DEAL 2022

CHALLENGES AND OPPORTUNITIES IN CONTEMPORARY ENGLISH APPLIED LINGUISTICS

EDITED BY
GYULA TANKÓ & ATTILA M. WIND

AN EDITED VOLUME OF THE DEPARTMENT
OF ENGLISH APPLIED LINGUISTICS,
SCHOOL OF ENGLISH AND AMERICAN STUDIES,
EÖTVÖS LORÁND UNIVERSITY, BUDAPEST

DEAL 2022:
Challenges and Opportunities in Contemporary English Applied Linguistics

Edited by
Gyula Tankó & Attila M. Wind

DEAL 2022: Challenges and Opportunities in Contemporary English Applied Linguistics

An edited volume of the Department of English Applied Linguistics,
School of English and American Studies,
Eötvös Loránd University, Budapest

Edited by
Gyula Tankó & Attila M. Wind

Budapest, 2022

The publication of the volume was financed by the Textbook Grant Application of Eötvös Loránd University.

Editors:
Gyula Tankó
Attila M. Wind

ISBN 978-963-489-550-3
ISBN 978-963-489-551-0 (pdf)

DOI: 10.21862/ELTE.DEAL.2022

EÖTVÖS UNIVERSITY PRESS

ELTE | FACULTY OF HUMANITIES

Executive Publisher: the Dean of the Faculty of Humanities of Eötvös Loránd University
Editorial Manager: Nóra Csanádi-Egresi
Publishing Editor: Györgyi Majoros
Layout: Gyula Tankó
Cover: Ildikó Kmotrik
Printed by: Multiszolg Ltd.

Table of Contents

Contributors 7

Editorial 11

Chapter I:
Current Issues in Language Education Policy and Assessment

Disambiguating the Concept of Language Education Policy
Enikő Öveges 15

Probing the Reading Paper of the Advanced Level EFL School-Leaving Examination
Gyula Tankó and Zsuzsanna Andréka 23

Chapter II:
Theoretical and Practical Insights for Language Development

Multiple and Single-Source Text Integrated Writing: A Comparative Study of Task Characteristics and Composition Processes
Gyula Tankó 59

The Development of Syntactic Complexity and Fluency in an Advanced Writing Course
Attila M. Wind 97

Students' Opinion on the Usefulness of Netflix, Instagram, and TikTok in EFL Learning—A Pilot Study
Boglárka Spissich 119

Chapter III:
Expanding the Understanding of Individual Differences

Individual Differences Revisited
Éva Illés 141

Teacher Trainees' Self-Efficacy Beliefs in Light of Their Perceived Language Aptitude and Explicit—Implicit Language Learning Behavior
Anna Zólyomi 163

Flow Experiences During Speaking Tasks in the Hungarian EFL High School Classroom
Katalin Piniel and Fanni Ritecz 197

English Majors' Flow Experiences in L2 Writing: An Interview Study
Duha Alsayed-Ahmad and Ágnes Albert 227

The Charismatic Teacher: An Interview Study on the Motivating Agency of Charismatic Language Teachers in Hungarian Higher Education
Szabolcs Peták and Csaba Kálmán 259

Contributors

Ágnes Albert (ORCID ID 0000-0001-8339-7119) holds a PhD in Language Pedagogy. Her research interests include task-based language learning and individual differences in foreign language learning, in particular learner creativity and positive emotions associated with the learning of foreign languages. Recently, she was involved in a research project investigating secondary school learners' motivation, autonomy, and positive and negative emotions in connection with their English classes. She is currently a member of the MTA-ELTE Foreign Language Teaching Research Group.

Zsuzsanna Andréka (ORCID ID 0000-0001-6357-8895) is a PhD student in the Language Pedagogy and English Applied Linguistics PhD program at Eötvös Loránd University, Hungary. Her main research area is language testing. She is currently teaching academic skills courses at the university, focusing on improving students' reading and writing skills. She is also a PhD student researcher member in an assessment project funded by the Hungarian National Research, Development and Innovation Office.

Duha Alsayed-Ahmad (ORCID ID 0000-0002-3492-4637) holds an MA in English Studies and is a PhD student in English Applied Linguistics and Language Pedagogy. Her research interests include translation studies and positive psychology, in particular language learners' experiences of flow. She was involved in a research project studying the flow experiences of EFL learners in writing tasks from different genres. She taught a Skills Development course at ELTE, and she is currently a member of the MTA-ELTE Foreign Language Teaching Research Group.

Éva Illés (PhD, ORCID ID 0000-0001-5998-3716) is an associate professor at the Department of English Applied Linguistics. She has a wide range of experience, including teaching in adult, secondary, and higher education in Hungary and Britain. Her current research areas are pragmatics, English as a lingua franca, translation in language teaching, and teacher education. She is currently working on projects which aim to create a

mutually beneficial, close connection between research, be it theoretical or empirical, and language teachers' classroom practice.

Csaba Kálmán, PhD, works at the Department of English Applied Linguistics at Eötvös University, Budapest. His research interests include foreign language motivation, adult education, corporate language education, and self-regulation in both educational and corporate contexts. He is a member of the MTA-ELTE foreign language teaching research group. In addition, he has been running his own business offering a variety of corporate language training and skills courses for 29 years.

Enikő Öveges was an associate professor at the Department of English Applied Linguistics at Eötvös Loránd University, Budapest. Her major interest was language policy, in which she earned her PhD and habilitation degrees. She specialized in foreign language teaching frameworks and curricula in the Hungarian public education sector. She was involved in and led various national and international projects, participated in large-scale empirical surveys, and authored several publications. She was the head of the Department of English Applied Linguistics.

Szabolcs Peták graduated from Eötvös University as a History and English Language Teacher in 2022. Shortly after his graduation, he started working as a Creative Copywriter for a Hungarian advertising agency. He is still loosely associated with education, as in his free time, he also works as a private teacher, and he is still interested in research, as it is an important part of his job.

Katalin Piniel (PhD, ORCID ID 0000-0001-9225-3301) is an assistant professor at the Department of English Applied Linguistics at Eötvös University, Budapest, where she obtained her PhD in Language Pedagogy. She is interested in the dynamic interrelationship of individual differences in language learning, with a special focus on emotions. She is currently involved in a project supported by the National Research, Development, and Innovation Office in Hungary, exploring high school

students' emotions, motivation, and autonomy in their English classes both at macro and micro levels.

Fanni Ritecz is a recent graduate from Eötvös Loránd University, Budapest, where she obtained her MA in the 6-year Double Major Teacher Training Program. She is a teacher of English and German as Foreign Languages and has just started as a Technical Writer at Semcon. She is particularly interested in teaching and helping children find the beauty in language learning and culture because she believes that the ability to speak other languages broadens one's world, and thus presents amazing opportunities.

Boglárka Spissich (ORCID ID 0000-0002-8110-9133) is a PhD student at Eötvös Loránd University in the Language Pedagogy and English Applied Linguistics program. She is currently teaching academic skills and advanced writing courses for BA students at the university. Her main research interest is EFL learning and teaching with the help of films and social media, focusing on the use of well-known applications.

Gyula Tankó (PhD, ORCID ID 0000-0002-0721-8171) is an associate professor at the Department of English Applied Linguistics at Eötvös Loránd University, Budapest, and has taught in upper-primary, secondary, and higher education institutions in Italy, Germany, the U.K., and Romania. He teaches and researches Language Assessment, English for Academic Purposes, and Mediation Studies. He has published books and articles on language assessment and academic writing and is the principal researcher in a large-scale assessment project funded by the Hungarian National Research, Development and Innovation Office (Hungarian Scientific Research Fund, OTKA; Project number: 142536).

Attila M. Wind (PhD, ORCID ID 0000-0002-2702-5211) is an assistant lecturer at the Department of English Applied Linguistics at Eötvös Loránd University. He completed his PhD at Lancaster University (UK) in 2019. His research interests are in the area of second language development, more broadly, with a special focus on second language

writing development. He is also interested in the application of Complex Dynamic Systems Theory to the study of second language development.

Anna Zólyomi (ORCID ID 0000-0002-9280-5775) is an applied linguistics and language pedagogy PhD student. She is teaching at Eötvös Loránd University, Hungary. She is taking part in state-funded research projects involving the New National Excellence Program (ÚNKP), a research grant which she has won in four consecutive years, and the Hungarian Scientific Research Fund (OTKA) by the National Research, Development and Innovation Office. She is also a member of the MTA-ELTE Foreign Language Teaching Research Group. Her main areas of interest include language aptitude and explicit–implicit learning processes.

Editorial

It is with great joy that we share with you the third volume in the applied linguistics series of the Department of English Applied Linguistics at Eötvös Loránd University. Since Enikő Öveges, our late Head of Department, proposed that the we should launch a series to showcase the wide range of topics from the field of applied linguistics that we teach and research, a large number of empirical and theoretical investigations conducted by colleagues, often jointly with post-graduate students, have been made available electronically and in print to a wider readership. In agreement with the original broader scope set for the initiative, we have been engaged for almost a year in a vigorous professional cooperation by means of which we have gained valuable knowledge and further improved our reviewing, editing, and project managing competencies.

We organized this year's contributions into three strands. In the first one, focusing on language education policy and language assessment, Enikő Öveges presents a systematic theoretical overview of the approaches to language education in the national and international literature and discusses the disparate interpretations, terminological overlaps, and diverse cross-disciplinary conceptualizations in the field. Following the review of the numerous facets of the construct, a comprehensive definition is proposed for language education policy. In the second study, Tankó and Andréka explore the reading paper of the Hungarian advanced level English as a foreign language (EFL) school-leaving examination. The investigation was prompted by the insufficient reading ability of first-year English major students and confirmed that the reading paper does not measure as effectively as it should the aspects of language ability that it is intended to assess. The findings provide critical insights into the flaws of the measurement and should help improve the examination.

The second strand features three studies that contribute to topical discussions in the field of Second Language Acquisition. With the shift in preference away from individual to integrated tasks for teaching and assessing English for academic purposes, discourse synthesis has become a central research topic. Tankó's study substantiates the claim that

discourse synthesis needs to be reconceptualized, shows that it can also be elicited by a single-source text integrated writing task, and points out that it is the special-purpose task schema, not the number of texts that triggers the process of intra-textual synthesis, which is an original concept that substantially advances our understanding of synthesis. In the second study, Wind investigates the second language writing development of undergraduates. Using computational tools, he analyzes essays written at the beginning and end of an advanced writing course. The findings show that, unlike the students' syntax, their fluency showed statistically significant changes during the semester, and based on these results Wind formulates practical recommendations for the classroom. In the third article, Spissich investigates the way Hungarian secondary school students relate to films and the highly topical social media tools as means for foreign language learning. She discusses the differential beneficial effects experienced by students across various language skills and means of learning.

The last strand contributes to Individual Differences (IDs) research with one theoretical and four empirical studies. In her theoretical study, Illés offers a thorough critical overview of the main trends and developments in the conceptualization and research of individual differences in second language acquisition. In accordance with developments in other fields of applied linguistics, she argues for a dynamic conceptualization of IDs and embraces the encouraging reversal in the information flow between researchers and classroom practitioners. Zólyomi, in a complex exploratory study on IDs among teacher trainees, sets out to identify the factor structure and reliability of an instrument being developed to measure teachers' self-efficacy beliefs. The findings provide insights into the underlying components and their associations. In their study on individual differences in flow experiences related to speaking tasks, Piniel and Ritecz present a conceptual and partial replication of a research study on task-specific flow in the Hungarian EFL high-school context. Their study corroborates the findings of the initial investigation, namely that a large proportion of learners are likely to experience flow in the classroom while completing language learning tasks. Also focusing on flow experiences, in their interview study Alsayed-Ahmad and Albert investigate English major's flow experiences during

the completion of writing tasks. They report that English university students participating in the study do experience flow and specify the optimal conditions that induce the experience. In the final study, Peták and Kálmán explore how university foreign language teachers and EFL teacher trainees consider charisma as an implicit motivational tool. The interview study reveals that subject knowledge, methodological knowledge, and positive character traits are the most distinctive characteristics of charismatic L2 teachers.

We would like to commend our colleagues for their valuable contributions. We would also like to express our profound gratitude to the reviewers who provided prompt and detailed constructive criticism despite their busy schedules. In closing, we pay tribute to the invaluable proof-reading work of Jamil Toptsi and James Griffin and to Eötvös Loránd University for funding our project (Textbook Grant, BTK/6355).

We hope that you will enjoy reading the articles in this volume just as much as we have.

Gyula Tankó and Attila Wind

1

Disambiguating the Concept of Language Education Policy

Enikő Öveges
oveges.eniko@btk.elte.hu

Abstract

Language education policy is a field that overlaps with language policy, language planning, and language pedagogy. The present article first explores the concept of language education policy in an attempt to distinguish it from language pedagogy based on a review of the available national and international canon of literature, which reveals that the academic discourse on this topic has been fraught with several terminological ambiguities and divergent interpretations of the scope of these two fields. The subsequent examination of the relationship between language education policy and language education reveals that the use of the concepts in research is fuzzy, and some attempts have been made at differentiating them through various disciplinary pursuits. Finally, the difficulties inherent in the definition of language policy as a construct are addressed. They are believed to originate from the interchangeable use of language policy and language planning, as well as from the fact that language policy lacks a standard theoretical framework and is construed inconsistently across disciplines. The article ends with a definition proposed for language education policy.

Keywords: language education policy, language pedagogy, language education, language planning, terminological ambiguities

DOI: 10.21862/ELTE.DEAL.1.152.1.15

Disambiguating the Concept of Language Education Policy

What is Language Education Policy? How, if at All, Does it Differ From Language Pedagogy?

Bárdos (2000) described the latter as a discipline studying language education, and defined it as follows: language pedagogy "investigates the process of language teaching and language learning from the aspects of linguistics, pedagogy, and psychology" (p. 32). On this basis, we could say that these two fields do not overlap; however, in the same volume the author presented several models of language pedagogy. The first of these—reminiscent of an earlier theory of his (Bárdos, 1984)—contains several elements and levels connected to language education planning and implementation that seem to go beyond language pedagogy. An example among the institutions listed can be the ministry in charge of this specific field, while among the problems to be tackled we could mention the range of languages learnt or the question of contact hours, both of which are planning issues—that is, they may fall outside the scope of a definition that focuses on the teaching and learning process. Education policy decisions also appear in Medgyes' (1995) language pedagogy model, but there they can be found at the first level, isolated from subsequent elements that can be directly linked to language pedagogy.

Bárdos' (2000) interpretation of language pedagogy can be most transparently outlined from the figure entitled "Language pedagogy and its sister disciplines" (p. 32), which reveals the levels or "shell structure" of language pedagogy. The figure is in harmony with his statement that language pedagogy "examines the process, content, organizational forms, control, and evaluation of language learning and language teaching" (p. 31), and consequently identifies the following levels: (1) the theory and practice of teaching foreign languages; (2) the mediation of language content; (3) the development of language skills; (4) the planning, implementation, and assessment of teaching materials; (5) the knowledge and use of teaching materials; (6) the assessment and evaluation of foreign

language competencies and test taking strategies; and (7) language acquisition and language learning, bilingualism, and age specific features or errors in language development. Taking these "shells" into account, language pedagogy and language education policy are indeed interrelated in some ways—for example, in terms of how language competencies are measured—but to assert this for sure we would need to know whether the author considered the items listed above in the context of school or higher levels of education. In his 2006 paper, Bárdos reinforced this view with the figure entitled "Language pedagogy within the set of sister and basic sciences" (p. 9), in which the language pedagogy set overlaps with the following sets in ascending order: language policy, language acquisition, applied linguistics, and technical language and testing. Furthermore, continuing Petneki's (2001) deliberation on the category membership of methodology, Bárdos (2006) indicated that methodology is just one component of language pedagogy as a discipline since "it represents only a tiny slice of what language pedagogy actually means" (p. 8). The same approach can be perceived in Einhorn's (2015) monograph, where the developments in the field of language pedagogy are discussed in relation to language teachers, and is also reflected in the themes explored in the special issue of *Iskolakultúra* (2002) focusing on language pedagogy or in the study by Illés (2013), in which the author investigated the place of English as a lingua franca in language pedagogy relative to both the classroom and the language teacher. The international literature also reflects this trend, as shown by the works of several authors from different periods, such as Prabhu's (1987) book on second language pedagogy, Ellis' (2000) paper on task-based language teaching, or Pütz et al.'s (2001) volume on the aspects of linguistics in language teaching.

Based on the above, the answer seems straightforward: language education policy is not identical with language pedagogy. How do we then define language education policy? Is the term interchangeable with language education? Phillips' study (2007) investigated education policy issues since it focused, among other things, on the choice of a second language; yet, the discourse is framed around language education and not

the term language education policy. In her choice of definitions, Einhorn (2015) relied on European documents and recommendations. In her interpretation, language education policy covers the system of language levels, the measurement of the effectiveness of language education, the number and range of learnt languages, the role of the English language, and the start of foreign language learning or bilingual education. Among the expected challenges of language education policy, Hidasi (2017) mentioned similar examples, but her definitions represent a kind of borderline between activities taking place at the school and education policy levels. These examples also illustrate how difficult it is—if not impossible—to separate language education and language education policy from each other. As is the case with language pedagogy, language education policy and language education constitute two partially overlapping sets.

How Can We Define Language Policy as a Construct?

