen as a powerful effect of those wider discourses that shape our being.

Does this mean that all the perspectives are authentic, and all the conclusions help in understanding the same phenomenon from different dimensions? Up to that point in time, I had not experienced that yet. I needed to take a social stance to 'find out' students' learning experiences in the ESL program. Before I explored my social stance, I needed to reflect on what knowledge is and how I, as a researcher, could acquire this knowledge.

What is Knowledge and How Can I Acquire It?

To understand students' ESL learning experiences, it was crucial to first question my own assumptions about the nature of knowledge and the world, and how I, as a researcher, express or interpret the reality. This process of interrogation began with questions such as 'what is knowledge?', and 'how is it acquired?'. Seeking answers to these questions led me to epistemology. Heylighen (1997) asserts that epistemology essentially attempts to answer basic questions such as, "what distinguishes true (adequate) knowledge from false (inadequate) knowledge?" (p.1).

> X [anonymized supervisor] said (1) "you are examining students' experience, which means you are trying to know something new or constructing new knowledge. So, what do you, as a researcher, mean by knowledge? How do you conceptualize knowledge? What is knowledge for you? [...] (2) how do you acquire knowledge? [...]" He [supervisor] pointed out that the answer to all questions should align. The most important point to today's meeting [with supervisor] is to clearly explain myepistemological and ontological stance and link it with my methodology and social stance.
>
> (Reflective note, 17th February, 2009)

In the context of social research, such questions translate into issues of scientific methodology: how can one develop theories or models that are better than competing theories? My primary focus was not to develop a theory or model but to select a theory or model that supports a holistic account of students' learning experiences from their perspectives and within their contexts. This approach can be seen in the following discussion where I shifted from a psychological perspective to the sociological constructs of identity, community, and institutional influences. The departure from a psychological perspective was influenced by the approach's detachment of the individual from his or her context (Ashwin, 2012); from my perspective, knowledge of students' learning experiences is inadequate when it is not understood within its specific context. This approach defined my epistemological position as interpretivist, due to my underpinning assumption that "all human actions are meaningful and have to be interpreted and understood within the context of social practices. To make sense of the social world, the researcher needs to understand the meanings that form and are formed by interactive social behavior" (Scott & Usher, 1996, p.18). However, I later acknowledged the fact that interactive social behaviors are also formed within social structures (as discussed in detail in the following discussion).

As an interpretivist, I believed that in order to understand an individual's behavior, attempts should be made to view the world from that individual's viewpoint. However, any individual's viewpoint is inevitably shaped by their particular context. The researcher must obtain access to the individuals' contexts in order to interpret their reality from their point of view. And within an interpretive framework, the researcher attempts to make sense of what he or she is researching. Bryman (2016) labels this process double hermeneutic—through the process of social research, both the subject (the researcher) and the object (other participants in the study) of the research share the position of an interpreter or sense seeker. This means that a researcher must understand how participants view their reality, but also make sense of what he or she makes out of the participants' reality and how he or she defines their findings in the light of existing literature. As Bryman (2016, p.14) asserts,

> When social scientists adopt an interpretivist stance, they are not simply laying bare how members of a social group interpret the world around them. The social scientists will almost certainly be aiming to place the interpretations that have been elicited into a social scientific frame. There is a double interpretation going on: the researchers are providing an interpretation of others' interpretation. Indeed, there is a third level of interpretations going on, because the researchers' interpretations have to be further interpreted in terms of the concepts, theories, and literature of the discipline.

I also realized that the questions of social ontology cannot be detached from issues related to conducting social research. Bryman (2016) argues that ontological assumptions and commitments feed into the formulation of research questions and the way in which research is carried out. Following an interpretivist approach, constructivism forms the basis of the ontological position from which I approached reality, as well as the ontological assumptions that had guided the formulation of my research questions. Unlike objectivism, which argues that social phenomena and their meanings have an existence that is independent of social actors, constructivism asserts that social phenomena and their meanings are continually accomplished by social actors. It implies that social phenomena and categories are not only produced through social interaction, but are also in a constant state of revision, and therefore produce multiple and amendable realities (Bo et al., 2020; Fry et al., 2008; Rind & Ning, 2020). This approach was, therefore, particularly appropriate for my research, which sought to understand the hows and whys of students' ESL learning experiences.

Reconstruction Phase (Informed Conceptualization)

Soon, I realized that the constructivist ontological and interpretivist epistemological stances traverse the fields of phenomenology, hermeneutics, ethnography, and phenomenography. The phenomenologists believe that in order to understand an individual's behavior, one should attempt to view the world from that individual's viewpoint. The job of the researcher is to gain access to the respondents' unique contexts in order to interpret their reality from their point of view (Bryman, 2016). Hermeneutics holds that research is about understanding human behavior rather than explaining it. Ethnographic studies are holistic, originating from the idea that humans are best understood in the fullest possible context (Richardson, 2000). The assumption that underlines phenomenography is that one can understand students' learning and their learning situations from students'

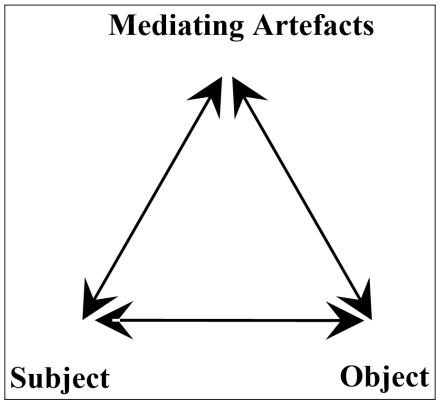
points of view (Marton & Säljö, 1976). Marton and Säljö (1976) also succinctly describe phenomenography as "a change in someone's capability for experiencing something in certain ways" (p. 208).

I chose the phenomenographic approach as it rejects a positivistic view of learning and portrays earlier quantitative research as too "detached" (Entwistle, 1997, p.13). In fact, the phenomenographic perspective emerged as a reaction to traditional, behaviouristic approaches to learning second languages such as B. F. Skinner's Stimulus-Response Theory and Robert Lado's Contrastive Analysis Hypothesis (Ashwin, 2012). From a behaviouristic perspective, language is perceived as a system of habits and second language learning proceeds by producing a response to a stimulus and receiving either positive or negative reinforcement. This is what I believed before I questioned the logic of my research inquiry. I realized that I was a behaviorist and tried to approach students' ESL experiences from a behaviouristic perspective. I assumed that students developed bad habits that hindered their second language learning. I, therefore, needed to develop an intervention that would work as reinforcement to generate desired behaviors. However, locating myself in constructivism and interpretivism, I found phenomenography to be a suitable theoretical framework for exploring students' learning in ESL.

I quickly realized that the phenomenographic perspective has been importantly critiqued for its individualistic approach to understanding students' learning in ESL (see details in Ahmed, 2012); nevertheless, unlike previous behaviouristic and neurological approaches to understanding ESL teaching and learning, the phenomenographic research tradition has tried to consider issues relating to the context. These contextual issues are only covered to a very limited extent in phenomenographic research, but this research tradition can be credited with moving from a notion of fixity to the notion of variability in ESL learning. I, therefore, aimed to locate myself within this theory of variability in learning, but at the same time sought to use this notion of variability within a more context-specific understanding of ESL teaching and learning. In particular, I was interested in understanding how certain variables impact ESL teaching and learning from the students' perspective. Thus, I needed to find a theory that could explain different variables affecting teaching and learning in the context of ESL.

To this end, I found Engeström's (2001) Activity Theory to be a suitable analytical tool for conceptualizing students' ESL learning experiences from their perspective and within their contexts. Activity Theory was developed from the work of Vygotsky (1980), whose triangular model of "a complex, mediated act" (p.40) was fundamental for illustrating the relationship between the human agent and objects in their environment, as mediated by cultural means, tools, and signs. As the model in Figure 1 depicts, the mediation process is two-way (represented by two-sided arrows); individuals actively shape the forces that are active in shaping them, and this process "lies at the heart of the many attempts to develop our understanding of the possibilities for interventions in processes of human learning and development" (Daniels, 2004, p.121).

Figure 1
Vygotsky's triangular model of a complex, mediated act

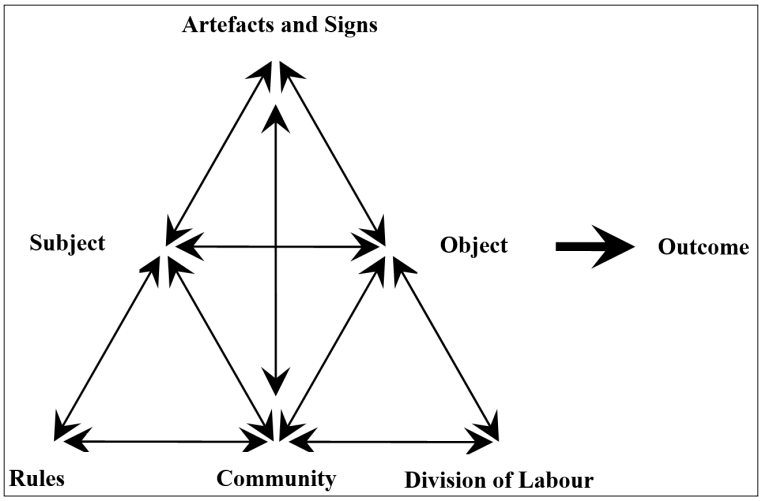


Leont'eV (1982) developed the second generation of Vygotsky's model, which distinguished between individual action and collective activity. Leont'eV viewed the activity system as incorporating three hierarchical processes driven by a different individual or communal motive. These ideas were further developed by Engestrom (1987) as illustrated in Figure 2.

Within the Activity Theory approach, learning and other human activities are understood as being integrated into a larger system, or an activity. An activity is a theoretical entity, discerned, described, and used by the researchers to consider the socially-based nature of the human activity. The interacting constituents of the activity continuously develop into new forms that together create an activity. Used in this way, Activity Theory serves as a clarifying and descriptive tool rather than a strongly predictive theory (Nardi, 1996) and does not prescribe any particular way of understanding learning or a fixed ontology(Berglund, 2004; Havnes, 2004). Several researchers found this theoretical approach useful for understanding ESL teaching and learning (Hawkins, 2008; Lantolf, 2014; McCafferty, 2002; Roebuck, 2000). I also found it useful and, therefore, adopted it to generate a model for a student learning activity in the ESL program by identifying different variables (see Figure 3). This model highlighted important factors such as students' identities, the community within the ESL classroom, and institutional influences on different variables that shape students' learning experiences.