Language policy is often used interchangeably with language planning (Szépe, 2001), most often in regard to its function of regulating official language use and as a conscious, intentional intervention (Hornberger, 2006; Ricento, 2006). The approaches and definitions differ because whereas for some language refers to a specific official or minority language, for others it also signifies a range of foreign languages. These differences are not coincidental given that language policy lacks a standard theoretical framework and can be approached from several disciplines (Darquennes, 2013; Ricento & Hornberger, 1996). Based on a theoretical review, Darquennes (2013, p. 12) formulated the following definition:

> language policy can be understood as the whole body of oral and/or written (in)formal texts that aim at (re)affirming or changing the language dynamics in (a part or different parts of) society [so as to distinguish it from language planning, which] can be considered as

> an attempt to influence the language dynamics in (a part or different parts of) society by means of concrete measures that address the corpus (…), the status (…), the acquisition (…) and/or the prestige (…) of a single or more language varieties.

Baldauf (2004) considered language policy to be a statement of intent and language planning as its implementation. Although these definitions do not extend to foreign languages, fitting examples have been cited: Darquennes (2013) discussed the European Day of Languages and the need for foreign language provisions, while Baldauf (2004) addressed planning early foreign language learning as a prerequisite for a knowledge economy. Cooper's (1990) initial model of language planning consisted of two elements: (1) status planning, that is the "deliberate efforts to influence the allocation of functions among a community's languages" (p. 9) and corpus planning, namely the "activities such as coining new terms, reforming spelling, and adopting a new script", or more specifically, "the creation of new forms, the modification of old ones, or the selection from alternative forms in a spoken or written code" (p. 31). He expanded his model with a third type of planning, *acquisition planning*, which is described as the "organized efforts to promote the learning of a language" (p. 157). By making the field relevant to applied linguists engaged in foreign language education, he connected it to language teaching and language learning. Baldauf (2004) equated acquisition policy (acquisition planning according to Cooper, 1990) and language-in-education policy, and assigned to them key areas such as curriculum or assessment policy. In addition, he talked about education planning, which in his view includes foreign/second language learning. Therefore, based on the cited sources, language policy also overlaps with language education policy since it aims to shape language use in the same way as language education policy does; however, the latter does so with taught and learned languages.

Conclusion

In conclusion, language education policy is a field that overlaps with language policy, language planning, and language pedagogy. It is an often used but less precisely defined construct that is interpreted in different ways. Based on the overview of the conceptualizations of the term in the highly salient body of literature discussed above, language education policy can be construed as a consciously developed and intentional intervention system that describes, regulates, and implements the aims, frameworks, and components of the foreign language education of a community, most often that of a country, against the cultural, political, historic, economic, and geographic backdrop of the given community. Its main subfields include (1) the elaboration of concepts and strategies; (2) regulation of organizations and content; (3) implementation, including the implementing groups; and (4) evaluation of the efficiency of foreign language teaching, as well as empirical data collection and analysis regarding the above.

References

Baldauf, R. (2004, May). *Language planning and policy: Recent trends, future directions* [Conference presentation]. American Association of Applied Linguistics Conference, Portland, US. https://espace.library.uq.edu.au/view/UQ:24518

Bárdos, J. (1984). Nyelvek és elvek [Languages and principles]. *Pedagógiai Szemle*, *14*(5), 471–475.

Bárdos, J. (2000). *Az idegen nyelvek tanításának elméleti alapjai és gyakorlata* [The theory and practice of teaching foreign languages]. Nemzeti Tankönyvkiadó.

Bárdos J. (2006). Módszertan vagy nyelvpedagógia? [Methodology or language pedagogy?] *A magyar nyelv és kultúra tanításának szakfolyóirata/Journal of Teaching Hungarian as a 2nd Language and Hungarian Culture*, *1–2*, 5–14.

Cooper, R. L. (1990). *Language planning and social change*. Cambridge University Press.

Darquennes, J. (2013). Language policy and planning in indigenous language minority settings in the EU. *Revue française de linguistique appliquée*, *18*, 103–119. https://doi.org/10.3917/rfla.182.0103

Einhorn, Á. (2015). *A pedagógiai modernizáció és az idegennyelv-tanítás* [The modernisation of pedagogy and the teaching of foreign languages]. Miskolci Egyetemi Kiadó. https://mek.oszk.hu/15500/15518/15518.pdf

Ellis, R. (2000). Task-based research and language pedagogy. *Language Teaching Research*, *4*(3), 193–220. https://dx.doi.org/10.1177/136216880000400302

Hidasi, J. (2017). The impact of interconnectedness: Culture, language, communication. *Nyelvvilág*, *19*, 62–69.

Hornberger, N. H. (2006). Frameworks and models in language policy and planning research. In T. Ricento (Ed.), *An introduction to language policy: Theory and method* (pp. 24–41). Blackwell.

Illés, É. (2013). Az angol mint lingua franca – új nyelvpedagógiai kihívás [English as a lingua franca – A new pedagogical challenge]. *Modern Nyelvoktatás*, *1–2*, 5–16. http://epa.niif.hu/03100/03139/00021/pdf/EPA03139_modern_nyelvoktatas_2013_1-2_005-016.pdf

Medgyes, P. (1995). *A kommunikatív nyelvoktatás* [Communicative language teaching]. Eötvös József Könyvkiadó.

Petneki, K. (2001). Nyelvpedagógia – az idegen nyelvek tanításának tudománya [Language pedagogy – The science of teaching foreign languages].

Iskolakultúra, *11*(1), 103–106. http://www.iskolakultura.hu/index.php/iskolakultura/article/view/19408

Phillips, J. K. (2007). Foreign language education: Whose definition? *The Modern Language Journal*, *91*(2), 266–268. https://doi.org/10.1111/j.1540-4781.2007.00543_9.x

Prabhu, N. S. (1987). *Second language pedagogy*. Oxford University Press.

Pütz, M., Niemeier, S., & Dirven, R. (Eds.) (2001). *Applied cognitive linguistics*. De Gruyter Mouton.

Ricento, T. (2006). Methodological perspectives in language policy: An overview. In T. Ricento (Ed.), *An introduction to language policy. Theory and method* (pp. 129–134). Blackwell.

Ricento, T. K., & Hornberger, N. H. (1996). Unpeeling the onion: Language planning and policy and the ELT professional. *TESOL Quarterly*, *30*(3), 401–427. https://doi.org/10.2307/3587691

Szépe, Gy. (2001). *Nyelvpolitika: Múlt és jövő* [Language policy: Past and future]. Iskolakultúra.

2

Probing the Reading Paper of the Advanced Level EFL School-Leaving Examination

Gyula Tankó
tanko.gyula@btk.elte.hu

Zsuzsanna Andréka
andreka.zsuzsanna@btk.elte.hu

Abstract

The two-level school-leaving examination system introduced in Hungary in 2005 and modified in 2017 is the most important gatekeeping examination for English major tertiary education programs. Experience shows that the language proficiency of the students admitted into the English major education programs is insufficient for effective learning and teaching. Therefore, this exploratory investigation analyzed the reading paper of the advanced EFL school-leaving examination to find out the extent to which it is suitable for the level-appropriate assessment of the aspects of language ability it is intended to assess. For this purpose, 32 tasks from eight reading papers administered in the 2017–2020 period were analyzed. With the use of statistical data available on the scores obtained by test takers alongside a close critical analysis and blind double-coding of test items, several shortcomings of the reading paper were revealed. The flaws uncovered undermine the construct validity of the examination and call into question the generalizability of its scores and its predictive validity. The findings may partly explain the ineffective reading ability of the students admitted into English major programs and should aid in the development of better reading papers.

Keywords: Hungarian EFL school-leaving examination, university admission, construct validity, reading assessment task analysis

DOI: 10.21862/ELTE.DEAL.2022.2.23

Probing the Reading Paper of the Advanced Level EFL School-Leaving Examination

The current two-level school-leaving examination system was introduced in Hungary in 2005. It originally measured foreign language proficiency at the intermediate (A2/B1) and advanced (B2) levels (Council of Europe, 2020), but in 2017, amongst other smaller modifications, the level of the intermediate examination was changed to B1. The foreign language school-leaving examination is compulsory for all secondary school leavers in Hungary (Hungarian Government, 2021), who must start learning a foreign language from grade 4 and a second one from grade 9 (Hungarian Government, 2020)—with English being the most popular language in all secondary level school types (Hungarian Central Statistical Office, 2022)—and who must take either an intermediate or advanced level foreign language school-leaving examination in grades 11 or 12 (Hungarian Government, 2020). In the spring of 2022 alone, 64,958 secondary school leavers took the English school-leaving examination, of which 18,791 chose the advanced examination (Educational Department, 2022b).

The aim of foreign language education in Hungary, as stated in the National Core Curriculum (Hungarian Government, 2020) is to develop the communicative competence of language learners so as to enable the appropriate realization of their communicative goals in authentic language use situations. Further foreign language education goals set by the curriculum are to capacitate language learners to access information efficiently, to assist their mobility for study purposes, to facilitate their entry into the labor market, and to aid them in the pursuit of their tertiary level studies.

An advanced level EFL secondary school leaving certificate is necessary for admission to English major tertiary education BA programs in Hungary, and either an intermediate or advanced level certificate is required for MA teacher training programs depending on the combination of subjects selected by the applicants (Educational Department, 2022c).

For example, secondary school graduates applying for an English-German teacher training MA program can meet the English language proficiency admission requirements even with an intermediate level certificate if they have an advanced level school-leaving certificate in German. However, even if the admission decisions are made on the basis of an advanced level school-leaving certificate in English, the question remains as to what the reason for the low language ability of first-year English major students admitted into English medium instruction programs may be. One of the likely explanations suggested by research findings (e.g., Dávid, 2008; Szabó & Kiszely, 2010; Tankó & Andréka, 2021) is that the advanced level EFL school-leaving examination itself does not assess language ability as intended. Given that further empirical investigation of the advanced level EFL school-leaving examination is needed, especially because high-stakes admission decisions are made based on it, the current study was carried out to add to the rather limited body of research (see Illés, 2011; Szabó, 2019) available on the reading paper of the advanced level EFL school-leaving examination.

Theoretical Background

Reading in Language Assessment

As Alderson (2000) discussed, assessing reading ability is a complex process determined primarily by the model of reading based on which assessment developers define the construct of reading for a specific purpose or setting. In order to differentiate poor and good readers or to make predictions based on their test performance about how they are likely to perform in other settings on other reading tasks, assessment developers must also make decisions on whether they wish to assess the reading process (i.e., how comprehension is reached) or product (i.e., the poor, fair, or good comprehension that is reached). This requires them to consider what levels of understanding they want to measure (i.e., whether they wish to assess the comprehension of literal or inferred text meaning,

or the ability to critically evaluate a text) and how to distinguish acceptable and unacceptable interpretations of a text. Furthermore, so as not to test them explicitly, assessment developers have to separate reading ability from other cognitive abilities or functions (e.g., reasoning or working memory); from the readers' background, topical, and cultural knowledge; as well as from other components of language ability—such as grammar and vocabulary—that have been found to correlate with reading ability measures. This list of factors, variables, and decisions to be made is far from complete, but it is sufficient to illustrate the complex nature of reading proficiency assessment and why assessment scholars tend to shy away from formulating a definition of reading ability.

A rare exception is the definition proposed by Urquhart and Weir (1998), according to whom "reading is the process of receiving and interpreting information encoded in language form via the medium of print" (p. 22). However, a mere cursory look at this definition is enough to appraise the difficulties it poses for an assessment developer who attempts to operationalize it. Instead, reading assessment developers typically measure reading strategies, skills, or sub-processes that are part of the model of reading they adapt (see Alderson, 2000).

An influential and modern taxonomy of reading behaviors used to assess general and academic English reading ability was proposed by Khalifa and Weir (2009). It is part of the cognitive processing model for reading comprehension and differentiates reading behaviors across two dimensions: careful versus expeditious and local versus global. The faster, goal-driven, selective reading behaviors—namely, skimming, search reading, scanning, and browsing—are expeditious reading strategies. The slower ones—which include the understanding of lexis and grammar which is usually not explicitly assessed—are careful reading skills.

Context in Reading Tasks

Alderson and Cseresznyés (2003) described readers in a modern language examination as being actively engaged with a variety of texts similar to the

ones they may potentially encounter in real life situations. Their reading process can be selective due to either the reading goal based on which they determine whether an idea or stretch of text is irrelevant or not, or because they have to skip unknown lexis. Moreover, the reading process is also understood to be flexible; that is, it varies according to the reading goals set by the task instructions. These reading purposes also aim to be authentic, which means that they simulate real life texts and the purposes for which they are normally read.

In modern examinations, readers are expected to use their language ability meaningfully in context. Therefore, much the same as in the case of the assessment of Use of English (UoE), the role of context is crucial for assessing reading ability. However, in the case of a UoE task, test takers have to read and understand a context (i.e., the text surrounding the item) that is markedly easier than in the case of a reading task. As Alderson and Cseresznyés (2003) stated, in a UoE task the context "will normally not contain unknown words, apart from those being tested, and will not have complex structures, other than those that might be being tested" (p. 27). In such a task, it is the meaningful use of a grammatical structure or lexical item that is in focus. In contrast, in a reading task the focus is on the use of various reading behaviors while processing a text, or on specific abilities such as reading critically and being able to differentiate fact from opinion or main idea from supporting detail. The context of the items in a reading test task can be below, at, or even above and, consequently, considerably more difficult than the assessed level. However, the part of the text targeted by the test item must be at the assessed level. Finally, it is also based on context that readers may be expected to make plausible inferences about meanings not explicitly expressed in the text.

Inferencing and Reading Comprehension

A simple definition of *inference* is "any piece of information that is not explicitly stated in a text" (McKoon & Ratcliff, 1992, p. 440). During the reading process, information present in a text can be condensed and

information not explicitly present in the text can be added by readers. Information reduction as well as accretion processes (i.e., generation and construction) are controlled by macrorules and result in macropropositions (van Dijk, 1980), namely, inferences that reduce information in and add information to a text (see Kintsch, 1993). Both types of inference aid reading comprehension and the storage of information in short- and long-term memory (Kintsch & van Dijk, 1978).

Research evidence unequivocally revealed that comprehension was obstructed without the use of relatively simple bridging inferences (Singer, 1994). These backward inferences help create discourse coherence by relating an idea to the discourse that precedes it (e.g., the explicitly signaled anaphoric reference). Contrarily, elaborative or forward inferences are not a requirement for comprehension, but empirical research findings also demonstrated that they do improve it (Singer, 1994). Moreover, whereas deductive inferences are controlled by formal rules operating on information explicitly present in the text and result in verifiable presumptions, elaborative inferences, also referred to as pragmatic inferences, depend on the reader's world knowledge (e.g., schemata, scripts, frames, memory organization packages, or stereotypes; Graesser & Kreuz, 1993) and add probable information to a text, such as a prediction about the contents of the text based on its title or a conjecture concerning the writer's attitude (Schmalhofer et al., 2002; Singer, 1994; Singer & Lea, 2012).

Reading Assessment Task Features Affecting Comprehension

Two components of the reading assessment task, titles and images, function as advance organizers. When they are clear, age- and content-appropriate, and functional both in terms of thematic relevance and reproducibility through test booklet printing (Alderson & Csereszny és, 2003; Tankó, 2005), both foretell the topic of the reading passage and by priming it help readers to relate the topic to their background knowledge. Meyer (1982) noted that a title has a signaling function as it "prematurely

reveals information abstracted from the content occurring later in the text" (p. 77). A title, therefore, aids comprehension as it facilitates the construction of a hypothetical coherent mental representation of the macrostructure (i.e., predicted content) of a text (Kintsch, 1988; Soederberg Miller & Stine-Morrow, 1998). Specifically, macrostructures achieve this effect by activating (i) relevant background information (i.e., schema) stored in long-term memory (Kintsch, 1998) and relevant vocabulary—both of which ease the comprehension of even less well composed texts—as well as (ii) cognitive frameworks that can be modified with new information (van Dijk & Kintsch, 1983). A number of research studies confirmed that on-line comprehension and recall (e.g., Bransford & Johnson, 1972; Miller et al., 2006; Smith & Swinney, 1992; Wiley & Rayner, 2000), reading time (e.g., Soederberg Miller & Stine-Morrow, 1998; Wiley & Rayner, 2000), processing of ambiguous words (Wiley & Rayner, 2000), and working memory demands (Miller et al., 2006) improve if a title is provided prior to reading and if it activates background knowledge. Naturally, what is implied here is that the title must be rhetorically functional, as it was intended by the writer of the text. Research evidence showed that altering or changing original text titles resulted in the construction of different mental representations of the same text (Bock, 1980); this should be avoided in assessment as it affects the justifiability of the results.

In a meta-analytic study of experimental research covering 33 years (i.e., 1985–2018), Guo et al. (2020) reported that graphics had a positive effect on reading comprehension. Notably, when graphic types were compared, pictures were found to have the most pronounced effect. However, pictures that were not organically related to the content of the text were also found to impact comprehension, albeit negatively (Wiley, 2019). A study on the effect of decorative pictures (i.e., those with a mainly aesthetic function) and instructional ones (i.e., those with an informative function) revealed that the participants paid little attention to decorative pictures and that such pictures had no effect on comprehension and learning. In contrast, informative pictures had an effect that, interestingly,

was enhanced by the presence of decorative ones (Lenzner et al., 2013). In a similar study investigating readers' metacomprehension accuracy of expository science texts, Jaeger and Wiley (2014) found that decorative images negatively affected metacomprehension accuracy.