Figure 2

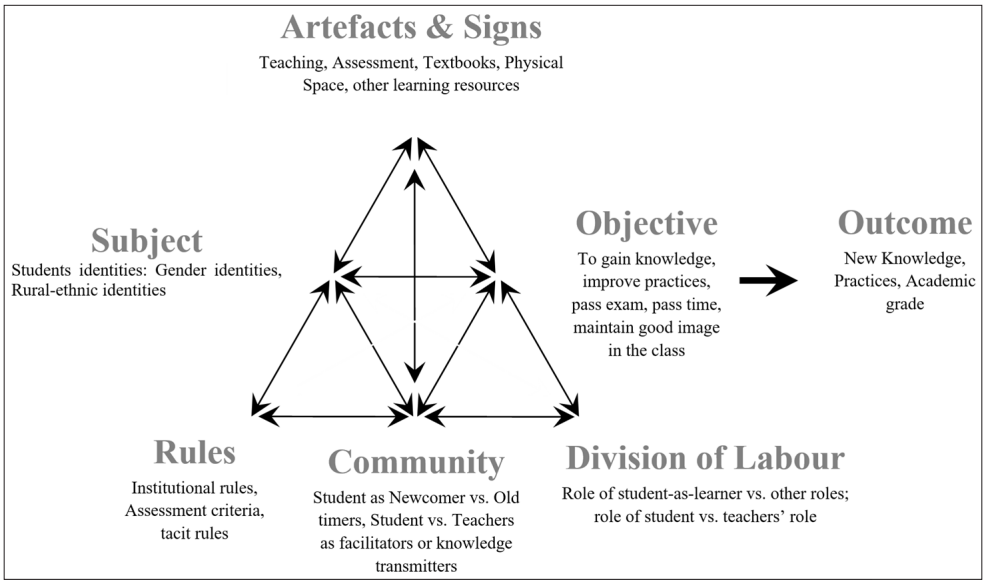
Engeström's model of an activity system



Note. Adapted from "Learning by expanding: An activity-theoretical approach to developmental research," by Y. Engestrom, 1987, p. 78. Copyright 1987 by the Orienta-Konsultit.

Figure 3

My conceptual understanding of students' learning experiences using Activity Theory

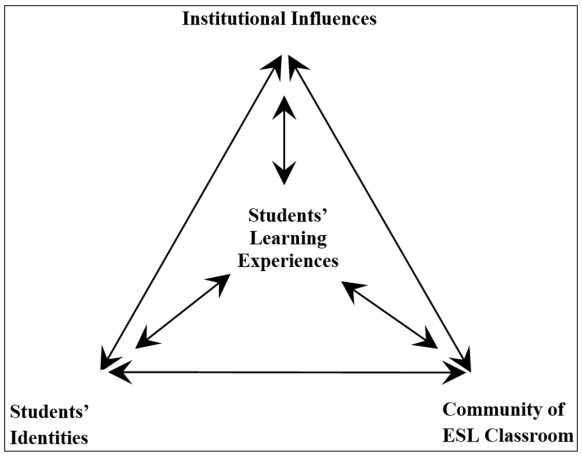


Note. Adapted from "Investigating students' experiences of learning English as a second language at the University of Sindh, Jamshoro, Pakistan" by I. Ahmed, 2012, The Sindh University Journal of Education-SUJE, 45(1), 1–20. Copyright 2012 by I. Ahmed.

However, Activity Theory does not offer conceptual insights into the notions of institutional influence, community, and identity. I, therefore, decided to use Activity Theory to a limited extent, i.e., to explain variables in students' learning experiences (Rind, 2016a). I did not adopt it as an overarching theoretical framework for my research. Nevertheless, my brief interactions with Activity Theory were useful for enabling me to finally focus on the use of institutional influence, community, and identity for further analysis. I wanted to employ the concept of institutional influence to examine the impact of the university's policies and practices on the teaching and learning in the ESL program (for details see Rind & Kadiwal, 2016). I saw students and teachers in the ESL classroom as a community, therefore I wanted to use the concept of community to examine the student-student and student-teacher interactions (for details see Rind, 2016b). I also wanted to use the notion of identity to examine the interactions among students' gender identities (for details see Rind, 2015), rural-ethnic identities (for details see Rind et al., 2016), and disciplinary identities (for details see Rind & Alhawsawi, 2013) within different aspects of the ESL program (see Figure 4).

Figure 4

My conceptual understanding of students' learning experiences using the sociological lenses of identity, community, and institutional influences



Note. Adapted from "Investigating students' experiences of learning English as a second language at the University of Sindh, Jamshoro, Pakistan" by I. Ahmed, 2012, The Sindh University Journal of Education-SUJE, 45(1). Copyright 2012 by I. Ahmed.

This conceptual framework offered a three-dimensional view (i.e., at a micro level, the interactions of students' different identities within the ESL program; at the meso level, the interactions of students with teachers and peers; and the macro level, the interactions of students' experiences with the policies and practice of the institute) into students' ESL learning experiences. These experiences are presented in detail in different published works. The interactions of students' different identities are presented in Rind and Alhawsawi (2013), Rind (2015), and Rind et al. (2016). Likewise, the detailed findings on the interactions of students with peers and teachers are presented in Rind (2016b). Finally, the details of how institutional influences shape students' experiences in ESL are presented in Rind and Kadiwal (2016).

Conclusion

In this paper, I attempted to reflect upon the different processes that shaped my positionality, perspective, beliefs, and values in relation to the research that I conducted to evaluate students' learning experiences with ESL in higher education. It is important because, as a qualitative researcher, I have often constructed the reality of the participants as the human research instrument. Therefore, reflexivity becomes a defining feature for me as a qualitative researcher. Considering that a qualitative researcher should attempt to be aware of their roles in the (co)-construction of knowledge, I tried to make it clear in my narratives how intersubjective elements impacted my data collection and analysis to enhance the trustworthiness, transparency and accountability of my research. However, the process of engaging in reflexivity is full of muddy ambiguity and multiple trials as researchers are needed to negotiate the swamp of interminable deconstructions, self-analysis, and self-disclosure. This ambiguity is more confusing for novice qualitative researchers who struggle with issues of subjectivity in their research.

Dealing with these issues as a novice qualitative researcher, I reflected upon how I have developed as a researcher in different stages while navigating my way to a piece of research that is both in-depth and trustworthy. Earlier research focusing on the dilemma of qualitative researchers promoted subjectivity as an opportunity rather than a problem if researchers effectively use reflexivity. However, no universal guidelines are given for the effective use of reflexivity. This paper contributes by presenting examples of how self-conversations and reflective notes became a great opportunity for me as a novice qualitative researcher to keep a check on my subjectivity. At the same time, they also demonstrated how I used these self-conversation and reflective notes to show my presence in my research to produce reflective yet authentic research. Presenting more and more reflective narratives to novice researchers may help them understand this process. This paper attempts to do the same, by presenting my reflections on my research processes and the ways I dealt with my subjectivity when developing a theoretical framework to understand students' learning experiences. Following some of these steps, many novice researchers may balance their subjectivity in their research, ensure the trustworthiness of their methodology, and achieve in-depth context specific analysis of their topics.

Acknowledgments

I would like to thank Prof. Sayed Yusuf, my Ph.D. supervisor, who guided me along this entire research journey. I would also acknowledge the contribution of Dr. Sajjadllah Alhawsawi for being an exceptional critical friend on this journey.

Reference

Ahmed, I. (2012). Investigating students' experiences of learning English as a second language at the University of Sindh, Jamshoro, Pakistan. *The Sindh University Journal of Education-SUJE*, *45*(1), 1–20. http://sujo.usindh.edu.pk/index.php/SUJE/article/view/2320

Ashwin, P. (2012). *Analysing teaching-learning interactions in higher education: Accounting for structure and agency*. A&C Black.

Berglund, A. (2004). A framework to study learning in a complex learning environment. *Association for Learning Technology Journal*, *12*(1), 65–79. https://doi.org/10.3402/rlt.v12i1.21457

Bo, N., Rind, I. A., & Asad, M. (2020). Influence of teacher educators on the development of prospective teachers' personal epistemology and tolerance. *Sage Open*, *10*(1), 1–14. https://doi.org/https://doi.org/10.1177/2158244020914639

Bryman, A. (2016). *Social research methods*. Oxford University Press.

Cohen, D., & Crabtree, B. (2006). *Qualitative research guidelines project*. http://www.qualres.org/index.html

Daniels, H. (2004). Activity theory, discourse and Bernstein. *Educational Review*, *56*(2), 121–132. https://doi.org/10.1080/0031910410001693218

Dunne, M., Pryor, J., & Yates, P. (2005). *Becoming a researcher: A research companion for the social sciences: A companion to the research process*. McGraw-Hill Education.

Engestrom, Y. (1987). *Learning by expanding: An activity-theoretical approach to developmental research*. Orienta-Konsultit.

Engeström, Y. (2001). Expansive learning at work: Toward an activity theoretical reconceptualization. *Journal of Education and Work*, *14*(1), 133–156. https://www.tandfonline.com/doi/abs/10.1080/13639080020028747

Entwistle, N. (1997). Contrasting perspectives on learning. *The Experience of Learning*, *2*, 3–22.

Finlay, L. (2002). Negotiating the swamp: the opportunity and challenge of reflexivity in research practice. *Qualitative Research*, *2*(2), 209–230. https://doi.org/10.1177/146879410200200205

Fry, H., Ketteridge, S., & Marshall, S. (2008). *A handbook for teaching and learning in higher education: Enhancing academic practice*. Routledge.

Glesne, C. (2016). *Becoming qualitative researchers: An introduction*. ERIC.

Havnes, A. (2004). Examination and learning: an activity-theoretical analysis of the relationship between assessment and educational practice. *Assessment & Evaluation in Higher Education*, *29*(2), 159–176. https://doi.org/10.1080/0260293042000188456

Hawkins, B. (2008). Using sociocultural theory to examine the context (s) of language learning and teaching. *Studies in Applied Linguistics and TESOL*, *8*(1). https://doi.org/10.7916/salt.v8i1.1498

Heylighen, F. (1993). *Epistemology: Introduction*. http://pespmc1.vub.ac.be/EPISTEMI.html

Hopman, J. (2021). Fieldwork supervision: Supporting ethical reflexivity to enhance research analysis. *International Journal of Research & Method in Education*, *44*(1), 41–52. https://doi.org/10.1080/1743727X.2019.1706467

Lantolf, J. (2014). (S) econd (L) anguage (A) ctivity theory: Understanding second language learners as people. In *Learner contributions to language learning: New directions in research* (pp. 141–158). Taylor and Francis.

Leont'eV, A. N. (1982). Problems of the development of the mind. In *Progress Publishers*. Progress.

Malterud, K. (2001). Qualitative research: standards, challenges, and guidelines. *The Lancet*, *358*(9280), 483–488. 10.1016/S0140-6736(01)05627-6

Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I—Outcome and process. *British Journal of Educational Psychology*, *46*(1), 4–11.

McCafferty, S. G. (2002). Gesture and creating zones of proximal development for second language learning. *The Modern Language Journal*, *86*(2), 192–203. https://doi.org/10.1111/1540-4781.00144

Nardi, B. A. (1996). *Context and consciousness: activity theory and human-computer interaction*. MIT press.

Richardson, L. (2000). Evaluating ethnography. *Qualitative Inquiry*, *6*(2), 253–255. https://doi.org/10.1177/107780040000600207

Rind, I. A. (2014). Investigating students' experiences in an ESL programme at the University of Sindh, Jamshoro, Pakistan. *Compare*, *44*(6), 1002–1004. https://doi.org/10.1080/03057925.2013.822244

Rind, I. A. (2015). Gender identities and female students' learning experiences in studying English as Second Language at a Pakistani University. *Cogent Education*, *2*, 1115574. https://doi.org/10.1080/2331186X.2015.1115574

Rind, I. A. (2016a). Conceptualizing students' learning experiences in English as second language in higher education from structure and agency. *Cogent Social Sciences*, *2*(1), 1191978.

Rind, I. A. (2016b). How newly appointed ESL teachers' beliefs are translated into their pedagogic strategies. *Journal of Research & Method in Education*, *6*(1), 2320–7388. https://doi.org/10.9790/7388-0613124132

Rind, I. A., & Alhawsawi, S. (2013). The effects of students' previous learning on graph-comprehension in the English as second language (ESL) textbooks in a Pakistani university. *EducationalFutures*, *5*(2), 44–58. https://educationstudies.org.uk/?p=630

Rind, I. A., & Kadiwal, L. (2016). Analysing institutional influences on teaching–learning practices of English as second language programme in a Pakistani university. *Cogent Education*, *3*(1), 1160606. https://doi.org/10.1080/2331186X.2016.1160606

Rind, I. A., & Ning, B. (2020). Evaluating scientific thinking among Shanghai's students of high and low performing schools. *The Journal of Educational Research*, *113*, 364–373. https://doi.org/10.1080/00220671.2020.1832430

Rind, I. A., Shahriar, A., & Fatima, S. (2016). Rural-ethnic identities & students'learning experiences in english as second language programme in a public sector university of Pakistan. *The Sindh University Journal of Education-SUJE*, *45*(1), 1–20.

Roebuck, R. (2000). Subjects speak out: How learners position themselves in a Psycholinguistic task. *I*n J.P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 79-85). Oxford University Press.

Scott, D. (1999). *Values and Educational Research*. Institute of Education.

Scott, David, & Usher, R. (1996). *Understanding educational research*. Psychology Press.

Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes*. Harvard University Press.

National Research University Higher School of Economics
Journal of Language & Education Volume 7, Issue 1, 2021

Ibna Seraj, P. M., & Hadina, H. (2021). A Systematic Overview of Issues for Developing EFL Learners' Oral English Communication Skills. *Journal of Language and Education, 7*(1), 229-240. https://doi.org/10.17323/jle.2021.10737

A Systematic Overview of Issues for Developing EFL Learners' Oral English Communication Skills

Prodhan Mahbub Ibna Seraj, Hadina Habil
Universiti Teknologi Malaysia

Correspondence concerning this article should be addressed to Prodhan Mahbub Ibna Seraj, TESL, School of Education, Universiti Teknologi Malaysia, 81310, Johor Bahru, Malaysia.
E-mail: mahbub@graduate.utm.my

The objective of this systematic review is to present a critical overview of current studies to explore issues such as the factors causing EFL learners' poor oral performance and the teaching and assessment methods of oral English communication skills (OECSs) for developing tertiary level learners' OECSs in EFL contexts. For this purpose, 51 empirical studies of the 2907 retrieved from SpringerLink, SCOPUS, Web of Science, and the Google Scholar database that were published between 2010 and 2019 in different EFL contexts were analyzed. This study followed the Preferred Reporting Items for Systematic Review and Meta-analyses (PRISMA) guidelines and was analyzed thematically using NVIVO 12, followed by the Mendeley reference management software. Studies that were conducted in native English contexts and non-empirical studies were excluded from consideration for this study. The findings showed that the environmental factor was the primary factor for learners' poor OECS performance in EFL contexts. For the method of teaching and assessment of learners' OECSs, the use of technology is rapidly increasing in different EFL contexts. This study suggests some implications for both future researchers and academics for developing EFL learners' oral English communication skills dealing with environmental, psychological, and linguistic factors along with teaching and learning resources at the tertiary level in EFL contexts.

Keywords: oral English communication skills, EFL learners' problems, teaching method, assessment method, systematic literature review

Introduction

For EFL (English as a Foreign Language) learners, gaining mastery over oral English communication skills is always more challenging than other language skills like reading, listening, and writing. In the Fourth Industrial Revolution era, the demand for oral English communication skills is inevitably increased for the use of IoT as a means of communication globally. The word 'communication' originated from the Latin word 'communicare' (meaning to share), and the French word 'communis' (common) means one's experience sharing with others (Okoro, 2007). Oliver and Philp (2014) defined oral communication as "the speaking and listening that occurs in real-time (i.e., in the present) in communicative exchanges (i.e., interactions)" (p. 5). This kind of exchange depends on the process of meaning-making and understanding of both communicators (Adler & Rodman, 2009). The significance of oral communication skills in mastering a second language was accentuated by Ellis (2003) when he highlighted that second and foreign language acquisition encompasses the ability to use the sound and grammar systems to communicate meanings. Oral communication skills make face-to-face interactions possible, as well as long-distance interactions such as online conversations.

In this techno-drenched world, for maintaining online and offline communication, EFL learners need to be skilled in oral communication including various vital aspects. In today's global context, they often need oral English communication skills for expressing opinions, making arguments, offering explanations, transmitting information, and making impressions upon others in their personal lives, future workplaces, social interactions, and political endeavors (Rahman, 2010).

Moreover, it has become an aphorism in EFL contexts that a graduate with a good command of oral English communication skills has a better opportunity in their professional life than one who does not (Idrus, 2016;

Kunioshi, Noguchi, Hayashi, & Tojo, 2012). Hence, preparing for the competitive and explosive nature of market value in both EFL and native contexts, graduates need to improve their skills (Subramaniam & Harun, 2012). Moreover, OECSs require distinct attention for learning and teaching in education programs for all disciplines so that students could be operative in both their academic and professional lives (Kunioshi et al., 2012; Malthus & Lu, 2012; Mitchell & Eng, 2010; Nikolic et al., 2018; Tuomaitė & Zajankauskaitė, 2017).

However, globalization increasingly demands the scholarly use of English for multiple purposes (Jim, 2018). Despite the importance of OECSs in the EFL contexts, tertiary-level learners' oral performance is not up to the mark, and in some cases it is still incomprehensible (Afshar & Asakereh, 2016). Studies have pointed out a number of factors that are responsible for developing learners' OECSs performance in the EFL context(Chen & Goh, 2011; Fitria & Salwa, 2018; Sinwongsuwat, & Bojanić, 2015; Soomro et al., 2019). Other studies employed different teaching and assessment methods for bringing innovation to existing traditional teaching and assessment methods in different EFL contexts for improving tertiary level learners' oral skills (Darmi & Albion, 2017; Lana et al., 2018)

Thus, the objectives of this systematic literature review (SLR) critically presents an overview of the factors responsible for the poor oral performance of EFL learners and the methods for teaching and assessing oral English communication skills at the tertiary level in EFL contexts from published empirical studies from 2010 to 2019 while maintaining the principles of the PRISMA model as inclusion and exclusion criteria of the studies. The significance of this SLR study is to present different issues like common problems, teaching strategies, and assessment methods for developing learners' OECSs in EFL contexts. This study could also provide insights for EFL policymakers, teachers, and educators to help their learners improve by provising an overview of the findings of this study for redesigning oral English communication pedagogy for future EFL contexts.

The researchers set the following research questions for this systematic literature review:

1. What are the factors responsible for EFL learners' poor oral English communication skills performance?
2. What are the teaching methods employed for developing learners' oral English communication skills in EFL contexts?
3. What types of assessment methods are employed for developing learners' oral English communication skills in EFL contexts?

Materials and Methods

Literature Search

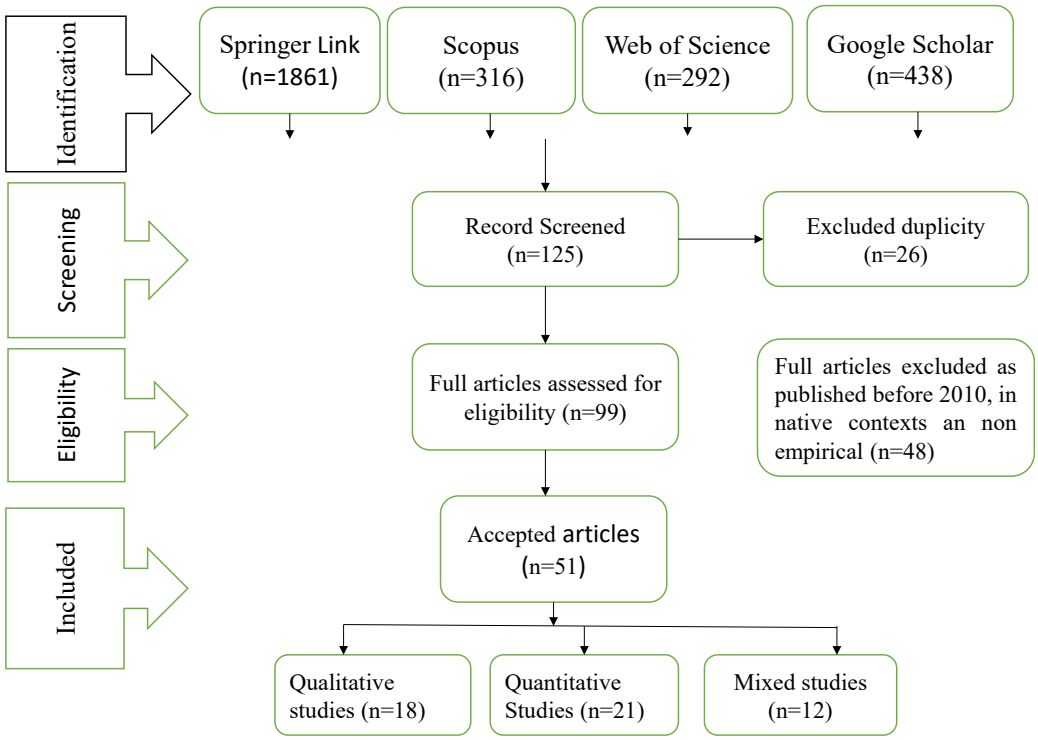
Reputable online databases such as SpringerLink, Scopus, Web of Science, and Google Scholar from 2010 through December 2019 were queried for studies addressing the teaching and learning of oral English communication skills in EFL contexts. All articles published in native English-speaking countries, those with non-empirical or anonymous contexts, and articles written earlier than 2010 were excluded. We included articles related to oral skills, oral communication skills, and oral English communication skills teaching and learning. Search terms were the phrases "oral communication skills" and "oral English communication skills." For only the Google Scholar database, the phrases were kept within an inverted comma for identifying topic related to literature. The Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines were followed (Moher et al., 2009) for selecting more relevant research studies on the topic and sorting for scrutiny was done with NVIVO for developing themes systematically on the issues for this study.

Selection of Articles

A PRISMA flow diagram for the search and inclusion/exclusion process is presented in Figure 1. A total of 2907 titles and/or abstracts assembled from the databases were reviewed by the researchers and 2782 articles were excluded, as they did not address any of the research questions after being screened by their titles. The remaining 125 were checked for duplicity resulting in 99 useable articles. For finding the most relevant articles to the SLR questions for full review, the studies such as reviews, conceptual articles, and those that were conducted in native contexts were excluded from the consideration of the analysis. Finally, 51 studies met the

criteria of this SLR for quality of evidence and pertinence to the research questions. Some articles were analyzed for multiple questions.