In addition to titles and images, the task instructions and the linguistic accuracy of the input also affect comprehension. When provided in the assessed language, the level of the instructions must not be higher than the assessed proficiency level (Bachman & Palmer, 2010). Furthermore, operational test tasks must be reviewed to ensure that the efficient processing of the input is not hindered by spelling errors, incorrect grammar, or formatting problems (Fulcher, 2010). Additionally, to help processing, the input must also be well-formed and correctly punctuated because, as Tankó (2022) noted, punctuation aids text processing and affects comprehension through its disambiguating function. Another reason why accurate input must be provided is that test takers are believed to learn even while taking a test (Bachman & Palmer, 2010).

Given that they affect comprehension, the task features discussed above jeopardize the justifiability of the interpretations made regarding reading ability. Justifiability being a validity issue, the last part of the review discusses construct validity.

Construct and Criterion-Related Validity

Messick (1995) differentiated between two major types of threats to construct validity that can occur simultaneously: (i) construct underrepresentation incurred by the narrow and therefore ungeneralizable assessment of a construct and (ii) construct-irrelevant variance induced by the broad assessment of a construct. The latter threat has two subtypes: construct-irrelevant difficulty, which is caused by chance factors unrelated to the measured construct that make the completion of a task difficult; and construct-irrelevant easiness, which is caused by task formats that allow a test taker to answer an item correctly

without engaging the assessed construct or processes. Construct-irrelevant variance is especially important in the case of assessments where context is important as it "matters whether the contextual clues that people respond to are construct-relevant or represent construct-irrelevant difficulty or easiness" (p. 743).

Messick (1980) also described two types of criterion-related validity: "concurrent validity and predictive validity, which differ respectively in terms of whether the test and criterion data were collected at the same time or at different times" (p. 1016). Given that the advanced level examination certificate is used for making admission decisions to universities, the implied claim is that it indicates the test takers' future level on the criterion, namely how well they will function in an English medium education context. The relationship between construct and criterion-related validity is that if the former is undermined, the latter collapses.

Purpose of the Present Study

The English major programs at Eötvös Loránd University are popular. In the autumn semesters of 2021 and 2022, 223 and 209 (N = 432) students were admitted to the MA in English teacher training programs in addition to the 270 and 265 (N = 535) students admitted to the BA in English program (Educational Department, 2022a). This means that close to half of the admitted students were not required to have a B2 level certificate in English, which is a problem in itself because the minimum proficiency level needed for academic purposes is B2 (Kirkland & Saunders, 1991). This can partly explain the low language ability of first year English major students that has been causing problems for both students and teachers. However, an additional concern is the increasing body of evidence indicating that the B2 level EFL school-leaving examination does not assess language ability as intended (e.g., Dávid, 2008; Szabó & Kiszely, 2010; Tankó & Andréka, 2021).

Given that high-stakes admission decisions are made on the basis of the advanced level EFL secondary school leaving certificate, the

justifiability of the assessment needs to be investigated. As a consequence, this research study was carried out in order to analyze the reading paper of the B2 level EFL school-leaving examination. The research question it proposed to answer was the following: To what extent is the reading paper in the advanced level EFL school-leaving examination suitable for the level-appropriate assessment of the aspects of language ability that it intends to assess?

Methods

To answer the research question, a qualitative content analysis study was carried out to analyze reading papers from past EFL school-leaving examinations. The first section in this part gives a brief introduction to the EFL school-leaving examination. The second section describes the reading papers analyzed, and the last section presents a summary of the data analysis.

The EFL School-Leaving Examination

The EFL school-leaving examination is administered for secondary school students in Hungary twice a year, in May and in October, at two levels: The intermediate level is intended to be at level B1 and the advanced level at level B2 (Council of Europe, 2020). The examination consists of a written part (which includes Reading, Use of English, Listening, and Writing papers) administered in one sitting with a break and an oral part administered on a separate day after the written part.

A B2 level state-accredited language examination certificate is issued if a test taker achieves a minimum of 60% on both the written and oral parts of the advanced level EFL school-leaving examination. A test taker whose score is between 40%–59% receives a B1 level state-accredited language examination certificate (Hungarian Government, 2022). Admissions officers award additional points for those applying to tertiary university programs who hold a B2 level EFL school-leaving examination

certificate stating that they have achieved a minimum score of 45% (Educational Department, 2022c).

The Reading Papers Analyzed

The advanced level reading papers investigated (representing the 2017–2020 period) were administered together with the Use of English paper in the first half of the written examination. Test takers had 70 minutes to complete it and—depending on the number of items—could get maximum 28 or 30 raw points, which were converted to 30 final points. The reading paper accounts for 25% of the total score for the written part, in which the four papers are equally weighted. For the successful completion of the written part of the EFL school-leaving examination, test takers must achieve a minimum of 12% (Hungarian Government, 2022). Therefore, assuming that test takers can pass the written part of the examination by scoring 12% on each paper, depending on the number of items in the paper, a minimum of three or four reading items must be answered correctly, the equivalent of merely four converted points, which is a disconcertingly low cut score.

Data Analysis

The first version of the codebook used in this study and a set of analytical decision rules were created based on the specifications available for the advanced level reading paper and the relevant literature on the reading construct and assessment. An example of an analytical decision based on the "*minimal effort necessary to solve the item correctly"* principle (Tankó & Andréka, 2021) is the one stating that in cases when the correct response to an item can be given based on a semantic, syntactic, or form matching decision, the decision type to be recorded is the one that requires the least effort. For example, in 2019-i-T1-I25 (i.e., year 2019, first take, task one, item 25), the co-text before the gap "These sites are built to be engaging, (25) _____ is addictive for others." clearly cues the option "... and what's

engaging for some ..." because of the word present in both, so making the semantic lexical repetition link between "some" and "others" or the grammatical link with the repetition of the preposition "for" may only be needed as reassuring check possibilities.

Following this, a pilot sample of the tasks (25%) was solved by the two authors, the items were coded, and following a consensus-building discussion the codebook and analytical rules were revised. Then each author independently coded the items in the remaining papers. During the coding process, the codebook and the coding rules were updated when new coding issues emerged. The coding was conducted with the assistance of various tools, including the MS Excel software to count words and edit the coding form; the CEFR-based Vocabulary Level Analyzer (ver. 2.0; Uchida, 2022), which estimates the CEFR level of an input text; Multimodal Analysis Image, a software for image annotation and analysis (trial version; Tan et al., 2012); Textinspector, a web-based linguistic analysis tool that produces metrics benchmarked to the CEFR (Weblingua, 2022); and English Vocabulary Profile Online, which is a reference database based on the Cambridge Learner Corpus that assigns a CEFR level to the lexis in texts (Cambridge University Press & Assessment, 2015). The results of the coding were compared, codings which did not match were discussed, and a final set of jointly approved codings were created for analysis.

Reading Paper Test Specification

According to the detailed school-leaving examination specification (Ministry of Education, 2002) from which several of the categories were selected for the code book, the reading paper aims to measure test takers' ability to read independently and comprehend various kinds of real-life authentic texts with the use of appropriate strategies and at the level of specificity appropriate to the set reading purposes. Although in the poorly organized specification this information is added in a seemingly random manner to a thematically unrelated section, the types of comprehension to

be measured are global, selective, and detailed comprehension. It needs to be noted here that none of the constructs are defined, which raises questions about their operationalization. Furthermore, the specification not only fails to link the measured reading abilities with the types of comprehension named, but the listed abilities all denote reading activities that can only be achieved through global careful comprehension (see the taxonomy of Khalifa & Weir, 2009, described in the Theoretical Background section), which is a non sequitur. The reading paper is supposed to measure the test taker's ability to follow a train of thought, opinions, and arguments; understand information in sufficient detail (NB whether at the global or local level remains unspecified); and infer the writer's point of view as well as the feelings and emotions of the writer or characters (i.e., formulate elaborative inferences). Using Gray's (1960) phrasing, the paper aims to assess the test taker's ability to read *the lines* as well as *between the lines*.

The input is to be authentic (but may be edited), so it may contain "words, phrases and structures whose level exceeds that of the examination, *but which are not necessary for the successful completion of a task* [emphasis added]" (Ministry of Education, 2002, Advanced level examination section); in addition, the input should be straightforward in content; well-organized; concrete or abstract; and thematically suited for the experience and general interest of the age group. In terms of prior knowledge demands, it must be at the level of the general knowledge of a secondary school leaver, on a topic specified in the detailed requirements, and finally, level-appropriate as regards linguistic and content complexity. A variety of genres are also specified, ranging from user's manuals and newspaper articles to academic and fictional literature.

The specification names an impressive array of task types from which reading item writers can choose freely and which can be used in the reading paper in any combination. The list comprises matching (at least 14 subtypes), ordering (three subtypes), multiple choice, true/false/not stated statements, short answer questions, open or banked cloze, gapped summary, and grouping according to given categories task types.

The reading paper may consist of 3—4 tasks, each with an English language instruction and with one longer or several shorter input texts per task. The total input length must be between 1,300—1,500 words, and the paper must consist of 25—30 items.

Codebook

Due to the lack of attention to technical detail and incoherence problems, the specification summarized above had to be elaborated on the basis of the available literature on reading comprehension before a coding scheme could be designed. The codebook that was written for this study consists of two main parts (i.e., task characteristics and item characteristics) featuring eight and six variables, respectively. Of these, 10 were nominal and four were interval variables. The coding scheme, indicating measurement levels and offering brief descriptions of the variables together with information on whether the coding was human only or computer assisted, can be found in the Appendix.

Coder Agreement

To check the reliability of the coding, Cohen's κ reliability coefficient was used; if this was not possible, percentage agreement was calculated instead. There was perfect agreement (Landis & Koch, 1977) in two cases (Scope of relationship, $\kappa = .845$, 95% CI [.79 to .92], $p < .001$; CEFR level of title, $\kappa = .848$, 95% CI [.76 to .93], $p < .001$]), almost perfect agreement in three cases (Reading behavior type engaged–Category B, $\kappa = .858$, 95% CI [.80 to .92], $p < .001$; Linguistic decision required by response, $\kappa = .903$, 95% CI [.84 to .96], $p < .001$), and substantial agreement in two cases (Reading behavior type engaged–Category A, $\kappa = .747$, 95% CI [.67 to .82], $p < .001$; CEFR level of item, $\kappa = .633$, 95% CI [.56 to .70], $p < .001$). A high percentage agreement was found in three cases (Comprehension level, 99%; Task type, 97%; and Image-text intersemiotic sense relations, 91%). In the case of those

variables where only computer-generated indices were used (i.e., Length per task/paper and CEFR level of input) or where there was 100% agreement between the coders (e.g., Number of items per task/paper), no intercoder reliability index was calculated.

Results and Discussion

The outcomes of the analysis are presented in this section according to the task and item characteristics variables investigated.

School-Leaving Examination Results

Within the 2017–2020 period investigated, altogether 60,691 secondary school students registered for the advanced level EFL school-leaving examination. From these, 59,976 took the examination and 57,686 (96%) passed (Educational Department, 2022b). Based on the analysis of the results of the test takers who had a reading score recorded, the majority of the test takers managed to receive fairly high scores on the reading paper (N = 59,976; M = 22, Mdn = 23, SD = 5.548; $Q1$ = 19, $Q2$ = 23, $Q3$ = 26). Furthermore, altogether 63 students had a converted reading score of 4 points, and of these 35 (56%) passed the EFL school-leaving examination and became eligible—some (n = 4) with additional points awarded for a minimum of 45% achievement—for admission to English major programs offered in Hungary. This serves as evidence that a student with the minimum acceptable EFL reading score can become eligible for a tertiary English major program, but the low number of such cases found is moderately reassuring.

Task Characteristics

Task Types

The advanced level reading papers were selected for this study from the eight examinations administered in the 2017–2020 period. Each reading

paper analyzed consisted of four tasks, and the number of items per paper ranged from 28 to 30 ($f_{30} = 5$, $f_{29} = 2$, $f_{28} = 1$). The task types included were matching sentence segments (i.e., clauses or phrases) to gaps and true/false/not stated (each $n = 7$, 22%); matching lexical items to gaps (i.e., open cloze), multiple choice, and matching sentence beginnings to ends (each $n = 4$, 13%), filling in a list of gapped sentences ($n = 2$, 6%) or a gapped summary ($n = 1$, 3%) based on the input; and matching complete sentences to gaps, paragraphs to gaps, or questions to answers (each $n = 1$, 3%).

In spite of the fact that on the basis of the specification nine main task types were included in the codebook, the reading papers only contained five of these. As could be expected given that the matching main task type had 12 subtypes, matching tasks were used most frequently in the reading papers. Of these, three (i.e., matching sentences/paragraphs to gaps and questions to answers) tested global text organization, namely coherence, which—although not irrelevant in terms of reading comprehension—is also tested in the writing and speaking parts, which should be sufficient for decision making. Instead, other task types like short-answer would contribute more relevant information for reading comprehension assessment and improve the generalizability of the results. The true/false/not stated task type was also frequent.

What is difficult to explain is the switch from gapped summaries ($n = 2$, 2018-i-T2, 2019-ii-T1) to gapped sentences (2020-i-T2) over the years. In the case of a gapped summary, the test taker reads a continuous text and contrasts its macrostructure with that of the input text. This task is cognitively more demanding and arguably much more authentic than comparing the content of sentences from a list to an input text. The cognitive load derives from the complexity and number of operations to be performed. Its authenticity becomes obvious if we consider the relationship between the headline-and-lead advance organizer dyad and the body of a news article, or between an abstract and the full text research paper. The headline and the lead together add up to a selective (also known as guided, Tankó, 2019) summary (Bell, 1998), whereas a research

article abstract is a global summary of the paper. Both these tasks illustrate common, real life reading activities from the general and academic target language use domains. Furthermore, the variation in terms of the main task types used from 2017 to 2018 is considerable ($n_{2017} = 2$, $n_{2018} = 5$), which raises justifiability issues concerning consistency across different assessment administrations.

Length of the Input

According to the specification, the overall length of the input text must be between 1,300–1,500 words per reading paper. The average length of the input per paper was 1,470 words, with a narrow range of 1,434 to 1,495 words, which is consistent with the specification. The average length of the input per task was 367 words, with a large range of 292 to 461. The multiple-choice tasks, however, add a considerable reading load with their verbose options, leading to inconsistency in the amount of input to be processed across years. Furthermore, the amount of input to be processed in the test items also varied markedly within the multiple-choice tasks; in fact, it more than doubled in the 2020 spring task compared to 2019 (2019/i/T4, $n = 158$; 2018-ii-T3, $n = 221$; 2019-ii-T3, $n = 270$; 2020-i-T3, $n = 330$ words). This raises concerns in terms of the consistency of the assessment across different takes.

CEFR Level of the Input

The overall CEFR level of the input texts was assessed with the CEFR-based vocabulary level analyzer (Uchida, 2022). The levels were found to range from B1.1—the lowest level within the B1 band (Uchida & Negishi, 2018)—to C2, the highest defined in the CEFR. As Table 1 shows, the overall CEFR level of 20 (62%) reading input texts was above the B2 band level, while five (15%) were at levels below it.

Table 1

Reading Input Text CEFR Levels in the B2 Level Examination

Level	*f*	%	*cum %*
C2	11	34	34
C1	9	28	62
B2.1	5	16	78
B1.1	3	9	87
B1.2	2	6	94
B2.2	2	6	100
Total	32	100	

Trained item writers can construct B2 level reading comprehension items for an input text whose overall difficulty level exceeds the level of the examination. However, their job becomes challenging and maybe even impossible when they have to write items for input that is barely at the B1 level as they are not supposed to counterbalance the low difficulty level of the input with a high difficulty level item. In fact, the exact opposite is recommended (Alderson, 2000).

Input Text Titles

The titles of eight input texts (26%) were above B2 level ($n_{C1} = 4$; $n_{C2} = 4$) according to the coding rule which specified that the level of the highest CEFR level lexical item should be recorded as the indicator of overall title difficulty level. The rule was formulated with awareness of the fact that readers skip lexis they do not understand (Alderson & Cseresznyés, 2003); however, this is not exactly the case with titles, which are macropropositions with important discourse functions. As discussed in the review of the literature, understanding a title is important because it loads knowledge frames and activates vocabulary that enhances comprehension. The problem above could be mitigated with the inclusion of the topic of the input text in the task instructions. However, not all the instructions were found to do this (e.g., the discourse topic announcement

is missing from the instruction of task 2017-i-T1), and it is common knowledge that—most likely due to a sense of security deriving from testwiseness acquired through classroom test preparation—most students do not read the instructions.

Since the level of some of the lexical items appearing in the titles could not be estimated with the English Vocabulary Profile, the analysis was most likely unable to reveal all the level-related problems with titles (e.g., an extreme instance of this is "THE **A1** PERILS **??** OF 'TABOO' **??** GIFTS **A2**;" 2017-ii-T2 eventually coded as A2, where *taboo* is guessable, so easy, but *perils* is more likely a C1 level item like *hazard* or *threat*). Nevertheless, the analysis did reveal several other issues of which the most important are presented here: One input text had no title at all (2020-ii-T1). Modifications of the original titles and functionally related components resulted in distorted discourse topic signaling. For example, the title "The owl thieves of Sweden" (2020-ii-T2) should introduce a text about cash not being used anymore. However, it fails to do so because a fully functional lead present in the original "As the country ditches cash, criminals turn to stealing owls" was deleted, which disconnected the title from the text and raised the difficulty level of the input in an inauthentic way. Another modification type compromised the macroproposition function of titles. For example, only the first three of the six paragraphs in the text entitled "What's in the queen's handbag?" (2017-i-T1) discuss what is in the handbag; the remaining ones provide explanations about the functions of the bag (e.g., signaling device). The input bears close similarity to an online article entitled "What's inside the Queen's handbag and why is it so significant?" (Hello! magazine), which—unlike the test task version—does anticipate the discussion of reasons. Furthermore, several titles contained mistakes introduced by item writers. A title which was originally "It's a WET wedding! Hero groom jumps into a river during photoshoot with the bride to save drowning boy" (Daily Mail Online) was changed to "Canada groom rescues boy from lake" (2018-i-T1)—note the incorrect use of a noun instead of the adjective. A punctuation mark such as a colon might have been intended to be added after the first word.