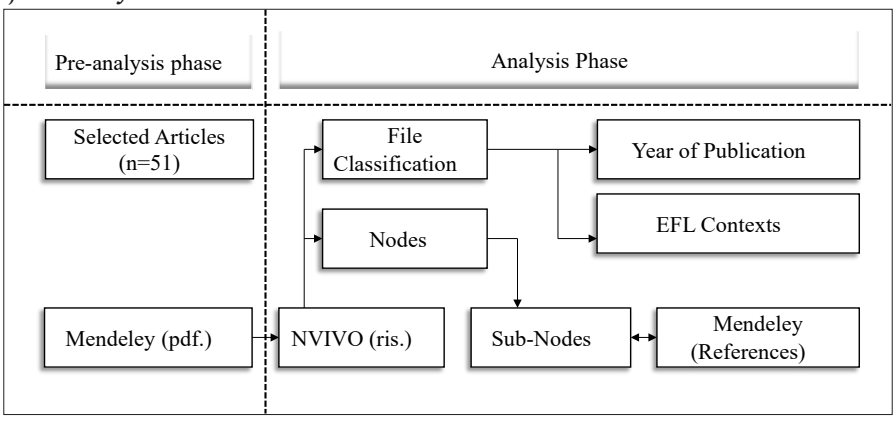
Figure 1
PRISMA flow diagram



Preparing Phase

The preparation phases of the analysis comprised all the processes necessary for the analysis (Figure 2). In the pre-analysis phase, Mendeley was used for making references database for all the primary studies (n=50). From the Mendeley, all the articles were exported as .ris files to import into NVIVO 12 software. The imported articles were classified for ensuring the targeted year of publication and EFL contexts with NVIVO. Then, pre-codes were developed as primary nodes in NVIVO 12 for learners' poor performance and teaching and assessment methods. Under the primary nodes, sub-nodes were built to get a scrutinized analysis of the texts. These nodes and sub-nodes were the answers to the SLR questions.

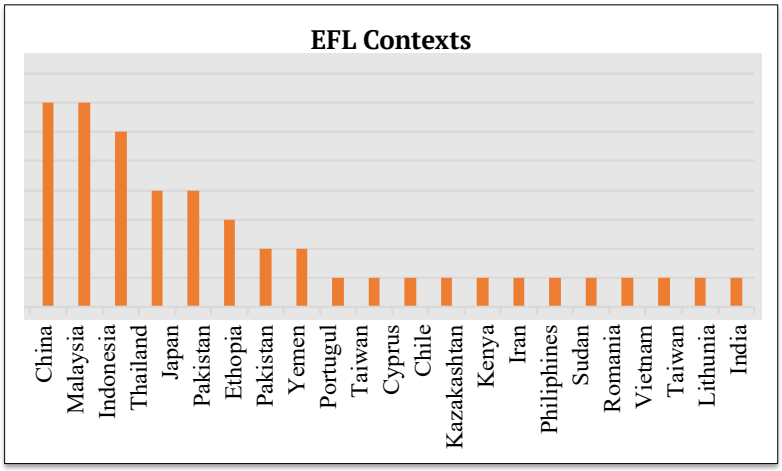
Figure 2
Preparing phases of the analysis



Analysis Phase

All the selected articles conducted in EFL contexts (Figure 3) were prepared for thematic analysis using the NVIVO program for finding out the answers to the SLR questions. NVIVO is a software program that provides tools for analyzing qualitative data in qualitative or mixed-method research. Thematic analysis was used to categorize facts for uncovering meaning (Fereday & Muir-Cochrane, 2006). In this study, thematic analysis was used to analyse texts for identifying themes pertaining to factors causing learners' poor performance, teaching strategies, and assessment methods for developing learners' OECS performance in EFL contexts. Researchers believe that if EFL teachers and practitioners understand the factors causing learners' poor oral performance, effective teaching strategies, and assessment methods they could concentrate more on developing learners' oral skills efficiently. Therefore, the SLR is considered indispensable for reviewing all of the related articles thoroughly and investigating the gathered data via thematic analysis. Consequently, from the SLR questions, three concepts, such as factors causing learners' poor performance, teaching strategies, and assessment methods, were selected as the major themes.Under these themes, there were sub-themes developed such as learners' passiveness, limited oral practice, learners' mixed abilities, and large class sizes for environmental factors; shyness, confidence, anxiety, and emotions for psychological factors; accuracy, fluency, pronunciation, low vocabulary levels, and lack of language knowledge for linguistic factors; and teachers' low efficacy and lack of teaching resources for teaching resources factors. Similarly, under the theme of teaching methods, technology-based, task-based, interactive, microteaching and communicative methods were developed as sub-themes. Finally, for the assessment method, mobile-based, university-based, teacher-based, and mixed-panel assessment methods were developed as sub-themes. These themes and sub-themes guided the researchers to reach the objectives of this study.

Figure 3
EFL contexts of the studies



Results and Discussion

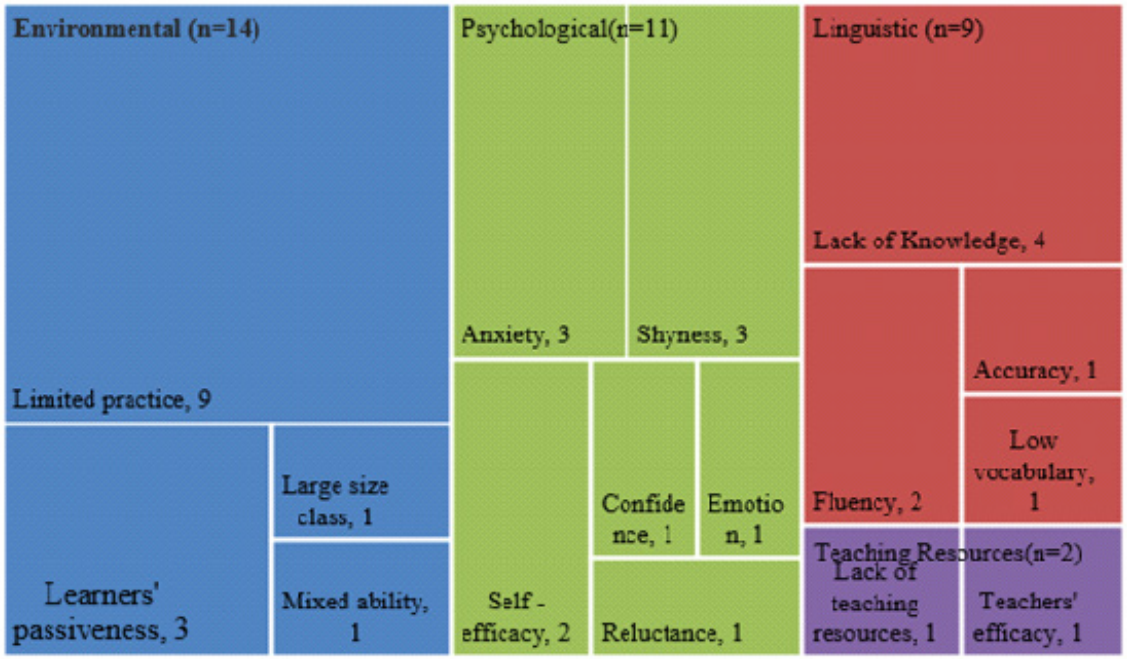
This section illustrates the detailed answers to the SLR questions for this review. The researchers identified the factors that were responsible for EFL learners' poor performance in oral English communication, teaching, and assessment techniques for developing oral skills in EFL contexts from the previous studies. The final section offers a discussion on insights for dealing with these problems, introducing teaching and assessment techniques for developing learners' oral English communication skills in EFL contexts.

Factors for EFL Learners' Poor Oral English Communication Skills

Based on the analysis of the studies, the results of the first research question are presented in Figure 4. Of the 51 studies, 34 studies reported different factors that were responsible for developing learners' oral skills. The hierarchy chart showed that there were four different factors, e.g., environmental (n=14), psychological (n=11), linguistic (n=9), and teaching resources (n=2) influencing the development of

learners' oral skills at the tertiary level in EFL contexts. The majority of the studies stated that environmental factors were the prime cause of EFL learners' oral skills developing. On the other hand, the fewest number of studies reported that the teaching resources factor was the significant factor for EFL learners' oral skills development at the tertiary level.Other studies reported more than one factor. Some of the sub-factors will be discussed below

Figure 4
Hierarchy chart of the factors



Environmental Factors

Environmental factors refer to the nature of non-native learning and teaching settings. The findings showed there were four different types of environmental factors: limited oral practice environment (n=9), learner passiveness (n=3), large class sizes (n=1), and learners' mixed abilities (n=1). Most of the studies illustrated that limited opportunities to practice oral skills inside and outside the classroom were the prime environmental factor for the poor development of EFL learners' oral skills in EFL contexts (Aeni et al., 2017; Alam & Uddin, 2013; Ella & Dapudong, 2014; Fitria & Salwa, 2018; Ismail et al., 2018; Rahman, 2010). As there was a lack of an oral practice environment in EFL contexts, learners rarely got the opportunity to practice outside the classroom (Soomro et al., 2019).

The findings showed that the traditional lecture-based teaching method still existed in the EFL contexts for teaching oral skills. In this method, learners played passive roles in their lessons and this hindered them from developing oral communication skills (Abid, 2018; Bruner et al., 2015). Moreover, the findings also showed that EFL learners were not motivated to learn oral skills. EFL learners' mixed abilities and inadequate educational backgrounds were also responsible for their poor performance (Bruner et al., 2015; Soomro et al., 2019). In an EFL class, there were different levels of learners, e.g., low, mid, and high. Whereas low-level learners tried hard and lost interest in the end, high-level learners developed quickly and received more attention. The last environmental factor was large class sizes (Chen & Goh, 2011). Large class sizes were a common phenomenon in many EFL contexts. In the large classes, it was difficult for the teacher to manage all the learners for engaging in oral tasks.

Under environmental factors, learners' limited practice, learners' passiveness, learners' mixed abilities, and large class sizes were found to be major causes of poor performance in EFL contexts. Other environmental factors like the influence of L1 on OECSs and learners' exposure to OECSs outside the classroom were unexplored.

Psychological Factors

The results showed that there were six types of psychological factors, e.g., anxiety (n=3), shyness (n=3), self-efficacy (n=2), confidence (n=1), emotions (n=1), and reluctance (n=1), that were responsible for causing EFL learners' poor OECSs in EFL contexts. Both anxiety and shyness were equally addressed in these studies. Some studies suggested that anxiety was a crucial psychological factor that led to learners' disruption, sweating, and trembling while they were speaking. These were due to a lack of practice inside and outside the classroom, insufficient vocabulary, low L2 motivation, and fear of criticism (Aeni et al., 2017; Kasbi & Shirvan, 2017; MacWhinnie & Mitchell, 2017).For developing oral skills, learners' shyness that resulting from nervousness, inferiority complexes, and having a fear of doing something wrong, was one of the major psychological factors in EFL contexts (Aeni et al., 2017; Alam & Uddin, 2013). The findings also showed that EFL learners' lack of confidence, resulting from a lack of oral practice and linguistic knowledge, was another challenging factor for oral skills learning (Fitria & Salwa, 2018).