Because it is beyond the scope of this paper to discuss punctuation errors in detail, it can only be noted here that the instructions in 23 tasks contained one punctuation mistake, 68 punctuation mistakes were found in the body of the input texts (e.g., $n = 7$ in 2017-i-T4 and 2019-ii-T2 each), and 26 were found in the test items (e.g., $n = 5$ in 2019-ii-T3). Not only do the tasks become unduly difficult when punctuation cannot perform its text disambiguating function (Tankó, 2022), but it also potentially teaches test takers incorrect English use—if not during the test, then when teachers use the tasks in their classes.

Image-Text Intersemiotic Sense Relations

Each reading task featured an image. Some of these ($n = 11$, 34%) were acceptable as they set the context and potentially helped the activation of the schema necessary for comprehension. Such images illustrated the input (e.g., a picture of Christopher Marlow with a text about the playwright, 2018-ii-T2; or a picture of a hornet that illustrated the insect discussed in the text, 2019-ii-T3). Attempts made to illustrate more complex text content failed, and the remaining images were not functional because they were indiscernible due to their size (e.g., 2019-i-T3), quality (e.g., 2018-i-T4), or because by turning color images into black and white ones, important information was lost (e.g., 2020-ii-T4, where the image is supposed to be a heat map illustrating climate change with colors). Other images required age-inappropriate prior knowledge and failed to cue the discourse topic (e.g., 2017-i-T3). Instead of being informative, some provided irrelevant and misleading details (e.g., 2020-i-T1, where an image depicting a meeting held in the Whitehouse accompanied a text about bureaucracy in the UK, Austria, and companies in general).

Item Characteristics

Altogether 236 regular test items and 32 example items provided in the reading tasks were double-coded; the coding was finalized and the dataset analyzed. The number of items per task ranged from five to nine

(M = 7.38, Mo = 7, SD = 1.212), and there were 28 to 30 items in a reading paper (only three papers had less than 30 items: $n_{2017\text{-i}}$ = 28, $n_{2018\text{-I \& 2020-ii}}$ = 29), so the number of items per paper matches the test specification.

Comprehension Level

The results of the analysis showed that except for four items (1%), the reading papers tested literal comprehension. This does not match the specification to a desirable extent because, as summarized in the methodology section, the specification emphasizes that the reading paper assesses an extensive range of inference types. One of the inference items found in a True/False/Not stated task targeted the last paragraph of the text entitled "Three ways to train your brain to cope with heavy travel." The paragraph and the item are the following:

> **Input text paragraph #6**: **[(1)** If you feel sleepy during daylight hours when you first arrive somewhere new, try and do some aerobics.**]** Even if you do not feel tired in the evening, try to sleep anyway. **[(2)** And avoid drinking a coffee when you hit that wall in the afternoon.**]** Caffeine will only make the process much harder when it's time for bed. Smartphone use before bed is the ultimate no-no. The blue light emitted from it can trick your brain into thinking it's daytime and therefore block the production of the hormone melatonin, which would normally help you sleep.
>
> **Item 28**: Exercising or having coffee will have similar effects if you feel sleepy during the day. (2018-i-T4)

The idea that exercise helps reduce daytime sleepiness is implied only in Text Segment 1 because it does not explicitly state that exercise will wake up the jetlagged traveler. Nor does Text Segment 2 explicitly state that a coffee in the afternoon has the same effect—it also only cues this information, so it needs to be retrieved from prior knowledge. However, additional inferencing ability is required from the reader to understand

the macro-level relationship between the item and the paragraph. The reader must infer the analogy implicitly present in the first part of the paragraph that the item targets: exercise and coffee will have the same positive effect during the day (but not in the evening). In order to answer such items, readers must combine information across sentences within a paragraph (i.e., engage in global reading).

Scope of the Relationship

Most of the items in the reading papers analyzed measured local comprehension ($n = 160$, 60%). This means that most items could be answered by reading individual sentences with little need to take into consideration the context. Given that global reading (which requires the construction of a coherent meaning representation across sentences) is more cognitively demanding, the relatively low frequency of items engaging global reading behaviors may make the reading paper easier than it is intended to be.

Reading Skills and Strategies

The ratio of items that engaged the test-taker's reading skills versus their strategies also reflects the narrow scope of most of the items discussed in relation to the previous variable. The majority of the reading items required the use of skills ($n_{\text{Skill}} = 170$, 63%), some necessitated the joint deployment of strategies and skills ($n_{\text{Strat. \& Skill}} = 97$; 36%), and there was very little emphasis on strategies alone ($n_{\text{Strat.}} = 1$; 4%). Therefore, it can be concluded that most items measured careful reading skills, namely, the parsing of lexis and syntactic structures. Given that these behavior types are processes, the coders categorized the items based on the "*minimal effort necessary to solve the item correctly"* principle (Tankó & Andréka, 2021) that they applied as they solved the tasks themselves. For more pertinent insights, actual performance data and information about actual test taker's comprehension processes would be needed as in the case of the next variable.

Expeditious and Careful Reading Behaviors

The findings about expeditious and careful reading behaviors confirm the general overemphasis on careful reading in the reading papers discussed in relation to the previous variable. All the items could be answered by engaging careful local and global reading behaviors ($n_{\text{Car. Local/Within sentence}} = 160, 60\%$; $n_{\text{Car. Global/Across sentences}} = 93, 35\%$; $n_{\text{Car. Global/Across paragraphs}} = 14, 5\%$; $n_{\text{Car. Global/All text}} = 1, 4\%$). The fact that all the items could be solved with careful reading is disquieting in light of the narrow scope of most items and with respect to the specification according to which the reading paper measures selective reading ability. Admittedly, it could be argued that any item measuring any type of expeditious reading behavior can be answered with careful reading, providing there is sufficient time given for the test taker to substitute skimming, scanning, and search reading with careful local and global reading, but this is unlikely to apply to most test takers under the time constraints of the examination.

CEFR Levels of the Items

In spite of the computer assisted coding, the most difficult variable to code was the one involving the assessment of the CEFR levels of the reading items. Most of the problems were caused by the lexical items for which no CEFR levels were available. Once again, the rule followed during coding was that the level of a test item was to be recorded according to the level of the highest CEFR level lexical unit it contained or according to the lowest if a list of solutions was provided in the marking key, the latter based once again on the *"minimal effort necessary to solve the item correctly"* principle (Tankó & Andréka, 2021). The use of the rule is justified given that the complete comprehension of a well-written reading test item is necessary for a correct response; otherwise, most likely the assessor will be faced with construct irrelevant variance issues. Partly for this reason, the rule was not applied in the case of those lexical items that could be guessed easily based on L1 knowledge (e.g., "pyramid," C1, 2017-i-T2, "piramis" in Hungarian; "clichés," C2, 2018-ii-T1, "klisé" in Hungarian).

In such cases, the second highest CEFR level was recorded for the item. Any instance of the use of a proper name in the item and cases when an item could be answered with one word for which no estimated CEFR level was available (e.g., "archery," 2019-1-T2) was coded as "NA." The results are summarized in Table 2.

Table 2

Reading Item CEFR Levels in the B2 Level Examination

Level	*f*	%	*cum %*
B2	112	42	42
B1	53	20	62
C1	42	16	77
C2	23	9	86
A2	21	8	94
A1	13	5	98
NA*	4	1	100
Total	32	100	

*Reading test item containing a lexical unit that was a proper name or a one-word lexical unit whose estimated CEFL level was not known.

Given that 25% of the items contained C1 and C2 level lexis, these items were above the level intended to be measured by the examination. Since text difficulty is best predicted by vocabulary difficulty (Alderson, 2000), the responses given to these items provided more information about the difficulty of the task induced by the item than about the state of the test takers' reading ability. This increased level of construct complexity most likely resulted in construct-irrelevant difficulty and does not match the specification as these lexical units are necessary for the successful completion of a task. The level of the items below B2 level could actually be optimal provided the input content they targeted was of the level of difficulty intended to be measured by the examination.

Linguistic Decisions Required by the Response

The analysis of the types of decisions based on which a correct response to an item could be given showed that the majority of the items required a semantic decision ($n_{\text{Sem.}} = 207$, 77%). The second most frequent decision type required the combination of form and meaning cues ($n_{\text{Sem.}} = 28$, 10%). This means that one expeditious reading strategy behavior, scanning, could be used—even if infrequently—to respond to an item as the lexis in the item and the input were identical in form, which allows for string search. Semantic decisions, enhanced by syntactic cues, represented the third and almost equally frequent decision type ($n_{\text{Sem. \& [Syntax]}} = 23$, 9%). The low frequency of such items is actually reassuring because the reading construct, as defined in the specification, does not include syntactic ability. It is also encouraging that only two instances were found when the reverse applied, and five instances when the decision was equally informed by semantic and syntactic cues. There were three instances when the test taker had to rely on inference, using the generalization ($n = 2$) and construction macrorules ($n = 1$) for decision making. One of these is reproduced here:

> Input text sentence: The winning team is the one that completes a catch over the furthest distance, with no breakage.
>
> Item 28: Winning the championship depends on only _____ basic criteria. (2017-i-T4)

To answer Item 28 correctly, the test taker had to infer (1) that the largest distance an egg travels in the air and (2) that it does not break are the two criteria based on which the winner can be found, and then construct a macroproposition using the generalization rule. Additional items triggering the use of these more complex decision-making mechanisms necessary for the construction of an integrated representation of the content of an input text are required for an enhanced construct representation in reading ability assessment.

Conclusion

The analysis of the high-stakes advanced level EFL school-leaving examination reading papers selected for the current study revealed a number of construct validity issues that call into question the generalizability and predictive validity of the results of the reading paper and by extension that of the entire examination—especially if the findings published on the Use of English paper are also taken into consideration (Tankó & Andréka, 2021).

Based on the main problems found in terms of construct underrepresentation, it can be concluded that the reading paper samples the construct that it is intended to measure in a markedly narrow way. In addition, the range of operational task types is poor. The items basically only test literal comprehension and disproportionately target local comprehension, making ineffective use of context. The majority of the items primarily require the use of reading skills rather than strategies or the use of inference or expeditious local reading behaviors. It is to be noted that the official test specification addresses all of these constructs in more or less detail.

The reading paper was also found to be lacking in terms of construct-irrelevant variance. Construct-irrelevant difficulty was induced partly by a lack of consistency in terms of overall length of the reading text (i.e., input text and items) due to specific task types not used in each paper. This resulted in a lack of equivalence between test forms and compromised consistency across different groups of test-takers.

The linguistic and non-linguistic weaknesses of the task context generated by the instructions, images, and titles of the reading tasks were additional sources of construct-irrelevant difficulty. The fact that a quarter of the test task items (i.e., multiple-choice items or sentences to be matched to the text) contained C1 and C2 level lexis which was likely necessary for the successful completion of the tasks were further causes of construct-irrelevant difficulty. Finally, the substantially lower CEFR level of some of the input texts than the level intended to be measured by the examination

resulted in both construct-irrelevant easiness—due to inappropriately easy input text selection during the development of these assessment tasks—and to construct-irrelevant difficulty induced by the exceedingly difficult test task items written to counterbalance the low difficulty level of the input. The disconcertingly low cut score established for the paper further aggravates the problems, and in spite of its apparent beneficial consequences to the test takers, it affects secondary school leavers—and thus potential English major students—negatively.

The direct effect of the school leaving examination on those secondary school leavers who become English majors is that, contrary to the goals set for the two-level school-leaving examination system, it fails to aid them in the pursuit of their tertiary level studies. Specifically in terms of reading ability, it does not benefit—as it should—only those prospective university students who are able to access information efficiently for study purposes. At university, students have to read long, complex texts and combine content extracted from these texts across paragraphs and texts by using their full range of expeditious and careful reading behaviors both globally and locally (see Tankó, 2019; Weir et al., 2000). The indirect effect of the school-leaving examination on the same stakeholder group is that while it may facilitate their mobility for study purposes or entry into the labor market through certification, because of its low generalizability and predictive validity it ultimately does a disservice to those who do not have the functional competencies required by either of these domains.

The limitations of the study are that the CEFR level of each test task item could not be determined with the right level of accuracy because an estimated CEFR level was not available in the English Vocabulary Profile database for every lexical unit used in the analyzed test items. Moreover, more pertinent insights could be gained with actual performance data and information about test taker's comprehension processes. With the above limitations considered, a crucial practical outcome of the study is that it will be easier to improve the examination with the help of the now

identified and explained shortcomings of the specifications and operational reading papers.

Note: This study was conducted as a pilot study for the research project (K 142536) financed by the Hungarian Scientific Research Fund (OTKA).

References

Alderson, J. C. (2000). *Assessing reading*. Cambridge University Press.

Alderson, J. C., & Csereszнyés, M. (2003). *Into Europe: Reading and use of English*. Teleki László Foundation.

Bachman, L., & Palmer, A. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.

Bell, A. (1998). The discourse structure of news stories. In A. Bell & P. Garrett (Eds.), *Approaches to media discourse* (pp. 64–104). Blackwell Publishers.

Bock, M. (1980). Some effects of titles on building and recalling text structures. *Discourse Processes*, *3*(4), 301–311. https://doi.org/10.1080/01638538009544494

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning & Verbal Behavior*, *11*, 717–726. https://doi.org/10.1016/S0022-5371(72)80006-9

Cambridge University Press & Assessment (2015). *EnglishProfile—The CEFR for English: English vocabulary profile online:* [Software]. https://www.englishprofile.org/wordlists/evp

Council of Europe (2020). *Common European framework of reference for languages: Learning, teaching, assessment. Companion volume.* Council of Europe.

Dávid, G. (2008). *Az emelt szintű idegen nyelvi érettségi és az államilag elismert nyelvvizsgák a vizsgázói teljesítmények tükrében: Összegző tanulmány, bővített változat* [The advanced level foreign language school-leaving examinations and the state accredited language examinations in the light of the candidates' performance: A synthesis study—extended version]. Nyelvvizsgáztatási Akkreditációs Központ.

Educational Department (2022a). *Elmúlt évek felvételi statisztikái* [University admission statistics of recent years]. https://www.felvi.hu/felveteli/ponthatarok_statisztikak/elmult_evek/!ElmultEvek/index.php/elmult_evek_statisztikai/

Educational Department (2022b). *Statisztikák, vizsgaeredmények* [Statistical data and school leaving examination results]. https://www.ketszintu.hu/publicstat.php

Educational Department (2022c). *Felsőoktatási felvételi tájékoztató - 2022. szeptemberben induló képzések* [Higher education admission bulletin for September 2022]. https://www.felvi.hu/felveteli/jelentkezes/korabbi_elj_archivum/felveteli_tajekoztatok/FFT_2022A

Fulcher, G. (2010). *Practical language testing* (1st ed.). Routledge.

Graesser, A. C., & Kreuz, R. J. (1993). A theory of inference generation during text comprehension. *Discourse Processes*, *16*(1–2), 145–160. https://doi.org/10.1080/01638539309544833

Gray, W. S. (1960). The major aspects of reading. In H. Robinson (Ed.), *Sequential development of reading abilities* (Vol. 90, pp. 8–24). Chicago University Press.

Guo, D., Zhang, S., Wright, K. L., & McTigue, E. M. (2020). Do you get the picture? A meta-analysis of the effect of graphics on reading comprehension. *AERA Open*, *6*(1). https://doi.org/10.1177/2332858420901696

Hungarian Central Statistical Office (2022). 23.1.1.16. *Idegen nyelvet tanulók a középfokú iskolákban* [Foreign language learners in secondary schools]. https://www.ksh.hu/stadat_files/okt/hu/okt0016.html?utm_source=kshhu&utm_medium=banner&utm_campaign=theme-oktatas

Hungarian Government (2020). *A Kormány 5/2020. (I. 31.) Korm. Rendelete a Nemzeti alaptanterv kiadásáról, bevezetéséről és alkalmazásáról szóló 110/2012. (VI. 4.) Korm. rendelet módosításáról* [Government decree 5/2020. (I.31.) on the modification of Government decree 1010/2012. (VI.4.) on the issuing, introduction and application of the new National Curriculum]. https://magyarkozlony.hu/dokumentumok/3288b6548a740b9c8daf918a399a0bed1985db0f/megtekintes

Hungarian Government (2021). *A Kormány 65/2021. (II. 15.) Korm. Rendelete egyes köznevelési és szakképzési tárgyú, valamint a jelnyelvoktatói névjegyzék vezetésével összefüggő kormányrendeletek módosításáról* [Government Decree amending the government decrees on public education, vocational training and management of the list of sign language teachers]. Magyar Közlöny, 2021. február 15.