According to the results, the emotional factor was another psychological factor that emerged from lack of emotional support from teachers, family members, and friends. The emotional factor influenced female learners more than male learners in Kazakhstan (Ismail et al., 2018). Additionally, self-efficacy was another psychological factor related to learners' aptitudes, attitudes, and aspirations that influenced learners' oral skill development (Idrus et al., 2011).Finally, the results showed that learners' reluctance to speak in English emerged from a lack of confidence for learning oral skills in EFL contexts (Şimon, 2014). All these psychological factors contributed to learners' nervousness when they are communicating using oral skills.

Regarding psychological factors, this study found learners' shyness, anxiety, self-efficacy, reluctance, emotions, and confidence rather than motivation, nervousness, and fear to speak from the previous studies. The findings would suggest that more studies need to be addressed in this regard for developing learners' OECSs.

Linguistic Factors

The results revealed that there were four different types of linguistic factors, e.g., lack of language knowledge (n=4), accuracy (n=1), fluency (n=2), and low vocabulary (n=1) were responsible for causing EFL learners' poor OECSs in EFL contexts. According to the results, learners' lack of grammatical knowledge was prominent. When learners spoke, they made mistakes using improper grammar (Chen & Goh, 2011; Soomro et al., 2019; Zukurnain & Kaur, 2014). The other linguistic factor related to accuracy revealed that learners were unable to produce accurate speech in terms of pronunciation, presentation, and communicating ideas (Alghail & Mahfoodh, 2019). Fluency was a linguistic factor related to speech rate, spontaneity, and frequency of pauses that was also responsible for EFL learners' poor oral skills (Alghail & Mahfoodh, 2019; Onoda, 2014). Finally, EFL learners' low vocabulary level was another linguistic factor for hindering the developing oral skills performance (Soomro et al., 2019).

Thus, the findings showed that the lack of language knowledge, fluency, accuracy, pronunciation and vocabulary were linguistic problems for developing learners' oral skills. But learners' problems related to the suprasegmental features of speech, especially stress and intonation, were not found. Therefore, the findings of this study would recommend more studies addressing these issues in future research.

Teaching Resources

According to the results, a lack of teaching resources was also a challenging factor for developing learners' oral skills at the tertiary level in EFL contexts. The findings reported that teaching resources were related to the lack of teaching materials and teachers' efficacy. A lack of teaching materials indicated that textbooks were not appropriate for developing learners' oral skills. Moreover, in EFL contexts, there was a lack of audio-video laboratories and multimedia equipment for teaching-learning oral skills. On the other hand, teachers had low efficacy as they had a deficiency in oral skills and pedagogical knowledge for developing learners' oral skills in EFL contexts (Chen & Goh, 2011). In some cases, both learners' and teachers' low efficacy was also responsible for poor development in learners' oral skills in EFL contexts (Idrus et al., 2013).

However, some of the other problems related to teaching resources factors, e.g. the curriculum, lesson plans, and oral lesson activities were not addressed. Thus, this study would suggest focusing on these issues in the future.

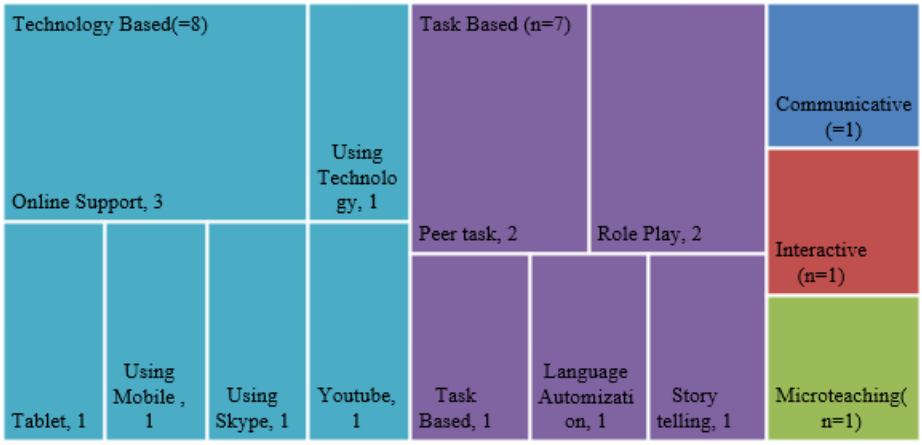
Teaching Methods

Based on the analysis of the studies, the results of the second research question are presented in Figure 5. Of the 51 studies, 18 studies reported five different types of teaching methods e.g., technology-based (n=9), task-based (n=7), communicative(n=1), interactive (n=1), and microteaching (n=1) employed for developing learners' oral skills in different EFL contexts. The findings showed that bringing innovation into traditional teaching methods for creating a new environment for oral practice inside and outside the classroom was essential for developing learners' oral skills (Bunjan & Suppasetseree, 2017; Darmi & Albion, 2017; Subramaniam et al., 2013).

According to the results, the highest number of studies used different types of technology e.g., online-based (n=3), tablet (n=1), Skype (n=1), YouTube (n=1), and mobile (n=1) as a teaching tool for developing learners' oral skills. The study found that technology from audiobooks to language apps helped develop oral skills through gaining correct pronunciation and the correct use of words in EFL contexts (Lana et al., 2018). The largest number of studies used online-based teaching methods that supported the virtual language community, which helped to develop learners' oral skills (Bunjan & Suppasetseree, 2017; Hart, 2016; Kunioshi et al., 2012). According to Darmi and Albion (2017), classes that employed mobile-based learning had significant improvement in enhancing OECSs in Malaysian contexts.

Using tablets in ELT teaching in the Portuguese context, iBook and other educational resources assisted learners in improving oral skills (Couvaneiro & Pedro, 2015). Integrating Skype with teaching methodologies was found to be very effective and useful for reducing learner anxiety when attempting to develop oral skills (Herlo et al., 2017). Moreover, the inclusion of YouTube into the classroom made learners active participants for learning OECSs (Subramaniam et al., 2013).

Figure 5
Hierarchy chart of teaching method



According to the results, there were five different types of task-based teaching methods, e.g., peer tasks (n=2), role play (n=2), tasks-based (n=1), language atomization (n=1), and storytelling task-based (n=1) found in different EFL contexts. Rahman (2010) reported that the task-based approach through involving learners in different tasks simulating real-life situations assisted EFL learners in enhancing their OECSs. The findings revealed that task-based teaching with peer-mediated strategies helped to develop learners' OECSs in different EFL contexts (Mamo et al., 2015; Tsai, 2019). The role-play teaching method also had a significant impact on increasing learners' motivation and confidence for developing OECSs (Nasihah, 2019; Tipmontree & Tasanameelarp, 2018). The findings showed that the storytelling task-based teaching method was found useful for the development of adult learners' oral communication skills in the Nepalese context. (Manzano, 2018).

The results also showed some other teaching methods were used to develop learners' OECSs in different contexts. For example, the communicative teaching method (n=1) was found effective for enhancing EFL

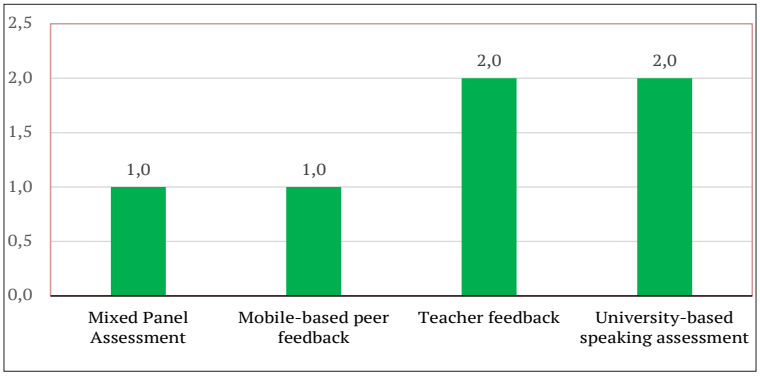
learners' OECSs. In the Malaysian context, Zukurnain and Kaur (2014) reported that a communication strategy that promotes spontaneity between speakers to help them agree on meaning in the target language could help them overcome the challenges faced by learners when learning OECSs. By developing communicative strategies, EFL learners gained oral competence (Thu et al., 2016).In the context of Kenya, the interactive teaching method (n=1) used by EFL teachers for providing opportunities for language use, accuracy, proficiency, and immediate feedback had a positive impact on learners' OECSs performance (Jessica et al., 2015). In the context of Chile, microteaching (n=1), which engaged learners actively in the learning session, had an influential impact on learners' oral skills development (Herrera & Vielma, 2018).

From the results of this study, there were five different teaching methods found: technology-based, task-based, microteaching, interactive, and communicative teaching. Although the results illustrated that the trend of using technology in teaching and learning OECSs in EFL contexts is rapidly increasing, there is still a lack of evidence for a solid theoretical framework in this regard. There are very few studies that dealt with the specific problems of learners' OECSs development using technology. Moreover, there is a need for research into which technology is suitable for OECS teaching and learning in EFL contexts.

Assessment Methods

The results of data analysis for the third research question showed four different types of assessment methods (n=4) that were used to evaluate learners' oral English communication skills in different EFL contexts. They are presented in Figure 6. They were university-based speaking assessment (n=2), teacher feedback (n=2), mobile-based feedback (n=1), and a mixed panel assessment method (n=1).

Figure 6
Assessment methods



The findings illustrated that mobile-based feedback monitored learners' oral production and the learners received corrective feedback that enhanced their overall communication performance (Fang, Cassim, Hsu, & Chen, 2018). Another study reported that teacher-based feedback, which involved the "praisecriticism suggestion" supported the development of learners' OECSs performance (Mulualem et al., 2019; Wang et al., 2017, 2018).In some EFL contexts, like in Indonesia, teacher feedback boosted learners' oral skills development (Saefurrohman, 2018). Kernel et al., (2017) reported in their study that mixed panel assessment prepared students for life after university through multidisciplinary thinking. It also provided self-assessment in which learners understood their mistakes for future use. In the context of mainland China, a university-based assessment was useful for eliciting candidate's oral performances. At the same time, it also revealed that oral features, e.g., length of turns, choice of words, hesitation markers, and topic coherence, had the potential to affect the scores of the candidate (Liu & Jia, 2017). Moreover, Gan, Oon, and Davison (2017) reported that instituition-based assessment made a significant change in developing learners' oral skills since learners were familiar with it.

The findings showed that there were few studies found on assessment methods for the evaluation of EFL learners' OECSs performance. These methods were university-based assessment, teacher feedback, mobile-based, the mixed panel assessment methods. This study suggests that more studies are needed on the assessment method to facilitate learners' oral performance.

Limitations and Implications

We reviewed and analyzed empirical studies on the issues surrounding the development of oral skills in EFL contexts e.g., problems of EFL learners, and teaching and assessment methods, published in the reputable online databases of SpringerLink, SCOPUS, Web of Science, and Google Scholar from 2010 to 2019. Based on the analysis, there are four main problems that are responsible for EFL learners' poor oral English communication skills, namely environmental, psychological, linguistic, and teaching resource issues. Although this study found four main problems, there are other problems such as the role of learners' L1, motivation, nervousness, stress, intonation, oral activities, lack of authentic materials, and the curriculum that are also responsible for EFL learners' oral skills development in EFL contexts but were less studied. This study has further found that the previous studies dealt with the lack of teachers' efficacy but no evidence was found regarding learners' efficacy as a barrier for EFL learners' OECSs development. This SLR also found that there is a lack of evidence on how these problems are solved for developing learners' OECSs in EFL contexts.