Hungarian Government (2022). A Kormány 100/1997. (VI. 13.) *Korm. rendelete az érettségi vizsga vizsgaszabályzatának kiadásáról* [Government Decree 100/1997. (VI. 13.) on the issuance of the examination regulations for the school leaving examination]. https://njt.hu/jogszabaly/1997-100-20-22

Illés, É. (2011). A szövegértés pragmatikája: Egy érettségi feladat elemzése [The pragmatics of text comprehension: The analysis of a school-leaving examination task]. *Iskolakultúra*, *21*(4–5), 144–156. http://www.iskolakultura.hu/index.php/iskolakultura/article/view/21150

Jaeger, A. J., & Wiley, J. (2014). Do illustrations help or harm metacomprehension accuracy? *Learning & Instruction*, *34*, 58–73. https://doi.org/10.1016/j.learninstruc.2014.08.002

Khalifa, H., & Weir, C. J. (2009). *Examining reading: Research and practice in assessing second language reading*. Cambridge University Press.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*(2), 163–182. https://doi.org/10.1037/0033-295X.95.2.163

Kintsch, W. (1993). Information accretion and reduction in text processing: Inferences. *Discourse Processes*, *16*, 193–202. https://doi.org/10.1080/01638539309544837

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363–394. https://doi.org/10.1037/0033-295X.85.5.363

Kirkland, M. R., & Saunders, M. A. (1991). Maximizing student performance in summary writing: Managing cognitive load. *TESOL Quarterly*, *25*(1), 105–121. https://doi.org/10.2307/3587030

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174. https://doi.org/10.2307/2529310

Lenzner, A., Schnotz, W., & Müller, A. (2013). The role of decorative pictures in learning. *Instructional Science*, *41*, 811–831. http://www.jstor.org/stable/43575400

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, *99*(3), 440–466. https://doi.org/10.1037/0033-295x.99.3.440

Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, *35*(11), 1012–1027. https://doi.org/10.1037/0003-066X.35.11.1012

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Meyer, B. J. F. (1982). Signalling the structure of text. In D. H. Jonassen (Ed.), *The technology of text* (pp. 64–89). Educational Technology Publications.

Miller, L. M., Cohen, J. A., & Wingfield, A. (2006). Contextual knowledge reduces demands on working memory during reading. *Memory and Cognition*, *34*(6), 1355–1367. https://doi.org/10.3758/bf03193277

Ministry of Education (2002). *40/2002. (V. 24.) OM rendelet az érettségi vizsga részletes követelményeiről* [Ministry of Education Decree on the detailed requirements of the secondary school leaving examination]. https://njt.hu/jogszabaly/2002-40-20-45

Schmalhofer, F., McDaniel, M. A., & Keefe, D. (2002). A unified model for predictive and bridging inferences. *Discourse Processes*, *33*(2), 105–132. https://doi.org/10.1207/S15326950DP3302_01

Singer, M. (1994). Discourse inference processes. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 479–515). Academic Press.

Singer, M., & Lea, R. B. (2012). Inference and reasoning in discourse comprehension. In Schmid, H.-J. & Geeraerts, D. (Eds.), *Handbook of cognitive pragmatics* (pp. 85–119). Mouton de Gruyter.

Smith, E. E., & Swinney, D. (1992). The role of schemas in reading text: A real-time examination. *Discourse Processes*, *15*(3), 303–316. https://doi.org/10.1080/01638539209544814

Soederberg Miller, L. M., & Stine-Morrow, E. A. L. (1998) Aging and the effects of knowledge on on-line reading strategies. *Journal of Gerontology: Psychological Sciences*, *53B*(4), 223–233. https://doi.org/10.1093/geronb/53B.4.P223

Szabó, G., & Kiszely, Z. (2010). Államilag elismert nyelvvizsgarendszerek, illetve az emelt szintű érettségi összevetése próbavizsgázói teljesítmények tükrében német és angol nyelvből [A comparison of the state accredited language examination systems and the advanced level school-leaving examinations in German and English in the light of mock test performance.]. *Modern Nyelvoktatás*, *16*(4), 19–38.

Szabó, G. (2019). A szövegnehézség vizsgálata az emelt szintű angol érettségi olvasáskomponensében [Examining text difficulty in the reading component of the advanced level school-leaving exam in English]. *Modern Nyelvoktatás*, *25*(3–4), 102–119.

Tan, S., E., M, K. L. E., & O' Halloran, K. (2012). *Multimodal analysis image* [Software]. Multimodal Analysis Company. https://multimodal-analysis.com/products/multimodal-analysis-image/index.html

Tankó, G. (2005). *Into Europe: The writing handbook*. Teleki László Foundation.

Tankó, G. (2019). *Paraphrasing, summarising and synthesising skills for academic writers: Theory and practice* (Rev. ed.). Eötvös University Press.

Tankó, G. (2022). *Professional writing: The academic context* (Rev. 2nd ed.). Eötvös University Press.

Tankó, G., & Andréka, Z. (2021). Probing the advanced level EFL school-leaving examination: The Use of English paper. In G. Tankó & K. Csizér, (Eds.), *DEAL 2021: Current explorations in English applied linguistics* (pp. 65–105). Eötvös Loránd University. https://edit.elte.hu/xmlui/handle/10831/62555

Uchida, S. (2022). *CVLA: CEFR-based vocabulary level analyzer* (ver. 2.0) [Software]. https://cvla.langedu.jp/

Uchida, S., & Negishi, M. (2018). Assigning CEFR-J levels to English texts based on textual features. *In Y. Tono & H. Isahara (Eds.), Proceedings of the 4th Asia Pacific Corpus Linguistics Conference (APCLC 2018)*, 463–467. APCLA.

Urquhart, A. H., & Weir, C. J. (1998). *Reading in a second language: Process, product and practice*. Longman.

Van Dijk, T. (1980). *Macrostructures: An interdisciplinary study of global structures in discourse interaction and cognition*. Erlbaum.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* Academic Press.

Weblingua (2022). *Textinspector* [Software]. https://textinspector.com/

Weir, C. J., Huizhong, Y., & Yan, J. (2000). *An empirical investigation of the componentiality of L2 reading in English for academic purposes* (Studies in Language Testing, Vol. 12). UCLES/Cambridge University Press.

Wiley, J. (2019). Picture this! Effects of photographs, diagrams, animations, and sketching on learning and beliefs about learning from a geoscience text. *Applied Cognitive Psychology*, *33*(1), 9–19. https://doi.org/10.1002/acp.3495

Wiley, J., & Rayner, K. (2000). Effects of titles on the processing of text and lexically ambiguous words: Evidence from eye movements. *Memory & Cognition, 28*(6), 1011–1021. https://doi.org/10.3758/BF03209349

Appendix

Reading Assessment Task Coding Scheme

TASK CHARACTERISTICS

- Task type [nominal]: *20 task types identified in the test specification (e.g., short answer, gapped summary, multiple choice) and subdivided into nine main types of which one, matching, had 12 subtypes (e.g., banked cloze)*
- Length per task [interval; computer assisted coding]: *total number of words in a complete input text (i.e., reconstructed text with the title also included)*
- Length per paper [interval; computer assisted coding]: *total number of words in all the complete input texts within one paper*
- Number of items per task [interval]: *total number of items in a task*
- Number of items per paper [interval]: *total number of items in a paper*
- CEFR level of input [nominal; computer assisted coding]: *A1–C2 (Uchida, 2022)*
- CEFR level of title [nominal; computer assisted coding]: *A1–C2 (CUP & Assessment 2015; Weblingua, 2022)*
- Image-text intersemiotic sense relations [nominal; computer assisted coding]: *four relationship types (e.g., illustration, contrast) (Tan et al., 2012)*

ITEM CHARACTERISTICS

- Comprehension level [nominal]: *literal / inference (Gray, 1960; Khalifa & Weir, 2009; Schmalhofer et al., 2002; Singer & Lea, 2012)*
- Scope of relationship [nominal]: *global-broad / local-narrow (Bachman & Palmer, 2010; Khalifa & Weir, 2009; Urquhart & Weir, 1998)*
- Reading behavior type engaged–Category A [nominal]: *strategy / skill (Bachman & Palmer, 2010; Khalifa & Weir, 2009; Urquhart & Weir, 1998)*

- Reading behavior type engaged–Category B [nominal]: *expeditious / careful (Khalifa & Weir, 2009)*
- CEFR level of item [nominal; computer assisted coding]: *A1–C2 (CUP & Assessment 2015; Weblingua, 2022)*
- Linguistic decision required by response [nominal]: *8 subtypes (e.g., correct answer can be given based on a semantic or syntactic decision, or a combination of the two is needed; or if a superordinate term is generated) (van Dijk, 1980; van Dijk & Kintsch, 1983)*

3

Multiple and Single-Source Text Integrated Writing: A Comparative Study of Task Characteristics and Composition Processes

Gyula Tankó
tanko.gyula@btk.elte.hu

Abstract

Integrated tasks (IntTs) are prominent in teaching and assessing English for academic purposes. The composition process of multiple-source text IntTs is discourse synthesis. However, there is a marked lack of discussion on it in relation to single-source text IntTs as the current understanding is that it is only elicited by multiple-source text tasks. As this has been challenged by recent research, this study undertook to substantiate the claim that discourse synthesis needs to be reconceptualized. In Phase 1 of the investigation, a guided summary writing task was analyzed, piloted with 28 participants, and the scripts were independently double-coded for content reproduction and macrorule use to document the input-related processes it engaged. In Phase 2, a comparative analysis of the task characteristics and processes engaged by synthesis and guided summary writing tasks was conducted. The results indicate that (i) the two tasks are very similar and engage appreciably matching processes, and (ii) discourse synthesis can also be engaged by a single-source text integrated writing task. An innovative difference found is that it is not the number of source texts but the special-purpose task schema that elicits discourse synthesis. This and the taxonomy of integrated task types proposed in this study are of practical relevance for researchers, teachers, and assessors.

Keywords: discourse synthesis, guided summarization, inter and intra-textual synthesis, integrated academic reading-into-writing task types

DOI: 10.21862/ELTE.DEAL.2022.3.59

Multiple and Single-Source Text Integrated Writing: A Comparative Study of Task Characteristics and Composition Processes

In the fields of teaching (e.g., Chan et al., 2015; Leki & Carson, 1994, 1997; Plakans & Gebril, 2012) and assessing (e.g., British Council et al. 2022; Pearson, 2022; ETS, 2022; ISE, 2022) English for academic purposes (EAP), there has been a distinct shift in preference from integrated tasks to independent tasks given that the former are believed to replicate more effectively the processes engaged by actual target language use domain tasks characteristic of a variety of educational settings, specifically tertiary education. The composition process elicited by integrated tasks is discourse synthesis, the conceptualization of which has evolved through several theoretical and empirical studies (Nelson, 2008; Nelson & King, 2022; Plakans, 2009, 2010, 2013; Spivey, 1984, 1990, 1991; Spivey & King, 1989) that investigated it in relation to integrated task types, the composition processes they elicit, the assessment construct redefinition and scoring problems they pose, assessment task design issues they raise, and their pedagogical implications.

However, in spite of the fact that integrated task types have been investigated in relation to summary writing tasks (e.g., Ascención, 2008), which is a single-source text reading-into-writing task type, discourse synthesis as a process is believed to be elicited only by multiple-source text integrated task types. Recent empirical research evidence on the guided summary writing task—a single-source text reading-into-writing task type—indicates that discourse synthesis needs to be reconceptualized (Tankó, 2021b, 2022b).

Therefore, a comparative analysis of the characteristic features and processes elicited by the classical multiple-source text synthesis writing task and the single-source text reading-into-writing guided summary writing task was conducted. The aim was to investigate whether—contrary to current potentially limited perceptions—the composition process elicited by both of these integrated task types is in fact discourse synthesis. The findings substantiate the claim that discourse

synthesis is not only characteristic of multiple-source text tasks and provide valuable insights into the processes elicited by synthesis and guided summary writing tasks.

Theoretical Background

The Conceptualization of Discourse Synthesis

The conceptualization of discourse synthesis has evolved through several empirical studies. Spivey (1984) coined the term discourse synthesis to describe a process of composition that combines reading comprehension and written production. The tasks that elicited discourse synthesis were relabeled as *hybrid* reading-into-writing tasks (Spivey & King, 1989). Designed for teaching and assessment processes, such tasks intend to replicate as much as possible the characteristics of actual language use tasks from the educational domain so as to guarantee their authenticity (e.g., Cumming et al., 2005; Gebril, 2018; Knoch & Sitajalabhorn, 2013; Plakans, 2013). According to Bachman and Palmer (2010), to achieve authenticity, teaching and assessment task designers must ensure that discourse synthesis tasks engage language learners' and test takers' language ability in the same way as actual target language use tasks do. This is necessary because (1) making meaningful interpretations about the language learners' or test takers' language ability based on their performance elicited with the task and (2) the generalizability of these interpretations to the target language use domain—that is beyond instructional or assessment settings—depend on the extent to which the characteristics of language teaching or assessment tasks correspond to those of target language use tasks. For this reason, discourse synthesis tasks consist of two or more input texts—sometimes delivered through different channels (i.e., aural and visual)—on various aspects of the same or topic (e.g., Knoch & Sitajalabhorn, 2013; Plakans & Gebril, 2013, 2017; Spivey, 1984, Spivey & King, 1989). For example, in Spivey's 1984 study, which investigated how university students with differing comprehension

skills performed on a reading-into-writing synthesis task, the participants were given descriptive texts on the same topic, namely three encyclopedia articles presenting facts about the armadillo. They were instructed to write an expository composition, specifically a report with the rhetorical goal of informing young adults through the integration of factual content from the source texts provided. In terms of propositional characteristics, each source text (i.e., ST1, ST2, and ST3) contained some unique propositions; some propositions were also present in one additional source text (e.g., proposition X was shared by ST1 and ST3, whereas proposition Y was shared by ST1 and ST2); and some occurred in all three source texts. The task visibly intended to replicate an academic writing scenario where writers have to read several source texts on a given topic and produce a source-based piece of writing that presents a synthesis of relevant and related content (e.g., a discursive essay or a review of the literature).

This replication endeavor is also captured by the more recent definitions of integrated writing tasks, that is "test tasks that combine two or more language skills to simulate authentic language-use situations" (Plakans, 2013, p. 1). Knoch and Sitajalabhorn (2013) further defined integrated writing tasks as having these two key features: "(a) the input material needs to include a significant proportion of language and, directly following from this, (b) the task needs to require that the language in the source material is used and transformed to complete the writing task" (p. 304). Both of these characterize source-based writing that students do in university content courses.

Due to the parallel deployment of reading and writing abilities in such tasks, Spivey (1990) described discourse synthesis as a hybrid act of literacy in which the cognitive operations performed during the reading and writing task completion phases, that is engaging in "textual transformations through composing" (p. 265), are mutually affective and cannot be separated easily—or maybe not at all. The process underlying both reading and writing was argued to be meaning making for the purposes of comprehension and composition (Nelson, 2008). Therefore, those engaged in discourse synthesis and henceforth referred to as

discourse synthesis writers are believed to construct meaning—that is mental text representations—during both the reading and writing phases (Spivey, 1991), which is in agreement with mainstream research on reading (Kintsch, 1998, 2009, 2012, 2018) and writing (Bereiter & Scardamalia, 1987; Hayes, 2012).

Language users engaged in the act of discourse synthesis actively construct a new piece of discourse. The condition for this is that the integrated task must authorize generative processes by making possible the creation of a novel configuration of meaning and by enhancing the "writer's own sense of authority in writing the piece" (Spivey, 1990, p. 281). Several attempts have been made at creating taxonomies of integrated tasks in which one of the organizing principles was the extent to which generative processes are required for task completion. Two such notable attempts are that of Plakans (2013) and of Gebril (2018)—however, as shown below, neither of these is adequate.

The classification proposed by Gebril (2018) fails to differentiate systematically between task types, the modalities of the input (e.g., non-verbal visual information, verbal visual information, or a mix), and language abilities or skills required for the completion of the tasks. It therefore contains confusing overlapping categories that render it unsuitable for analytic purposes. More acceptable—but affected by the narrow assessment perspective of the study in which it was proposed—is the taxonomy of integrated task types put forward by Plakans (2013). This taxonomy is a substantial adaptation of the academic writing task taxonomy compiled by Leki and Carson (1997). As a result, two of the categories, text- or content-responsible and stimulus-related tasks, are aptly differentiated based on the number of generative processes the tasks require. However, the third task added by Plakans (2013) does not match the organizing principle used in the case of the first two categories. It features thematically linked integrated writing tasks, and therefore it represents a type of writing task in a test paper whose topic is identical with that of the task(s) in the reading paper preceding the writing paper. In the case of such tasks, the relationship between the tasks cannot only be

stimulus-related or content-responsible, as Plakans (2013) stated, but the reading input may be intended (also) to serve as language input for the writing task. A comprehensive and multi-faceted classification taxonomy that is better suited for the purposes of this study is the one proposed in Figure 1. It differentiates between integrated tasks based on the degree to which generative processes are required for task completion, as well the language skills and the semiotic modes involved.

The analytic description of the process of discourse construction differentiated the cognitive operations of *organization*, *selection*, and *connection*. Discourse synthesis requires writes to select ideas from several input texts, reorganize them, and establish new connections between them through integration across input texts and the use of their background knowledge (Nelson, 2008; Spivey, 1984, 1990). The ideas selected for discourse synthesis were found to vary in terms of levels of importance, that is, according to their hierarchical position in the text base, namely the "sequence of propositions expressed by the whole sentence sequence of a text" (van Dijk, 1980, p. 32)—meaning that the higher position a proposition occupies in the text base, the more likely it is to be relevant for the task—and according to their prominence indicated by the recurrence of the same proposition across several texts (Spivey, 1984).

Cognitive Transformational Operations in Discourse Synthesis

As several studies have discussed in detail (Nelson, 2008; Nelson & King, 2022; Spivey, 1984, 1990; Spivey & King, 1989), there is an interaction between the mental representations derived by means of construction from input texts and those constructed for output texts during the organizing, selecting, and connecting operations. In this section, the brief descriptions of the operations based on these studies are elaborated with additional relevant theoretical and empirical research findings. While performing the *organization operation* in the reading phase, discourse synthesis writers carry out organizational transformations and as a result change the representation of the text meaning as it was intended by the

Figure 1

A New Taxonomy of Source-Based Writing Tasks

I.	Basis of categorization:	*Source use*
	Principle of categorization:	*Degree of reliance on input content*
	Task types:	▪ stimulus-related writing (i.e., in order to complete a task, writers must read the input and use it as a source of inspiration for the topic of their composition, but the input content does not have to be reproduced either partially or fully in their written product—it only serves as a "springboard" for writing, see Leki & Carson, 1997, p. 41); Microskills*: — ▪ text or content-responsible writing (i.e., in order to complete a task, writers must both read and provide evidence that they have understood the input; they must base their written product "on content acquired primarily from text", see Leki & Carson, 1997, p. 41); Microskills*: 1, 2, 3.1, 3.2
II.	Basis of categorization:	*Language skills*
	Principle of categorization:	*Types of macro and micro language skills engaged by the task*
	Task types:	▪ reading-into-writing ▪ listening-into-writing ▪ reading and listening-into-writing
III.	Basis of categorization:	*Form of the input*
	Principle of categorization:	*Semiotic modes*
	Task types:	▪ non-verbal visual input (e.g., picture description) ▪ verbal visual input (e.g., global summary or guided summary, see Tankó, 2022a) ▪ non-verbal and verbal visual input (e.g., graph description task)

*Microskills: 1. Note-taking; 2. Direct quotation; 3. Indirect quotation: 3.1 Summarization, 3.2 Paraphrasing

author. This was addressed also by Widdowson (1984), who considered the reading process to be an instance of reader interaction with the text controlled by the reading goal. According to him, depending on their reading goals, readers can adopt a specific stance towards the text, namely a reader is "free to take up whatever position suits his purpose on the dominance/dependence scale" (p. 91). Consequently, readers can be positioned along a cline with submissive and assertive positions at the two extremes. Submissive readers decode the message as intended by the writer, maintaining the hierarchical structure of the text base. These readers engage especially in global careful reading (Urquhart & Weir, 1998). Assertive readers, however, have their own well-defined reading aims and process a text with respect to these aims, so they can disregard the hierarchical structure of the text base: a low-ranked microproposition from the text author's point of view may become a seminal proposition for assertive readers due to their individual reading goals. This has been confirmed by Newton et al. (2018) as well as by Robinson (1987), who stated that:

> (a)n assertive reader brings his own purpose to the reading context and seeks to dominate the writer by interpreting with reference to the terms and conditions of this purpose. The dominant reader reconstructs only that aspect of the writer's overall intention which satisfies his purpose. (p. 91).

Tankó (2021b, 2022b) found that even those readers who assumed the dominant reader stance first carefully read the entire source text provided in a guided summary writing task: An informed dominant reader stance therefore depends on the careful global comprehension of the source text, which requires the reader to assume a submissive reader stance first.

Reading goals that require discourse synthesis writers to assume an assertive reader stance are set not by tasks that activate conventional genre schemas generating well-defined and predictable mental representations (e.g., taking the form of a Who? What? When? Where? Why? How? for a

newspaper article recounting an event, see Bell, 1998; or the IMRD superstructure for a research article, see Swales, 1990), but by tasks that activate "special-purpose schema[s]" (Kintsch & van Dijk, 1978, p. 373) which override canonical superstructures and generate unique mental representations. As a consequence, much the same as readers of inexpertly written input texts do (Kintsch & van Dijk, 1978), discourse synthesis writers engage in organizational transformations that result in task-specific, highly individual representations of the input texts according to their purposes (Lorch & van den Broek, 1997; Nelson, 2008).

While performing the organization operation in the writing phase, discourse synthesis writers are guided by their mental representation of the text they intend to write and generate new relations between the ideas derived from the source texts. As Spivey (1990) noted, the content organization required for discourse synthesis is determined by both the reading and writing processes—production is therefore determined by reception, as also confirmed by Tankó (2021b, 2022b).

In the course of the *selection operations* performed during the reading phase, the decisions of discourse synthesis writers can be guided by patterns of textual organization (e.g., the problem-solution pattern; see Hoey, 2001) as corroborated by empirical research evidence (Johns, 1988), and by the position held by propositions in the text base. However, both during reading and writing, the selection decisions of discourse synthesis writers are informed by one or more principles of relevance—referred to by van Dijk (1979) as differential relevance, "because it differentially selects items for 'special treatment' from among similar items (i.e., items on the same level)" (p. 118)—determined primarily not by textual organization and hierarchy considerations but by the structure of the emerging text (Spivey, 1984, 1990) controlled by the reading goal (i.e., the task schema) set by the task instruction (Kintsch & van Dijk, 1978; Tankó 2021b). Spivey (1990) noted that the formulation of such discourse goals that guide the textual transformation processes in discourse synthesis depends on writers' task management ability.

The *connection operations* performed during the reading phase allow discourse synthesis writers to integrate the content extracted from the source text with their own prior knowledge, a construct subsuming world, topic, and discourse knowledge such as perceptiveness of text structure (Spivey, 1990; Spivey & King, 1989), in order to form what is known as a situation model of text representation (see Kintsch, 2004; Perrig & Kintsch, 1985; Singer & Leon, 2007; van Dijk & Kintsch, 1983). Perceptiveness of text structure and task management ability were found to improve with cognitive development (Nelson, 2008) and writing skills training (Nelson, 2008; Plakans, 2010; Szűcs, 2020).

Discourse synthesis writers generate content in a number of ways: by inventing new content which is new "in terms of the way content was shaped and positioned" (Spivey, 1990, p. 280), by connecting ideas derived from the source texts in a novel way, and by inferring new compressed content. When they integrate source-text content with their own prior knowledge, the connections they create allow them to infer macropropositions with the use of the zero, deletion, generalization, or construction macrorules (see van Dijk, 1980).

Purpose of the Present Study

Based on the analysis of the characteristics of integrated reading-into-writing tasks and the nature of the composing processes they elicit, the present study aimed to investigate whether the guided summary writing task is in fact a type of discourse synthesis task that—contrary to current, potentially limited perceptions of discourse synthesis—integrates content not from multiple source texts but from a single source text. In order to achieve this aim, the research questions formulated for this study were as follows:

1. To what extent does the single-source text integrated guided summary writing task share the characteristic features of a multiple-source text integrated synthesis writing task?

2. To what extent is discourse synthesis, the process of composition required for the completion of a multiple-source text integrated synthesis writing task, also engaged by the single-source text integrated guided summary writing task?

Method

The study was conducted in two phases. In the first phase, a guided summary writing task developed for teaching/learning purposes but reproducing the key features of live guided summary writing assessment tasks—except for source text length and time constraints—was designed and piloted. The aim of the pilot was to ascertain whether the guiding statement incorporated in the task instruction activated the principles of relevance for the selection decisions which the task actually intended to activate. Consequently, the guided summary writing task was administered to a group of students whose language proficiency, background knowledge profile, and language use domain characteristics matched that of the target group and had completed an academic skills course in which they practiced academic reading, summarizing, paraphrasing, and writing. The summary scripts were analyzed by two independent coders for content point inclusion and, as a corollary, given that each content point was connected to the application of one or more macrorules, for macrorule use.

In the second phase, a comparative analysis was conducted. The characteristics of classic discourse synthesis tasks (i.e., those originally designed by Spivey, 1984, 1991; Spivey & King 1989) and guided summary writing tasks (Tankó, 2019, 2022a), as well as the processes activated by the two tasks were compared systematically. For this purpose, an analytical framework was designed based on (i) the task features and the composition processes required for the completion of the discourse synthesis and guided summary writing tasks reported in the body of theoretical and empirical works published on reading-into-writing task

types, and (ii) the framework of language task characteristics (Bachman & Palmer, 2010)—elaborated with (iii) the comprehensive and multi-faceted source-based writing task taxonomy proposed in this study (see Figure 1). The subsequent sections provide more details about these two phases.

Phase One

Instrument

In this section, first the guided summary writing task type is described generically. Then the actual guided summary writing task type which was designed, piloted, and used in this study is presented.

Characteristic Features of the Guided Summary Writing Assessment Task Type

The integrated reading-into-writing task investigated in this study is a reader-based (Hidi & Anderson, 1986) guided summarization task. Tankó (2022a) explained the difference between a complete or global source text and a selective or guided source text summary the following way:

> Writers may have to read complete texts (e.g., an article or a book) or parts of a larger text (e.g., a chapter) and summarize all the main ideas from them. In such cases, writers produce complete source text summaries. However, there are cases when writers are instructed or would like to extract only specific ideas from complete texts or parts of a larger text. In such cases, they write guided summaries. The difference, therefore, between a complete source text summary and a guided summary is that whereas for the first type all the main ideas of a text need to be extracted, for the second only specific ideas have to be extracted and written up in the form of a summary.
>
> Both types of summary writing processes are question driven. When writing a complete source text summary, the writer asks the

> questions: *What are the main ideas in this text?* and *What are the supporting details of the main ideas?* [...] Writers who write a guided summary use the following questions: *What are those (main) ideas in this text that are related to topic X / to aspect N of topic X?* and *What are the supporting details of the (main) ideas related to topic X / to aspect N of topic X?* (p. 119)

The task was designed to reflect changes in the field of academic English skills development in tertiary education, where integrated tasks have been reported to occur as course assignments with increasing frequency (Chan et al., 2015; Leki & Carson, 1994, 1997; Plakans & Gebril, 2012, 2017). This is most likely due to the knowledge-transforming feature (Bereiter et al. 1988) of the language use activities characteristic of the domain. Furthermore, the task also aimed to reflect the universal changeover to integrated language assessment tasks that has taken place in EAP assessment in all the major international academic English tests (e.g., IELTS Academic, British Council et al., 2022; Pearson Test of English Academic, Pearson, 2022; TOEFL iBT and the revised Paper-delivered Test, ETS, 2022; or Trinity College London's Integrated Skills in English test, ISE, 2022).

According to the test specification for stakeholders (Tankó, 2021a), the guided summary writing task intends to engage test takers' English academic reading, note-taking, summarizing, paraphrasing, argumentation, and writing skills in order to make possible the measurement of the students' ability to use English at a high level of proficiency in formal academic language use settings. Test takers have 60 minutes to complete the task without the use of dictionaries or any other reference materials or electronic devices.

The source text is an approximately 700-word-long reading passage on a general academic topic (e.g., using dictation in the language classroom) that discusses several aspects of the topic in varying details, in various parts of the reading passage, and with occasional repetitions. Test takers must find and summarize five or six thematically related aspects by

responding to a guiding question prompted by the instruction. They do not summarize the whole text (i.e., do not write a global summary) but only those parts that contain propositions that answer the guiding question (i.e., write a guided summary). The summary must be written up in the form of a complete paragraph of about 130 words in a neutral, semi-formal, or formal style typical of popular science magazines.

The process of task completion engages academic reading and writing in addition to a range of related academic skills, such as note-taking, summarizing, paraphrasing, or the use of academic register. It therefore integrates the tasks of reading an academic text globally and selectively, extracting and summarizing specific propositional content, rewording the summarized content, and writing it up in the form of a short academic text. The available empirical research evidence demonstrates that the guided summary writing task does in fact engage the above-mentioned language skills (Szűcs, 2020; Tankó, 2021b, 2022b).

The Guided Summary Writing Task Used in this Study

The guided summary writing task designed and piloted for this study was developed according to the Academic Skills Test specification for item writers (Tankó, 2011). The reading passage (see Appendix A) is a self-contained excerpt selected from E. M. Forster's *Aspects of the novel* (1956). The entire input text is 1,224 words long and contains several distinct sections distributed across the text with propositional content relevant for the task, which altogether comprise 440 words. Five content points were identified in the reading passage during task design with the guiding question: *Why are flat characters of use to the novelist?* The first of these content points represents a description. The text type of the remaining four is argumentation, and for the purposes of a finer grained analysis these were split up into the components of claim and supporting evidence. The content points intended to be elicited with the guiding question that operationalized the principles of relevance controlling the selection decisions for this task and the macrorules (MRs; van Dijk, 1980) to be

applied to the extracted propositional content are presented in Appendix B.

Administration of the Guided Summary Writing Task

The guided summary writing task designed for this study was piloted with 28 English majors who had completed the academic skills course. The task was administered in class, and the use of a dictionary or any other reference materials was not allowed. The students were given 90 minutes to work, but all of them handed in their summaries in less than 70 minutes.

Data Obtained from the Piloting of the Guided Summary Writing Task

The students' summaries were transcribed and then coded independently by two trained analysts in order to identify the content points included. The agreement between the two coders was calculated as a percentage value and was high (91%). In the case of each disagreement, the mismatching coding decisions were discussed, and a consensus code was recorded before the analysis of the data.

Phase Two

For the second phase of the study, an analytical framework consisting of two subsections was constructed (see Appendix C). In the first, the points of comparison were the discourse synthesis and guided summary writing task characteristics. In the second subsection, the comparison was guided by the characteristics of the composing processes elicited by the multiple and single-source text task types investigated. The results of the analyses conducted in Phases one and two of this study and their discussion are presented in the following section.

Results and Discussion

In this section, first the results of the piloting of the guided summary writing task are discussed, and then—with the help of the worked task

included in Appendix A—the characteristics of the two task types investigated and the composing processes they elicit are compared.

Guided Summary Writing Task Pilot

As shown in Table 1, although the number of content points (CPs) included varied to a large degree—whereas CP3A was included by all the students, CP4A only occurred in five summaries—overall the students identified and included each content point intended to be selectively extracted from the source text based on the guiding question. The close number of occurrences of the content points subdivided into the claim ("A") and supporting evidence ("B") components (see Appendix A) most likely indicates that the students recognized the claim—support organizational pattern of the task-relevant propositional content in the input and reproduced it in their summaries accordingly.

Table 1

Content Points Included in the Guided Summaries and the Macrorules Used

	GUS	Macrorule*	%
CP1	11	SEL	39
CP2A	25	SEL	89
CP2B	28	GEN	100
CP3A	25	SEL	89
CP3B	23	GEN	82
CP4A	5	SEL	18
CP4B	6	SEL	21
CP5A	12	SEL	43
CP5B	11	SEL	39

*SEL = selection, GEN = generalization

Based on the macrorules associated with the content points, without an analysis of the actual macrorule use quality, it can be conjectured that

summary writers used the selection and generalization rules. A discussion of the level of skill required for the competent application of the macrorules is not necessary for this analysis. What is important for the comparison of the composition processes elicited by the guided summary writing and discourse synthesis tasks is that during the completion of the guided summary writing task, students engaged not only in content selection and restructuring but also compressing.

As van Dijk (1980) pointed out, the deletion rule can be considered to be a selection rule: "In a more positive sense, then, the same rule may be taken as a SELECTION rule, which selects from a text base all propositions which are interpretation conditions (presuppositions) of other propositions in the text base" (p. 47). As the students selected specific propositional content for the summary, they also deleted irrelevant propositional content. Moreover, they also used their prior knowledge to compress meaning and in this way constructed macropropositions and generalizations—in this case from details. To do this, they had to recognize that certain propositions are semantically connected and that through inference a superordinate proposition could be abstracted. Specifically, students had to delete three of the four instances of CP1 ($f = 4$) (see Appendix A) and generate one content point from CP2A ($f = 4$) and CP2B ($f = 4$). Therefore, based on the above discussion of content points and macrorules, it can be concluded that each content point was identified by the summary writers in the source text and included in the summaries with the use of the selection and generalization macrorules.

Task Type Characteristics

A comparison of the characteristic features of discourse synthesis and guided summary writing task types indicates that the two are very similar despite the conspicuous difference in the number of source texts.

Both task types are integrated given that, unlike independent writing tasks, they combine reading comprehension and written

production (Cumming et al., 2005). The lengthy verbal input material in both tasks must be comprehended and transformed in order to complete the tasks (Spivey, 1984, 1991; Tankó, 2019). Both types have been used as teaching and assessment tasks (e.g., British Council et al., 2022; ETS, 2022; Pearson, 2022; Tankó 2019, 2022a), and both elicit an extended production response (Spivey, 1984, 1991; Tankó, 2019, 2021b). Writers are expected to generate well-organized, self-contained, stand-alone, and complete pieces of discourse that are structurally not isomorphic with the source text and have distinct functions in new contexts (Nelson & King, 2022; Tankó, 2019). In order to complete either task type, writers must rely substantially on the input content, so content-responsible writing is required by both (Spivey, 1990; Tankó, 2019, 2022a). A further similarity is that two language skills are needed for the completion of each task: reading and writing. The tasks also share the same semiotic mode regarding input: writers have to process verbal visual input in both cases. One key difference between the tasks, however, is the number of source texts provided as input, which ranged from two (Spivey, 1991) to three (Spivey, 1984; Spivey & King, 1989) in the case of discourse synthesis tasks, whereas there is only one source text in the guided summary writing task.