This study also revealed that the integration of technology into teaching methods had a significant impact on developing EFL learners' oral skills in EFL contexts. But there is still a lack of studies on the theoretical perspective of using technology for teaching oral skills in EFL contexts. Moreover, there is still a debate over which technology is more effective for teaching oral skills in EFL contexts. Finally, this study reported that there are few studies conducted on assessment methods for evaluating EFL learners' oral skills. Thus, this study suggests that it is time to think about how technology can be used effectively to develop EFL learners' OECS in EFL contexts since the world is now experiencing the Fourth Industrial Revolution.

Therefore, the findings of this analysis provide a critical and current overview of teaching and learning oral skills for practical implications for teachers, policymakers, and researchers. We anticipate that the identified problems and methods could provide insights for academics and practitioners for overcoming the challenges developing learners' OECSs in EFL contexts.Moreover, the findings of this study suggest that more empirical studies are needed to find ways to deal with these problems developing learners' OECSs in EFL contexts. For developing learners' OECS skills, there is a need to innovate oral skills teaching and assessment methods. Thus, these findings could extend the body of knowledge on language teaching, providing an overview of accumulating issues like problems and methods for developing EFL learners' oral English communication skills in EFL contexts.

Conclusion

Teaching and learning oral English communication skills in an EFL context is a problematic phenomenon for both teachers and learners. It would be an overstatement to say that EFL learners' OECS performance at the tertiary level in the EFL context is up to the mark after their graduation. This systematic literature review accumulated the factors responsible for EFL learners' poor OECSs in different EFL contexts. It also shined a light on the existing scenario for teaching and assessing methods of OECSs in EFL contexts employed in studies from 2010 to 2019. However, the past research on the factors leading to learners' poor performance and the methodfs for the teaching and learning of oral English communication skills in EFL contexts are scattered and haphazard.

The findings showed that the environmental factor was the primary factor, along with other factors impacting EFL learners' poor OECS performance in EFL contexts. The environmental factors entailed learners' passiveness, demotivation, mixed abilities, and large class sizes. The psychological factors entailed learners' anxiety, shyness, self-efficacy, confidence, emotions, and reluctance.Again, linguistics factors were related to learners' lack of language knowledge, accuracy, fluency, and low vocabulary levels, whereas a lack of teaching resources along with the lack of authentic materials for teaching and learning oral skills were also an important finding of this study for future researchers. Previous studies showed that these factors were explicitly or implicitly responsible for EFL learners' poor oral performance.

The findings further illustrated five different types of teaching methods, e.g., technology-based, task-based, communicative, interactive, and microteaching, employed for developing learners' oral skills in different EFL contexts. For the method of teaching learners' OECSs, the use of technology is rapidly increasing in different

EFL contexts. Hence, the pedagogical establishment of using technology for teaching and learning oral skills should be focused on. Furthermore, this study found four different methods of assessment of oral skills: university-based speaking assessment, teacher feedback, mobile-based feedback, and mixed panel assessment.

For developing EFL learners' OECS performance, more empirical studies are needed to address these factors for solutions and innovative teaching and assessment methods. Thus, this study provides insights based on the discussion of the selected studies for enhancing learners' oral English communication skills. As OECSs are important for EFL learners, this study provides a potential foundation for researchers to conduct and implement more research for dealing with the factors responsible for EFL learners' poor OECS performance. Nevertheless, the findings suggest that these factors and methods of teaching and assessment are interrelated for teaching and learning oral communication skills in EFL contexts. Further studies are needed for integrating and innovating teaching and assessment methods with the assistance of technology for dealing with the barriers affecting the development of EFL learners' oral communication skills in EFL contexts.

References

Abid. (2018). Indonesian pre-service english teachers' perceived challenges in improving English oral communication skills. *Journal of English Education and Linguistics Studies*, *5*(2), 147–170. https://doi.org/https://doi.org/10.30762/jeels.v5i2.716

Aeni, N., Jabu, B., Rahman, M. A., & Strid, J. E. (2017). English oral communication apprehension in students of Indonesian maritime. *International Journal of English Linguistics*, *7*(4), 158-165. https://doi.org/10.5539/ijel.v7n4p158

Afshar, H. S., & Asakereh, A. (2016). Speaking skills problems encountered by Iranian EFL freshmen and seniors from their own and their English instructors' perspectives. *Electronic Journal of Foreign Language Teaching*, *13*(1), 112–130. https://e-flt.nus.edu.sg/v13n12016/afshar.pdf

Alam, Q., & Uddin, A. B. (2013). Improving English oral-communication skills of Pakistani public schools students. *International Journal of English Language Teaching*, *1*(2), 17–36.

Alghail, A. A. A., & Mahfoodh, O. H. A. (2019). Academic oral communication difficulties encountered by yemeni postgraduate students in a Malaysian university. *Issues in Educational Research*, *29*(2), 301–325.

Bruner, D. A., Sinwongsuwat, K., Bojanić, B., & Radić-Bojanić, B. (2015). EFL oral communication teaching practices: A close look at university teachers and A2 students' perspectives in Thailand and a critical eye from Serbia. *English Language Teaching*, *8*(1), 11–20. https://doi.org/10.5539/elt.v8n1p11

Bunjan, A., & Suppasetseree, S. (2017). The video blog based role play lessons to enhance English oral communication skills for tourism students. *International Journal of Educational Administration*, *9*(1), 1–16.

Chen, Z., & Goh, C. (2011). Teaching oral English in higher education: Challenges to EFL teachers. *Teaching in Higher Education*, *16*(3), 333–345. https://doi.org/10.1080/13562517.2010.546527

Couvaneiro, S. R., & Pedro, N. (2015). Collaborative learning using tablets in EFL to develop oral communication. In *ACM International Conference Proceeding Series* (pp. 139–143). ACM. https://doi.org/10.1145/2808580.2808602

Darmi, R., & Albion, Peter R. (2017). Enhancing oral communication skills using mobile phones among undergraduate English language learners in Malaysia. In J. H. Murphy A., Farley H., Dyson L. (Eds.), *Mobile learning in higher education in the Asia- Pacific Region education* (Vol. 40, I pp. 359–375). Springer Singapore. https://doi.org/10.1007/978-981-10-4944-6

Ella, R. B., & Dapudong, R. C. (2014). Oral communication achievement of the Bachelor of Arts students major in English at Rajamangala University of Technology, Lanna, Thailand. *International Journal of Academic Research in Progressive Education and Development*, *3*(4), 209–224. https://doi.org/10.6007/ijarped/v3-i4/1252

Fang, W.-C., Cassim, F. A. K., Hsu, C.-N., & Chen, N.-S. (2018). Effects of reciprocal peer feedback on EFL learners' communication strategy use and oral communication performance. *Smart Learning Environments*, *5*(1). https://doi.org/10.1186/s40561-018-0061-2

Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods*, *5*(1), 80–92. https://doi.org/10.1177/160940690600500107

Fitria, R., & Salwa. (2018). Students ' perceptions of English oral communication usage during practice learning program. *Inovish Journal*, *3*(2), 155–162.

Gan, Z., Oon, E. P. T., & Davison, C. (2017). ESL students' oral performance in English language school-based assessment: Results of an empirical study. *Language Testing in Asia*, *7*(1). https://doi.org/10.1186/

s40468-017-0051-2Hart, T. (2016). Learning how to speak like a "native": Speech and culture in an online communication training program. *Journal of Business and Technical Communication*, *30*(3), 285–321. https://doi.org/10.1177/1050651916636363

Herlo, D., & Kurt, M. (2017). Improve your oral communication skills in English by using informal synchronous chat. Paper Presentae at *The 7th International Social & Behavioural Sciences EInternational Conference.*

Herrera, R. F., & Vielma, J. C. (2018). Impact of microteaching on engineering students ' communication skills. *International Journal OfEngineering Education*, *34*(6), 1768–1777.

Idrus, H. (2016). Enhancing oral presentation skills of ESL students: The use of oral communication strategies. In S. Fun Tang, & L. Logonnathan (Eds.), *Assessment for learning within and beyond the classroom* (pp. 437-446). Springer. https://doi.org/10.1007/978-981-10-0908-2

Idrus, H., Salleh, R., & Muhammad, R. L. T. (2011). Oral communication ability in English: An essential skill for engineering graduates. *Asia Pacific Journal of Educators and Education*, *26*(1), 107–123.

Idrus, H., Salleh, R., Wan Hassan, W. N., & Ali, R. M. M. (2013). Self-efficacy in English language oral communication skills of technical trainees. In *The Proceedings of 2013 IEEE International Conference on Teaching, Assessment and Learning for Engineering* (pp. 800–804). https://doi.org/10.1109/TALE.2013.6654548

Ismail, K., Tuspekova, A., & Mustaffa, R. (2018). English oral communication in public secondary schools in Kazakhstan: Understanding its practice and challenges. *3L: Language, Linguistics, Literature*, *24*(2), 122–126. https://doi.org/10.17576/3L-2018-2402-09

Jessica, M. M., Kemboi, G. C., & Andiema, N. C. (2015). A study on interactive teaching methods in developing oral communicative competence in learners of English language in Trans Nzoia West, District of Kenya. *International Invention Journal of Arts and Social Sciences*, *2*(1), 5–13.

Jim, C. Y. H. (2018). Contexts, problems and solutions in international communication: Insights for teaching english as a Lingua Franca. *Journal of Asia TEFL*, *15*(2), 257–275. https://doi.org/10.18823/asiatefl.2018.15.2.1.257

Kasbi, S. and, & Shirvan, E. M. (2017). Ecological understanding of foreign language speaking anxiety: Emerging patterns and dynamic systems. *Asian-Pacific Journal of Second and Foreign Language Education*, *2*(1). https://doi.org/10.1186/s40862-017-0026-y

Kernec, JL. Levrai , P., & Bolster, A. (2017). Bringing the outside world in: Using mixed panel assessment of oral presentations with Electrical and Electronic Engineering students. *Proceedings of 2016 IEEE International Conference on Teaching, Assessment and Learning for Engineering* (pp. 1–8). IEEE. https://doi.org/10.1109/TALE.2016.7851761

Kunioshi, N., Noguchi, J., Hayashi, H., & Tojo, K. (2012). An online support site for preparation of oral presentations in science and engineering. *European Journal of Engineering Education*, *37*(6), 600–608. https://doi.org/10.1080/03043797.2012.733681

Lana, H., Shehata, A., & Omer, M. A. E. (2018). Effective techniques used by teachers to improve English oral communication skills in sudanese secondary schools. *International Journal of Information Research and Review*, *05*(II), 5899–5909.