Both task types are designed to simulate and replicate as much as possible the characteristics of real life language use tasks typically occurring in the educational domain (Plakans, 2013; Stemmer, 2019; Tankó, 2020). Furthermore, both require prior experience with the task types and an understanding of the functions they fulfil in the educational domain (Nelson & King, 2022; Tankó, 2019). In terms of the topical characteristics of the input, the two task types are also rather similar in that the source texts in discourse synthesis tasks either all focus on exactly the same topic (Spivey, 1984; Spivey & King, 1989) or on closely related topics; for example, each of the two texts deals with a mollusc subspecies (Spivey, 1991). The input text in a guided summary writing task also provides closely related propositional content, or thematic aspects, on one specific topic but within one source text. However, irrespective of whether the propositional content relevant to the task is provided in one or more

source texts, its features and potential patterns of occurrence within and across texts are identical in both tasks. Each source text in the case of a discourse synthesis task (Spivey, 1990; Spivey & King, 1989) and each text segment including a content point in the case of a guided summary writing task can feature unique propositional content (e.g., CP3A or CP3B in the sample summary, see Appendix A). Moreover, just as the same propositional content can occur repeatedly in more than one source text in a discourse synthesis task (see Spivey, 1990; Spivey & King, 1989), as the analysis of the sample guided summary writing task revealed, the same content point can also recur in the source text (e.g., whereas CP2A occurs in two, both CP1 and CP2B occur in four text segments, see Appendix A). As discussed in the next section, the skilful manipulation of the propositional content requires the use of almost identical processes in the case of both task types.

Task Completion Process Features

The majority of the processes required for the completion of discourse synthesis are also present in guided summary writing task types. The fact that the two task types require that these processes be applied across different numbers of source texts is a formal one.

The interpretation of the discourse synthesis and guided summary writing task schemas is reported to be a notably more complex process than in the case of independent writing tasks or conventional genre schemata. Writers of both integrated task types re-read the instructions several times in order to understand what the task was (Plakans, 2010; Tankó, 2022b), how they were supposed to complete it (e.g., avoid plagiarism or monitor the process of synthesis, see Plakans, 2010; Tankó, 2019, 2022b), and what their written product was supposed to be like in terms of rhetorical function and genre (e.g., an informative report versus a stand-alone argumentative guided summary). Furthermore, in the case of the guided summary writing task, students had to formulate a guiding question on the basis of the instruction (Tankó, 2021b). For these reasons,

the creation of a task representation and the interpretation of task demands are equally complex and taxing processes in the case of both task types.

As both Plakans (2010) and Szűcs (2020) demonstrated, inexperienced and experienced writers approached each task type in markedly different ways and used different composition processes, which resulted in substantially different written products. Therefore, prior experience with these task types, including explicit instruction, is necessary so that writers develop the appropriate task management abilities. They must understand, for example, what amount of input is required and how that input needs to be processed for the expected response. The scope of the relationship in the case of both tasks is both broad and narrow. Writers must first read the input text(s) entirely and subsequently narrow the range of input to be processed for their written products to the task relevant propositional content only. This is done by consecutively assuming the submissive and authoritative assertive reader position. Tankó (2021b) found that some summary writers managed to deploy careful global and selective reading processes simultaneously.

Furthermore, writers also need to be aware that they are engaged in meaning construction both during reading and writing. During reading, they first construct a global representation of the source text(s). Following this, while composing the written product, they select the relevant propositional content and restructure it according to the requirements of the expected written product. The fact that writers construct mental text representations, often by integrating source text content with their prior knowledge resulting in "novel configuration of meaning" (Spivey, 1990, p. 281) during both the reading and writing phases in the case of both tasks has been documented in several studies (Plakans, 2009; Plakans et al., 2018; Szűcs, 2020; Tankó, 2021b). Also well documented is the difficulty to separate the reading and composing phases of task completion (Spivey, 1990; Tankó, 2021b, 2022b). The two mental representation generating processes overlap and affect one another as writers engage in recursive composition processes in the case of both task types.

The cognitive operations of organization, selection, and connection have been thoroughly described in the body of literature available on discourse synthesis, and ample empirical evidence is available on how they are employed in the course of the completion of integrated tasks (Cumming et al., 2005; Plakans, 2009; Spivey, 1984, 1990, 1991; Spivey & King, 1989). When engaged in organization, both discourse synthesis (e.g., Plakans, 2009; Spivey, 1990) and guided summary writers (Tankó, 2021b, 2022b) structure the content of the input texts they read, and their reading processes are influenced by the composing processes, namely the structure of the emerging product they are writing. Their selection processes are also determined by the task schema that sets a selective reading goal in the case of both task types. Writers assume an assertive reader position and select propositional content relevant for their expected products, and in so doing change the representation of the text meaning as it was intended by the author of each source text (Spivey, 1990; Tankó, 2021b, 2022b).

However, the relevance principles used in the case of the two task types are somewhat different. Whereas discourse synthesis writers select propositional content on the basis of propositional prominence determined by the recurrence and by the position that a proposition occupies in the text base, guided summary writers select propositional content based on differential relevance (van Dijk, 1979) as determined by the task instruction (Kintsch & van Dijk, 1978). Depending on the macrostructure of the source text, the writers of either task type may have to select one instance of recurring propositional content (e.g., in the case of CP1, which is repeated 4 times in the guided summary source text, see Appendix A). However, whereas prominence can be a shared relevance principle, in the case of the guided summary a microproposition—that is a low-ranking proposition in the text base—may be actually included in the summary without any changes, or it may have to be transformed with the generation or construction rules in order to formulate a macroproposition.

When they perform connection operations, both discourse synthesis and guided summary writers construct mental representations by means of inference and elaboration based on the input they read by linking propositional content in the source text with their prior knowledge. They also use their prior knowledge to identify and link propositional content relevant for the task across multiple texts or within a single text, as well as to invent and write up content that is novel in terms of degree of conciseness, rhetorical structure, and language use (Spivey, 1990; Tankó 2021b). In discourse synthesis, writing an informative report requires streamlining the content extracted from the source texts and framing it with an introduction and a conclusion typical of the report genre. Guided summary writers have to invent a topic and a concluding sentence, two summative new macropropositions that are inferred from the meaning and functions of the summarized and paraphrased content points included in the body of the summary (Tankó, 2019).

Finally, the use of macrorules for content processing occurs in the case of both discourse synthesis and guided summary writing (Spivey, 1990; Tankó, 2019, 2021b). Discourse synthesis and guided summary writers both use the deletion/selection macrorule when they identify repeated propositions relevant to the task, as they only include these propositions once in their written products. The same rule is used not only to eliminate redundancy but also to delete irrelevant propositions at all text base levels and unnecessary propositions for other macroprocesses (e.g., construction). When discourse synthesis writers combine source text content with their prior knowledge and infer content for their written products, they engage in the same processes that allow guided summary writers to infer macropropositions with the use of the generalization or construction macrorules.

The results of the comparison of the task features and of the task completion processes characteristic of the two task types indicate that they are markedly similar both in terms of task characteristics and completion processes. Both tasks require students to engage in discourse synthesis in order to complete them. The findings suggest that two types of synthesis,

namely inter- and intra-textual synthesis, should be distinguished. The outcome of the two phases of comparison are summarized in Table 2.

Table 2

Summary of the Comparison of Task Type Characteristics and Task Completion Processes

Criterion	Identical for the discourse synthesis & guided summary writing task types*
Task Type Characteristics	
Task type	✓
Purpose	✓
Type of response	✓
Directness of relationship	✓
Language skills	✓
Input form	✓
Number of source texts	×
Authenticity	✓
Familiarity with academic reading and writing	✓
Topical characteristics of the input	~✓
Propositional content	✓
Task Completion Processes	
Task representation	✓
Task management ability	✓
Scope of relationship	✓
Meaning construction	✓
Task completion phases	✓
Cognitive operation 1: Organization	✓
Cognitive operation 2: Selection	~✓
Cognitive operation 3: Connection	✓
Macrorule use	✓

* ✓- matching, ~✓- similar, × - mismatching task type characteristics or task completion process features

Conclusion

Given the unquestionably increasing importance and frequency of occurrence of integrated tasks in English for EAP instruction and assessment due to the authentic way these tasks are considered to replicate the characteristic features of target language use domain tasks as well as the processes engaged by them, this study undertook to investigate a multiple and single-source text integrated reading-into-writing task type. The aim was to compare the task characteristics and composing processes required for the completion of synthesis and guided summary writing tasks in order to determine whether the process underlying both is discourse synthesis, which earlier had been claimed to only occur in the case of multiple-source text tasks.

The analysis conducted partly on the basis of the novel, comprehensive, and multi-faceted taxonomy of integrated task types proposed in this study revealed that except for one key formal difference, the number of source texts given as input, some topical characteristics of the input, and the type of relevance principles activated by the cognitive operation of selection, the two task types are very similar and the processes required for their completion overlap to a large extent. Consequently, it can be stated that contrary to currently held limited perceptions, discourse synthesis as a composition process can also be engaged by a single-source text reading-into-writing task: we can distinguish here between *inter-textual* and *intra-textual* synthesis. What elicits discourse synthesis is not the number of source texts provided as input but the special-purpose schema set for the task and operationalized with the selective reading goal. This is a seminal difference thus far overlooked in the body of literature on discourse synthesis and integrated tasks. This finding should be of practical relevance for researchers, teachers, and assessors using integrated tasks to analyze, teach and test discourse synthesis. Researchers can conduct analyses informed by a more accurate conceptual definition of discourse synthesis, and EAP instructors can explain more clearly and effectively the composition processes

underlying single and multiple-source text writing tasks when they teach source-based academic writing. Finally, assessors can formulate better construct definitions and measure discourse synthesis more accurately with integrated writing tasks.

References

Ascención, Y. (2008). Investigating the reading-to-write construct. *Journal of English for Academic Purposes*, *7*, 140–150. https://doi.org/10.1016/j.jeap.2008.04.001

Bachman, L., & Palmer, A. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world.* Oxford University Press.

Bell, A. (1998). The discourse structure of news stories. In A. Bell & P. Garrett (Eds.), *Approaches to media discourse* (pp. 64–104). Blackwell Publishers.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Erlbaum.

Bereiter, C., Burtis, P. J., & Scardamalia, M. (1988). Cognitive operations in constructing main points in written composition. *Journal of Memory and Language*, *27*(3), 261–278. https://doi.org/10.1016/0749-596X(88)90054-X

British Council, IDP: IELTS Australia, & Cambridge Assessment English (2022). *Test format*. Retrieved from https://www.ielts.org/for-test-takers/test-format

Chan, S., Inoue, C., & Taylor, L. (2015). Developing rubrics to assess the reading-into-writing skills: A case study. *Assessing Writing*, *26*, 20–37. https://doi.org/10.1016/j.asw.2015.07.004

Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005). Differences in written discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing*, *10*, 5–43. https://doi.org/10.1016/j.asw.2005.02.001

ETS. (2022). *TOEFL iBT Test Content*. Retrieved from https://www.ets.org/toefl/test-takers/ibt/about/content

Forster, E. M. (1956). *Aspects of the novel*. Mariner Books.

Gebril, A. (2018). Integrated-skills assessment. In J. I. Liontas et al. (Eds.), *The TESOL Encyclopedia of English Language Teaching* (pp. 1–7). Wiley-Blackwell. https://doi.org/10.1002/9781118784235.eelt0544

Hayes, J. R. (2012). Modeling and remodeling writing. *Written Communication*, *29*(3), 369–388. https://doi.org/10.1177/0741088312451260

Hidi, S., & Anderson, V. (1986). Producing written summaries: Task demands, cognitive operations, and implications for instruction. *Review of Educational Research*, *56*(4), 473–493. https://doi.org/10.3102/00346543056004473

Hoey, M. (2001). *Textual interaction: An introduction to written discourse analysis.* Routledge.

Johns, A. M. (1988). Reading for summarising: An approach to text orientation and processing. *Reading in a Foreign Language*, *4*(2), 79–90. https://doi.org/10125/67006

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.

Kintsch, W. (2004). The construction-integration model of text comprehension and its implications for instruction. In R. B. Ruddell & N. J. Unrau (Eds.), *Theoretical models and processes of reading* (pp. 1270–1328). International Reading Association.

Kintsch, W. (2009). Learning and constructivism. In S. Tobias & T. M. Duffy (Eds.), *Constructivist instruction: Success or failure?* (pp. 223–241). Routledge/Taylor & Francis Group.

Kintsch, W. (2012). Psychological models of reading comprehension and their implications for assessment. In J. P. Sabatini, E. R. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how we assess reading ability* (pp. 21– 38). Rowman & Littlefield Education.

Kintsch, W. (2018). Revisiting the construction—integration model of text comprehension and its implications for instruction. In Alvermann, D. E., Unrau, N. J., Sailors, M., & Ruddell, R. B. (Eds.), *Theoretical models and processes of literacy* (7th ed., pp. 194–219). Routledge.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363–394. https://doi.org/10.1037/0033-295X.85.5.363

Knoch, U., & Sitajalabhorn, W. (2013). A closer look at integrated writing tasks: Towards a more focused definition for assessment purposes. *Assessing Writing*, *18*(4), 300–308. https://doi.org/10.1016/j.asw.2013.09.003

Leki, I., & Carson, J. G. (1994). Students' perceptions of EAP writing instruction and writing needs across the disciplines. *TESOL Quarterly*, *28*(1), 81–101. https://doi.org/10.2307/3587199

Leki, I., & Carson, J. G. (1997). "Completely different worlds": EAP and the writing experiences of ESL students in university courses. *TESOL Quarterly*, *31*(1), 39–69. https://doi.org/10.2307/3587974

Lorch, R. F., Jr., & van den Broek, P. W. (1997). Understanding reading comprehension: Current and future contributions of cognitive science. *Contemporary Educational Psychology*, *22*(2), 213–246. https://doi.org/10.1006/ceps.1997.0931

Nelson, N. (2008). The reading-writing nexus in discourse research. In C. Bazerman (Ed.), *Handbook of research on writing: History, society, school, individual, text* (pp. 435–450). Taylor & Francis Group/Lawrence Erlbaum Associates.

Nelson, N., & King, J. R. (2022). Discourse synthesis: Textual transformations in writing from sources. *Reading and Writing*. Advance online publication. https://doi.org/10.1007/s11145-021-10243-5

Newton, J., Ferris, D., Grabe, W., Goh, C., Stoller, F. L., & Vandergrift, L. (2018). *Teaching English to second language learners in academic contexts: Reading, writing, listening, speaking*. Routledge.

Pearson. (2022). *PTE Academic*. Retrieved from https://pearsonpte.com/pte-academic/

Perrig, W., & Kintsch, W. (1985). Propositional and situational representations of text. *Journal of Memory and Language*, *24*(5), 503–518. https://doi.org/10.1016/0749-596X(85)90042-7

Plakans, L. (2009). Discourse synthesis in integrated second language writing assessment. *Language Testing*, *26*(4), 561–587. https://doi.org/10.1177/0265532209340192

Plakans, L. (2010). Independent versus integrated writing tasks: A comparison of task representation. *TESOL Quarterly*, *44*(1), 185–194. https://www.jstor.org/stable/27785076

Plakans, L. (2013). Assessment of integrated skills. *Encyclopedia of applied linguistics*. Wiley-Blackwell. https://doi.org/10.1002/9781405198431.wbeal0046.pub2

Plakans, L., & Gebril, A. (2012). A close investigation into source use in integrated second language writing tasks. *Assessing Writing*, *17*(1), 18–34. https://doi.org/10.1016/j.asw.2011.09.002

Plakans, L., & Gebril, A. (2017). Exploring the relationship of organization and connection with scores in integrated writing assessment. *Assessing Writing*, *31*, 98–112. https://doi.org/10.1016/j.asw.2016.08.005

Plakans, L., Liao, J. T., & Wang, F. (2018). Integrated assessment research: Writing-into-reading. *Language Teaching*, *51*(3), 430–434. https://doi.org/10.1017/S0261444818000149

Robinson, P. J. (1987). Reading and coping: Code, communication and the development of second language reading abilities using content area texts. ERIC Doc. ED 294 146; CS 009 129. University of Illinois: Reading and Communication Skills Center. Retrieved from https://archive.org/details/ERIC_ED294146

Singer, M., & Leon, J. (2007). Psychological studies of higher language processes: Behavioral and empirical approaches. In F. Schmalhofer & C. Perfetti (Eds.), *Higher level language processes in the brain* (pp. 9–25). Lawrence Erlbaum.

Spivey, N. N. (1984). *Discourse synthesis: Constructing texts in reading and writing*. (Outstanding Dissertation Monograph Series). International Reading Association.

Spivey, N. N. (1990). Transforming texts: Constructive processes in reading and writing. *Written Communication*, *7*(2), 256–287. https://doi.org/10.1177/0741088390007002004

Spivey, N. N. (1991). The shaping of meaning: Options in writing the comparison. *Research in the Teaching of English*, *25*(4), 390–418. https://www.jstor.org/stable/40171203

Spivey, N. N., & King, J. R. (1989). Readers as writers composing from sources. *Reading Research Quarterly*, *24*(1), 7–26. https://www.jstor.org/stable/748008

Stemmer, G. (2019). *The use of summarisation as an educational tool in university content classes* [Unpublished master's thesis]. Eötvös Loránd University, Budapest.

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

Szűcs, Á. (2020). *Reading English for academic purposes: The reading processes of first-year EFL learner BA students* [Unpublished doctoral dissertation]. Eötvös Loránd University, Budapest.