Liu, L., & Jia, G. (2017). Looking beyond scores: Validating a CEFR-based university speaking assessment in Mainland China. *Language Testing in Asia*, *7*(1), Article 2. https://doi.org/10.1186/s40468-017-0034-3

MacWhinnie, S. G. B., & Mitchell, C. (2017). English classroom reforms in Japan: A study of Japanese university EFL student anxiety and motivation. *Asian-Pacific Journal of Second and Foreign Language Education*, *2*(1), Article 7. https://doi.org/10.1186/s40862-017-0030-2

Malthus, C., & Lu, H. (2012). "Not a Mission Impossible": The perceptions of successful graduates on the development of spoken English for the workplace. *Journal of Asian Pacific Communication*, *22*(1), 120–139. https://doi.org/10.1075/japc.22.1.07mal

Mamo, B., & Yigzaw, A. (2015). Fidelity of peer mediation and its role in improving students' oral English communication skills. *Science, Technology and Arts Research Journal*, *4*(1), 215-217. https://doi.org/10.1016/S0002-9394(26)91015-5

Manzano, B. A. (2018). Examining the oral communication strategies used by a group of nepalese adult learners in an ESL context. *3L: Language, Linguistics, Literature*, *24*(1), 84–96. https://doi.org/10.17576/3L-2018-2401-07

Mitchell, R., & Eng, T. L. (2010). Assessment of students' learning experience in an oral communication course at MIT for EECS majors. In *2010 IEEE Frontiers in Education Conference (FIE)*, F1F-1-F1F-5. https://doi.org/10.1109/FIE.2010.5673341

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Mulualem, Muchie, M., & Meles, Y. (2019). Assessing teachers' practices and challenges of teaching students' oral communication skills. *International Journal of Research in Social Sciences and Humanities*, *9*(II), 237–246.

Nasihah, M. (2019). The use of role play technique to improve students' english oral communication. *Journal of English for Academic and Specific Purposes*, *2*(1), 10–17. https://doi.org/https://doi.org/10.18860/jeasp.v2i1.7259

Nikolic, S., Stirling, D., & Ros, M. (2018). Formative assessment to develop oral communication competency using YouTube: Self- and peer assessment in engineering. *European Journal of Engineering Education*, *43*(4), 538–551. https://doi.org/10.1080/03043797.2017.1298569

Okoro, N. (2007). Strategiesfor Remedying poor communication skills of stupents. *An Interdisciplinary Journal of Communication Studies*, *6*, 13–22.

Onoda S. (2014) An exploration of effective teaching approaches for enhancing the oral fluency of EFL students. In T. Muller, J. Adamson, P.S. Brown, & S. Herder (Eds.), *Exploring EFL fluency in Asia*. Palgrave Macmillan London. https://doi.org/10.1057/9781137449405_8

Rahman, M. M. (2010). Teaching oral communication skills : A task-based approach. *ESP World*, *9*(1), 1–11. https://doi.org/10.1080/10572259809364631

Saefurrohman, S. (2018). EFL *Teachers Assessment Methods in Oral Communications*. *Advances in Social Science, Education and Humanities Research*, *267*, 268–272. https://doi.org/10.2991/aecon-18.2018.52

Şimon, S. (2014). Enhancing the English oral communication skills of the 1 st year students of the Bachelor's degree program "Communication and public relations." *Procedia - Social and Behavioral Sciences*, *116*, 2481–2484. https://doi.org/10.1016/j.sbspro.2014.01.596

Soomro, M. A., Siming, I. A., Channa, M. A., Shah, S. H. R., Naeem, N., & Abbasi, A. M. (2019). An analysis of English oral communication apprehension and anxiety of engineering undergraduates in Pakistan. *International Journal of English Linguistics*, *9*(2), 162-173. https://doi.org/10.5539/ijel.v9n2p162

Subramaniam, Jeevi K., Gunadevi A., Pathma Fathima Harun, R. S. N. R. (2013). Polytechnic students ' perceptions of youtube usage in the english oral communication classroom. *International Journal of Asian Social Science*, *3*(9), 1962–1966.

Subramaniam, G. K. J., & Harun, R. N. S. R. (2012). Marketing Students' Perceptions and Their Experiences during Industrial Training on English Oral Communication Skills. *Procedia - Social and Behavioral Sciences*, *66*, 283–289. https://doi.org/10.1016/j.sbspro.2012.11.270

Thu, N. T., Thi, N., & Thu, K. (2016). Oral English communication strategies among Vietnamese non-majors of English at intermediate level. *American Journal of Educational Research*, *4*(3), 283–287. https://doi.org/10.12691/education-4-3-9

Tipmontree, S., & Tasanameelarp, A. (2018). The effects of role-playing simulation activities on the improvement of EFL students' business English oral communication. *Journal of Asia TEFL*, *15*(3), 735–749. https://doi.org/10.18823/asiatefl.2018.15.3.11.735

Tsai, I. I. yin. (2019). The effect of peer collaboration-based learning on enhancing English oral communication proficiency in MICE. *Journal of Hospitality, Leisure, Sport and Tourism Education*, *24*, 38–49. https://doi.org/10.1016/j.jhlste.2018.10.006

Tuomaitė, V., & Zajankauskaitė, Z. (2017). Oral communication in a foreign language competence development in academic contexts. *Studies About Languages*, (31), 113–129. https://doi.org/10.5755/j01.sal.0.31.19049

Wang, B., Teo, T., & Yu, S. (2017). Teacher feedback to student oral presentations in EFL classrooms: a case study. *Journal of Education for Teaching*, *43*(2), 262–264. https://doi.org/10.1080/02607476.2016.1257507

Wang, B., Yu, S., & Teo, T. (2018). Experienced EFL teachers' beliefs about feedback on student oral presentations. *Asian-Pacific Journal of Second and Foreign Language Education*, *3*(1), Article 12. https://doi.org/10.1186/s40862-018-0053-3

Zukurnain, N., & Kaur, S. (2014). Oral English communication difficulties and coping strategies of diploma of hotel management students at UiTM. *3L: The Southeast Asian Journal of English Language Studies*, *20*(3), 93–112. http://ejournals.ukm.my/3l/article/view/6078/2971

National Research University Higher School of Economics
Journal of Language & Education Volume 7, Issue 1, 2021

Alruwaili, A. (2021). Corpus linguistics for education-A guide for research by Pascual Perez-Paredes: A Book Review. *Journal of Language and Education,* 7(1), 241-244. https://doi.org/10.17323/jle.2021.11951

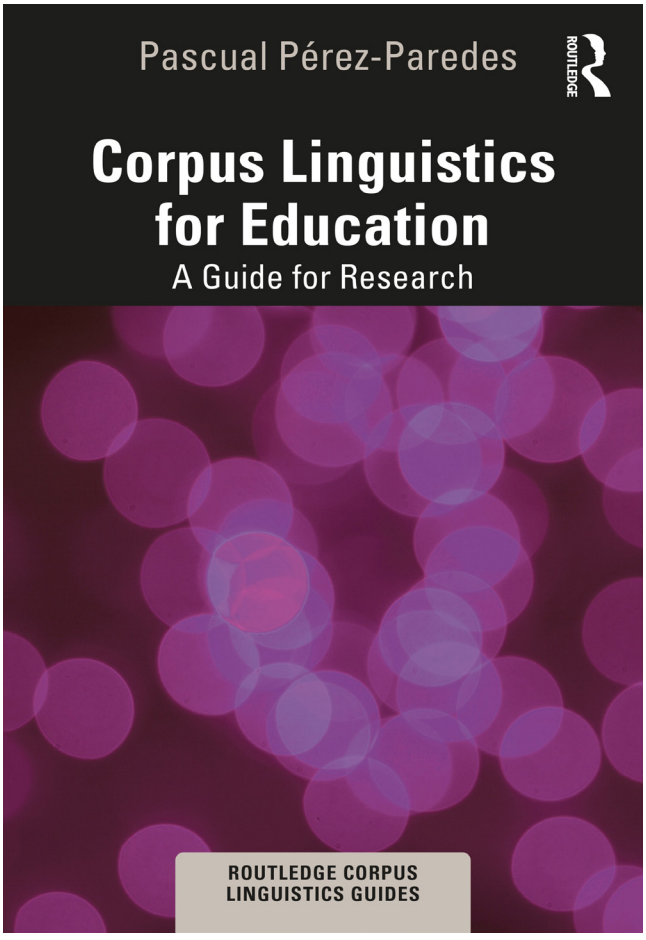
Corpus Linguistics for Education- A Guide for Research by Pascual Perez-Paredes: A Book Review

Awatif Alruwaili
Northern Border University

Correspondence concerning this article should be addressed to Awatif Alruwaili, Northern Border University, 9280, Arar Saudi Arabia, Saudi Arabia. E-mail: awatif.k.alruwaili@nbu.edu.sa

Pascual Perez-Paredes, P. (2020), *Corpus Linguistics for Education: A Guide for Research (Routledge Corpus Linguistics Guides)*. Taylor & Francis Ltd.

The book *Corpus linguistics for education* was written by Pascual Perez-Paredes (2020) and published by Routledge. The book aims to provide a guide for how corpus linguistics (CL) methods can be used in educational research. The book consists of seven chapters and an additional conclusion chapter. The book addresses different themes that are relevant for the inclusion of CL methods in the education research field such as frequency register (texts) and keyword analysis, among others. The book is distinguished by the inclusion of a good number of tables and figures providing step-by-step guides for all the selected methods of analysis. This book is very important for researchers and students who are interested in using CL methods in the field of education. This review consists of a brief summary of the eight chapters and a critical discussion of three key issues raised in the book. The review also provides an overall evaluation of the contributions this book has made to this particular discipline.



The first chapter is a thorough introduction to CL and how it can be used in educational research. The author discusses the concept of frequency and gives a number of examples of the use of frequency in learning an L1 and its use in public discourse and texts. This chapter presents two approaches for using CL in education research: the corpus is the primary data and CL is the research methodology, while the second approach uses a variety of methods to collect data that will be analysed by CL methods, so the corpus is considered secondary data. The second approach is the most convenient for use in education research.

The second chapter presents different approaches to text analysis. The author discusses each approach and provides examples of educational research studies using the approach. Content analysis uses a qualitative description approach, which is the 'classical example' (p.21) used in educational research, although it has the lowest level of interpretation. The author describes it thoroughly. In addition, examples of analysis tools such as NVIVO and MAXQDA are provided. The second approach is popular in educational research and has a high level of interpretation. The theme analysis approach is preferably used with a small dataset as it is time consuming because the data will be approached with no presumption (grounded theory). The third approach is conversational analysis, which is used

with spoken discourse. The author mentions the most famous uses of this approach (classroom interaction) as well as its disadvantages, such as being time consuming, a daunting task for analysing every single utterance, full of pauses, etc. The discourse analysis (DA) approach is used for text analysis, a focus different to that of the previous approaches. DA focuses on analysing ideas, stance, power, etc. The author links DA to educational research by stating that both are 'socially committed paradigms that address problems through a range of theoretical perspectives' (p. 25). The last approach is CL analysis of register (texts), which investigates texts from linguistics angles that concern the question of *how*. In a few lines, the author carefully reviews the two directions that respectively see CL as a science (branch of study) and as a mere methodology. The author then explains the types of methods of CL (qualitative by analysing concordance lines and quantitative by analysing collocations and keywords). This chapter attempts to link each approach to an educational research paradigm: grounded theory is a phenomological qualitative approach, conversational analysis is based on ethnomethodology, discourse analysis is linked to sociocultural theory, and CL methods are linked to different directions taking into account different schools of CL: 1) CL studies texts situated in a positivist research paradigm (Sinclair, 1991) and 2) situated in an objectivism paradigm (Hunston, 2002).

Chapter 3 provides a practical discussion of CL methods and research on language use such as frequency and reading concordance lines. The importance of frequency is discussed, as well as how it helps in discovering language patterns. CL research methods can be used by anyone interested in how language is used in texts. There are many aspects of examining words using CL methods, which expand the areas of inquiry in education research in particular. However, education research may focus on the language as an activity instead of the language as a genre, as linguists may prefer. The author provides two case studies where CL is used as a data analysis method to discover hidden patterns of language use: the first aims to compare the qualitative method (theme analysis) to the CL method (frequency, reading concordance lines). This case study (Fest, 2015) shows how CL methods can complement thematic analysis and go beyond the meanings of words (themes). The second case study uses content analysis and CL methods to analyse language polices of universities in Spain, in which the CL methods can add validity to content analysis findings through statistical testing. The author aims to demonstrate the strength of combing these two methods by selecting these case studies. The author introduces concordance lines by providing a simple definition then moves to the software that is used for reading concordance lines and/or analysing concordance lines (AntConc, Sketch engine). Further, the author introduces practical ways for examining concordance lines while bearing in mind the surrounding context and thinking beyond the word unit. The book introduces the Sinclair model (2003) for reading concordance lines, providing detailed explanation and a step-by-step guide for searching a corpus by using two small corpora.

Chapter 4 introduces CL methods for education research as an alternative for qualitative methods such as grounded theory. The chapter starts by presenting the principles for designing a corpus and how educational research can use the CL methods, either primary or secondary, as a part of the larger methodology of the research. Two case studies were discussed as examples for the use of CL methods as the main research methodology and secondary data as a part of the larger methodology. These two cases investigating education polices may be of interest to the readers of this book. In case study one, the author explains how to design a corpus according to research questions and the steps that needed to be considered before compiling the corpus. Further, it goes in detail through the process for creating and cleaning the corpus. As well as running the analysis (quantitatively and qualitatively), the author helps early-career researchers (postgrads) by showing them an example of breaking research questions into arguments that are linked to corpus building. Case study two uses embedded CL research methods to triangulate the results. The CL method's 'collocation analysis' is used to support the main research methods.

Chapter 5 discusses the interview data within the framework of CL methods and focuses on the process of the interview transcription. This chapter describes interview data and the data that happen during the interview itself. Transcription types are discussed along with the challenges and issues that might come up during the transcriptions. The chapter goes through many details of the transcription process that are rarely found in literature. In addition, an interview transcription can be prepared to be used as corpus data 'building corpus from interview transcription'. The author explains in depth how transcription can be done in a systematic way, reviewing quite a good number of software applications that could help in this process. Moreover, guidelines for orthographic transcription are explained and annotation is covered. The book emphasises the importance of including metadata in the corpus (interview transcription). Text Encoding Initiative (TEI) guidelines are discussed with examples of the metadata included. The book generally is quite good in giving the readers

options in all the discussed methods. For example, it gives the reader two options for annotating a corpus. One is annotation tags in the transcriptions itself, as a quick and easy way for cases where there is limited data. The second option is annotation taxonomy such as TEI, which is more time-consuming but provides a more sophisticated search, especially if the data set is very large. For each option, an example and detailed explanation is presented.

Chapter 6 presents CL methods to investigate the lexis in a corpus (textual data). The first method is keywords, which help to reveal embedded discourse among many other aspects of ideology. The book introduces keywords from a CL research methodology perspective. The author explains the methodological aspects of keywords, multi-keyword analysis, as well as software options used for running such analysis. Step by step, the author takes the reader through building a corpus, 'peace treaties worldwide', running the analysis quantitively, explaining different types of statistically significant tests such as log likehood, chi-square, and using reference corpora. The analysis is then continued qualitatively by examining the individual contexts in which keywords occur. Figures and tables are used for further clarification, allowing the reader to follow easily. After identifying the keywords, colligational analysis, another important aspect of CL methods research, is represented, which is an examination of grammatical relationships to discover 'the units of meaning' (p. 133). Noun and noun phrases are used for examining grammatical relationships. Sketch Engine is the recommended software for analysis but it is not the only one mentioned. The author explains another way to examine multi-words, which is n-grams, sequences of groups of words that occur frequently in a corpus. The methodological aspects of n-grams are discussed with examples and software that can extract n-grams from a corpus.

Chapter 7 illustrates a complex corpus search for educational research. It targets analysing talks, a way to integrate CL methods into education research. The author starts by reviewing theoretical aspects of spoken English language analysis from two perspectives: linguistics and educational research. The author explains how to perform a complex search in a very detailed manner with tables and figures. Firstly, a keyword analysis is run, POS tags searched, concordance lines established, and an interpretation drawn. Secondly, lemmas are searched for a specific word and the significant collocation of the searched words is identified. Thirdly, searches for lemmas and symbols are used to customise the findings.

Chapter 8 sums up the book with a reminder of its main aim, which is bridging the gap between CL and educational research. The author acknowledges the challenges that might be encountered in bridging the gap, including the difficulty of collaboration, interdisciplinary work, and the methodological differences between CL and research education. The former is based on positivism and the latter is interpretive paradigms, which have rarely been used to complement each other in published literature.

A Critical Discussion of Three Key Issues Raised in the Book

In this section, I will discuss three issues that the author raised in the book: the collaboration difficulty, education paradigms, and skills presented. The author points out the difficulty of collaboration between CL and education research disciplines because of the differences that might discourage researchers. The main difference is seen as CL being mainly quantitative while education research is mostly qualitative, possibly giving the impression that CL is of no use in educational research. Interdisciplinary work is acknowledged to be difficult in general by the author, and particularly rare between these two disciplines, as the CL discipline is relatively new.

Paradigms in educational research are discussed in one section in the first chapter. Although the explanation is clear and concise, the book could have benefited from the author discussing these paradigms to a greater extent. The author attempts to explain the role of CL methods in education research within the same section but it is not clear enough, and instead of explaining the role, he proceeds to justify why CL methods are not used widely in education research. However, to be fair, the author deals with CL not a science but as a collection of research methods throughout the book (i.e., mere instruments in educational research). Nevertheless, CL paradigms are mentioned in an indirect way in some sections. For example, the author mentions the linguistics-driven paradigm, which is situated in a positivist research paradigm. This paradigm is established by Sinclair, a pioneering corpus linguist (1991). It is worth mentioning that the linguistics-driven paradigm has an approach that is contrary to that of the corpus-based one. This distinction was first defined by Tognini-Bonelli (2001), who defined the linguistics-driven approach as a science and the corpus-based approach as a mere method.

The author covers the most relevant skills that might be needed when using corpus research methods in educational research as well as in the CL field. The book presents the most needed skills for using CL methods as a linguist or as an educational researcher. Eighteen skills were covered and included in the book to discuss the methods. Skill 1, 2, and 7 deal with frequency, its interpretation, and how to handle frequency. Skills 3 and 4 discuss register and textual data. Skill 5 is a practical guide for using an existing corpus. Skill 6 summarises the process of reading concordance lines. Skill 8 helps the reader to understand collocations within the CL perspective. Skill 9 explains corpus design features. Skill 10 explains what is involved in comparing two corpora, while skill 11 explains all the descriptive and inferential statistics mentioned in the book. Skill 12 provides the readers with hints for transcribing data. Skill 13 discusses how to use one's own taxonomy to annotate one's data. Skills 14, 15, and 16 are those related to keyword analysis, researching noun and noun phrases, and using n-grams. Skill 17 explains how to conduct a complex search. Skill 18 is a critical view of CL methods, which is significant for researchers.

A Final Evaluation of the Overall Contribution

This book is very valuable because of its contribution to educational research. CL methods are rarely discussed within the education research framework but this book is devoted to doing so. Each chapter starts without the assumption of prior knowledge by the readers. The book covers most of the important principles in CL such as accountability and how it has a great impact on the findings. The author discusses a very important issue to novice corpus-users, which is the size of the corpus and how it has an impact on the analysis. However, what is more important than the size is the representativeness. Principles of compilation and normalisation of frequency are key for having comparable results, as well as the significant statistical measurement of collocation, which is more valid than native intuitions. The tables and figures facilitate understanding of the CL methods as well as serving as guides to familiarise the reader with applying such methods (analysis). The author is concerned with when to use the CL methods in triangulation, which is critical in interdisciplinary work, so the approach used in this book is "sequential or cyclical" (p. 169). The sequential CL approach is used when CL methods are used as the main component of the data collection method, while the cyclical approach is used when CL methods are used as the starting point for the analysis, then a different method from education research is used, followed by CL methods again in a cyclical process.

References

Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge University Press.
Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
Tognini-Bonelli, E. (2001). *Corpus Linguistics at work*. John Benjamins Publishing.

Contents

Editorial

Elena Tikhonova, Lilia Raitskaya
Education for Sustainable Development: Glocal Implications for Universities 4

Research Articles

Minoo Alemi, Hessameddin Ghanbar, Atefeh Rezanejad
The Integration of Verbal Humor into EFL Classrooms: The Issues of Appropriateness and Relevance in Focus 14

Hesham Suleiman Alyousef
The Status of Theme in Research Article Abstracts in Seven Dentistry Subdisciplines: A Text-Based Study of Intradisciplinary Variations and Similarities in Thematic Choices and Thematic Progression Patterns..... 28

Aslihan Bagci, Belgin Aydin
Autonomy Orientations of English Language Teachers in Turkey..... 46

Tanju Deveci, Nader Ayish
Qualitative Adjectives in EFL Students' Reflective Writing Essays..... 64

Zoe Gavriilidou, Lydia Mitits
The Socio-linguistic Profiles, Identities, and Educational Needs of Greek Heritage Language Speakers in Chicago..... 78

Tahereh Heydarnejad, Azar Hosseini Fatemi, Behzad Ghonsooly
The Relationship between Critical Thinking, Self-regulation, and Teaching Style Preferences among EFL Teachers: A Path Analysis Approach 96

Ali Isik
Exploring How ELT Teachers Perceive and Practice English Language Assessment.....109

Mohsine Jebbour
Self-Disclosure and Moroccan EFL Learners' Writing Development: Effects on Complexity, Accuracy, and Fluency.....127

Maya Lambovska, Daniela Todorova
'Publish and Flourish' instead of 'Publish or Perish': A Motivation Model for Top-quality Publications....141

Tatiana Pospelova
The Collaborative Discussion Model: Developing Writing Skills through Prewriting Discussion.....156

Marta Segura Molla, Helena Roquet, Carmen Pérez-Vidl
The Effects of a CLIL Programme on Linguistic Progress at Two Different Points in Time.....171

Frankie Subon, Norseha Unin
The Efficacies of an Explicit Vocabulary Instruction Model on ESL Learners' Vocabulary Size and Writing Skills.....190

Vu Van Tuan
The Impact of Social Networking Sites on Study Habits and Interpersonal Relationships among Vietnamese Students 206

Opinion Article

Irfan Ahmed Rind
Becoming an ESL Researcher: A Personal Monologue.....219

Reviews

Prodhan Mahbub Ibna Seraj
A Systematic Overview of Issues for Developing EFL Learners' Oral English Communication Skills 229

Awatif Alruwaili
Corpus linguistics for education-A guide for research by Pascual Perez-Paredes: A Book Review 241