Tankó, G. (2011). *Guidelines for item writers: The guided summary writing task*. Department of English Applied Linguistics, Eötvös Loránd University.

Tankó, G. (2019*). Paraphrasing, summarising and synthesising skills for academic writers: Theory and practice* (Rev. ed.). Eötvös University Press.

Tankó, G. (2020). Eliciting written academic performance: An investigation of task effect on the written production of EFL students. In C. Kálmán (Ed.), *DEAL 2020: A Snapshot of Diversity in English Applied Linguistics* (pp. 1–37). Eötvös Loránd University.

Tankó, G. (2021a). *Academic skills test: Test specification*. Department of English Applied Linguistics, Eötvös Loránd University.

Tankó, G. (2021b). *Assessing academic reading ability with an integrated reading-into-writing guided summary writing task*. [Unpublished manuscript]. Department of English Applied Linguistics, Eötvös Loránd University.

Tankó, G. (2022a). *Professional writing: The academic context* (2nd Rev. ed.). Eötvös University Press.

Tankó, G. (2022b, June 3–5). *Validating the EAP reading component of the construct operationalised by the guided summary writing task* [Paper presentation]. 18th EALTA Conference, Budapest, Hungary.

Trinity College London. (2022). *ISE– Integrated Skills in English exams*. Retrieved from https://www.trinitycollege.com/qualifications/english-language/ISE

Urquhart, A. H., & Weir, C. J. (1998). *Reading in a second language: Process, product and practice.* Longman.

van Dijk, T. A. (1979). Relevance assignment in discourse comprehension. *Discourse Processes*, *2*(2), 113–126. https://doi.org/10.1080/01638537909544458

van Dijk, T. A. (1980). *Macrostructures: An interdisciplinary study of global structures in discourse, interaction, and cognition.* Erlbaum.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press.

Widdowson, H. G. (1984). Reading and communication. In C. Alderson & A. H. Urquhart (Eds.), *Reading in a foreign language* (pp. 213–230). Longman.

Appendix A
The Worked Guided Summary Writing Task
Used in the Study
(Tankó, 2022, pp. 115–117)

Write a paragraph of 140 words (+/-10%) in which you summarize in your own words as far as possible *the reasons why flat characters are of use to the novelist*, which are discussed in the reading passage below.

The actors in a story*

We may divide characters into flat and round. Flat characters were called "humorous" in the seventeenth century, and are sometimes called types, and sometimes caricatures. In their purest form, they are constructed round a single idea or quality: when there is more than one factor in them, we get the beginning of the curve towards the round. **[**The really flat character can be expressed in one sentence ~~such as "I never will desert Mr. Micawber."~~**] CP1** There is Mrs. Micawber—she says she won't desert Mr. Micawber, she doesn't, and there she is. Or: "I must conceal, even by subterfuges, the poverty of my master's house." There is Caleb Balderstone in *The Bride of Lammermoor*. He does not use the actual phrase, but it completely describes him; he has no existence outside it, no pleasures, none of the private lusts and aches that must complicate the most consistent of servitors. Whatever he does, wherever he goes, whatever lies he tells or plates he breaks, it is to conceal the poverty of his master's house. It is not his *idée fixe,* because there is nothing in him into which the idea can be fixed. He is the idea, and such life as he possesses radiates from its edges and from the scintillations it strikes when other elements in the novel impinge. Or take Proust. There are numerous flat characters in Proust, such as the Princess of Parma, or Legrandin. **[**Each can be expressed in a single sentence~~, the Princess's sentence being, "I must be particularly careful to be kind."~~**] CP1** She does nothing except to be particularly careful, and those of the other characters who are more

complex than herself easily see through the kindness, since it is only a by-product of the carefulness.

{[One great advantage of flat characters is that they are easily recognized whenever they come in—recognized by the reader's emotional eye, ~~not by the visual eye, which merely notes the recurrence of a proper name. In Russian novels, where they so seldom occur, they would be a decided help.~~ It is a convenience for an author when he can strike with his full force at once, and flat characters are very useful to him,**]** **CP2A-CLAIM + [**since they never need reintroducing, never run away, have not to be watched for development, and provide their own atmosphere— ~~little luminous disks of a pre-arranged size, pushed hither and thither like counters across the void or between the stars;~~ most satisfactory.**]}** **CP2B-SUPPORT**

{[A second advantage is that they are easily remembered by the reader afterwards.**]** **CP3A-CLAIM + [**They remain in his mind as unalterable for the reason that they were not changed by circumstances; they moved through circumstances, which gives them in retrospect a comforting quality, and preserves them when the book that produced them may decay.**]}** **CP3B-SUPPORT** The Countess in *Evan Harrington* furnishes a good little example here. Let us compare our memories of her with our memories of Becky Sharp. We do not remember what the Countess did or what she passed through. What is clear is her figure and the formula that surrounds it, namely, "Proud as we are of dear papa, we must conceal his memory." All her rich humour proceeds from this. She is a flat character. Becky is round. **[**~~She, too, is on the make,~~ but she cannot be summed up in a single phrase, ~~and we remember her in connection with the great scenes through which she passed and as modified by those scenes—that is to say, we do not remember her so easily because she waxes and wanes and has facets like a human being.~~**]** **CP1** **{[**All of us, even the sophisticated, yearn for permanence, and to the unsophisticated permanence is the chief excuse for a work of art. We all want books to endure, to be refuges, and their inhabitants to be always the same,**]**

CP4B-SUPPORT + [and flat characters tend to justify themselves on this account.**]} CP4A-CLAIM**

All the same, critics who have their eyes fixed severely upon daily life—as were our eyes last week— have very little patience with such renderings of human nature. Queen Victoria, they argue, cannot be summed up in a single sentence, so what excuse remains for Mrs. Micawber? One of our foremost writers, Mr. Norman Douglas, is a critic of this type, and the passage from him which I will quote puts the case against flat characters in a forcible fashion. The passage occurs in an open letter to D. H. Lawrence, with whom he is quarrelling: a doughty pair of combatants, the hardness of whose hitting makes the rest of us feel like a lot of ladies up in a pavilion. He complains that Lawrence, in a biography, has falsified the picture by employing "the novelist's touch," and he goes on to define what this is:

> It consists, I should say, in a failure to realize the complexities of the ordinary human mind; it selects for literary purposes two or three facets of a man or woman, generally the most spectacular, and therefore useful ingredients of their character and disregards all the others. Whatever fails to fit in with these specially chosen traits is eliminated—must be eliminated, for otherwise the description would not hold water. Such and such are the data: everything incompatible with those data has to go by the board. It follows that the novelist's touch argues, often logically, from a wrong premise: it takes what it likes and leaves the rest. The facets may be correct as far as they go but there are too few of them: what the author says may be true and yet by no means the truth. That is the novelist's touch. It falsifies life.

Well, the novelist's touch as thus defined is, of course, bad in biography, for no human being is simple. **{[**But in a novel it has its place: a novel that is at all complex often requires flat people as well as round,**] CP5A-CLAIM + [**and the outcome of their collisions parallels life ~~more~~

~~accurately than Mr. Douglas implies~~.**]}** **CP5B-SUPPORT** The case of Dickens is significant. Dickens' people are nearly all flat (Pip and David Copperfield attempt roundness, but so diffidently that they seem more like bubbles than solids). **{[**Nearly every one can be summed up in a sentence,**]** **CP1 + [**and yet there is this wonderful feeling of human depth.**]}** **CP2B-SUPPORT** Probably the immense vitality of Dickens causes his characters to vibrate a little, so that they borrow his life and appear to lead one of their own. It is a conjuring trick; at any moment we may look at Mr. Pickwick edgeways and find him no thicker than a gramophone record. But we never get the sideway view. Mr. Pickwick is far too adroit and well-trained. He always has the air of weighing something, and when he is put into the cupboard of the young ladies' school he seems as heavy as Falstaff in the buck-basket at Windsor. **{[**~~Part of the genius of~~ Dickens ~~is that he~~ does use types and caricatures, people whom we recognize the instant they re-enter**,]** **CP2A-CLAIM + [**and yet achieves effects that are not mechanical and a vision of humanity that is not shallow.**]}** **CP2B-SUPPORT** Those who dislike Dickens have an excellent case. He ought to be bad. **[**He is actually one of our big writers, and his immense success with types suggests that there may be more in flatness than the severer critics admit.**]}** **CP2B-SUPPORT**

(Forster, 1956)

* The crossed out text indicates those text segments within the CPs to which the deletion macrorule was applied. The generalization macrorule was applied to the segments with wavy underlining.

Appendix B
The Content Points in the Sample Guided Summary Writing Task and the Macrorules to be Applied to Them

CP1 (description): Flat characters can be described very briefly. [MR: Selection]
CP2A (claim): Flat characters can be easily recognized, so they are powerful tools for the writer. [MR: Selection]
CP2B (support): Flat characters are familiar, stable, evoke a specific mood, and are not truly simple. [MR: Generalization + Selection]
CP3A (claim): Flat characters are easily remembered. [MR: Selection]
CP3B (support): Flat characters are stable, consoling, and enduring. [MR: Generalization]
CP4A (claim): Flat characters satisfy an important reader expectation. [MR: Selection]
CP4B (support): Due to their stasis and regularity, flat characters are the safe havens readers of all sophistication levels need. [MR: Generalization + Selection]
CP5A (claim): Flat characters are needed in a complex novel. [MR: Selection]
CP5B (support): Novels with no flat characters lack realism. [MR: Selection]

Appendix C
The Analytical Frameworks Used for Task Analysis and Composition Process Comparison

1. Analytical framework for task analysis

Points of comparison:

- Task type: Integrated vs. independent writing task
- Purpose: language teaching/assessment task
- Type of response: the length and nature of the constructed response
- Directness of the relationship: text- or content-responsible vs. stimulus-related writing in terms of source use
- Language skills: number and type of skills involved
- Input form: one or mixed semiotic modes
- Number of source texts: number of different texts provided in the input
- Authenticity: degree of simulation of actual language-use situations
- Familiarity with academic reading and writing: degree of previous experience required
- Topical characteristics of the input: thematic relatedness of the source texts
- Propositional content: nature and distribution of the propositions in the text base

2. Analytical framework for task completion process analysis

Points of comparison:

- Task representation: task schema and interpretation of task demands
- Task management ability: amount of previous experience with the task type required

- Scope of relationship: amount of input to be processed for the expected response
- Meaning construction: mental text representations constructed during reading and writing
- Task completion phases: separability and mutual dependence of the task completion phases
- Cognitive operation 1: Organization (reading and writing purpose-dependent, individual mental representation of the input texts and of the relevant propositional content according to the text to be written)
- Cognitive operation 2: Selection (reading and writing goal dependent application of relevance principles)
- Cognitive operation 3: Connection (content generation through relating propositional content and previous knowledge by means of inference and elaboration during reading and invention during writing)
- Macrorule use: selecting and constructing (macro)propositions

4

The Development of Syntactic Complexity and Fluency in an Advanced Writing Course

Attila M. Wind
wind.attila@btk.elte.hu

Abstract

Second language (L2) writing development has been extensively researched over the past decades. However, most previous studies focused on how Asian learners improve their L2 writing; therefore, little is known about the Hungarian context. This study investigated the L2 writing development of undergraduates in the Undivided Teaching Training (UTT) and the English and American Studies (EAS) programs at a large university in Budapest. A total of 63 undergraduate students composed two essays: one at the beginning and one at the end of an advanced writing (AW) course. The essays (N = 126) were analyzed using the L2 Syntactic Complexity Analyzer (L2SCA) for syntactic complexity and fluency indices. This study found statistically significant changes in the fluency index in the combined groups' and EAS students' data. A number of pedagogical implications are presented based on the results of the study.

Keywords: second language writing development, advanced writing course, syntactic complexity, fluency

DOI: 10.21862/ELTE.DEAL.2022.4.97

The Development of Syntactic Complexity and Fluency in an Advanced Writing Course

Research on second language (L2) writing development has produced mixed results over the past 15 years. Some studies found that L2 writing improved over a short period of time (Mazgutova & Kormos, 2015), while other studies found that L2 writing did not improve over three years (Knoch et al., 2015). However, there are some consistent results. For example, previous studies found that low-proficiency learners tended to show greater improvements in more areas (Mazgutova & Kormos, 2015; Polat et al., 2020), while high-proficiency learners tended to demonstrate minor improvements in fewer areas of complexity (Knoch et al., 2015; Storch, 2009).

Most of the earlier studies on L2 writing development have tended to focus on Asian leaners of English. For example, Storch (2009) and Knoch et al. (2015) investigated the L2 writing development of Asian learners at an Australian university. Jiang et al. (2019) examined Chinese learners' L2 writing development at a Chinese university, while Mazgutova and Kormos (2015) concentrated on Asian learners studying at a British university. Recent studies seem to continue this tradition, such as Polat et al.'s (2020) study focusing on the L2 writing development of Turkish learners. These studies reveal an urgent need in the field to examine L2 writing development in other national settings, such as the Hungarian context.

This study investigated the L2 writing development of 63 Hungarian EFL learners studying at a large university in Budapest. The research focused on the development of syntactic complexity and fluency, as previous studies indicated that high-proficiency learners often struggle to improve in these areas (Knoch et al., 2015; Storch, 2009).

Complexity and Fluency in L2 Writing

In the literature on L2 writing, there are two approaches to the notion of complexity which can be distinguished: (a) relative and (b) absolute (Dahl, 2004; Miestamo et al., 2008). According to Bulté and Housen (2012), both approaches refer to characteristics of language features (such as items, patterns, constructions, and rules). Language complexity is defined by the absolute approach in objective and quantitative terms, meaning that the absolute approach considers the number of elements (e.g., total number of words in a clause) that a language feature constitutes. However, an alternative approach has been recommended by Mazgutova and Kormos (2015), who claimed that linguistic complexity should be analyzed in line with the mode, genre, and communicative demands of the specific task that the participants are to perform. In other words, they recommended that the linguistic features of the specific genre or task-type need to be considered in the operationalization of linguistic complexity in written and oral performance. For example, in the domains of corpus linguistics (Biber & Gray, 2010; Biber et al., 2011) and systemic functional linguistics (Halliday & Martin, 1993/1996), it was shown that different linguistic features are characteristics of the two different modes: speech and writing. For example, more phrasal embedding can be found in academic writing than in speech. In addition, academic writing is characterized by complex nominalization and the use of abstract and compound nouns (Fang et al., 2006; Norris & Ortega, 2009). Furthermore, the complexity demands of writing and speech are also different across genres: Nippold (2004) and Berman & Nir-Sagiv (2007) showed that expository texts contain a higher number of relative clauses and passive constructions as well as more complex noun phrases than narrative texts.

According to Bulté and Housen (2012), syntactic complexity is a multidimensional construct consisting of at least three dimensions: (i) sentence complexity, (ii) clausal complexity, and (iii) phrasal complexity. However, sentence complexity is further divided into subconstructs: (a) coordination, (b) subordination, and (c) sub-sentential

complexity. Sentential coordination refers to clauses that are connected with a coordinating conjunct (e.g., *and*, *but*, *or*, or *so*), while sentential subordination is a structure that includes two clauses connected with a subordinating conjunction (e.g., *because*, *when*, *while*, or *though*). Finally, sub-sentential complexity denotes both clausal and phrasal complexity and is usually measured with the mean length of clause.

As far as the measurement of syntactic complexity is concerned, overall syntactic complexity in L2 writing is usually measured using the mean length of T-unit (MLTU) index. T-units are defined as "the shortest grammatically allowable sentences into which the theme could be segmented" (Hunt, 1965, p. 21). Regarding sentential complexity, coordination is gauged by applying the coordinate clause/clause index, while subordination is usually measured by the clauses per T-unit, dependent clauses per clauses, the number of subordinate clauses, the subordinate clauses per clauses, the subordinate clause per dependent clause, the subordinate clauses per T-unit, the relative clauses per T-unit, or the verb phrases per T-unit indices. Sub-sentential (clauses + phrasal) complexity is generally measured by the mean length of clause index. Clausal complexity can be measured by the syntactic arguments per clause index, while phrasal complexity is generally measured by the dependents per phrase index. In addition, the frequencies of passive forms, infinitival phrases, conjoined clauses, wh-clauses, imperatives, auxiliaries, comparatives, and conditionals might also be measured (Bulté & Housen, 2012).

There are numerous definitions of writing fluency in the literature (Abdel Latif, 2013). Consequently, several indices have been recommended to measure writing fluency. For example, Miller (2000) used pauses during writing as a measure of writing fluency, while Knoch (2007) counted the changes made to the text. Sasaki (2000) measured the composing rate of writers, while Storch (2009) gauged the number and length of T-units. In another study, Baba (2009) used the text quantity as an indicator of writing fluency. In the next section, the studies on the

development of syntactic complexity and fluency in L2 writing are reviewed.

The Development of Syntactic Complexity and Fluency in L2 Writing

Research on L2 writing development dates back to the 1990s (Henry, 1996; Valdés et al., 1992). For example, Valdés et al. (1992) examined Spanish as a foreign language learners' L2 writing development by using the American Council on the Teaching of Foreign Languages (ACTFL) proficiency guidelines to analyze written samples, whereas Henry (1996) researched early L2 writing development by analyzing autobiographical essays written by university-level students of Russian. Therefore, Valdés et al. (1992) and Henry's (1996) studies are not particularly relevant to the current context. In addition, L2 writing developmental research went through major changes in the past decades which makes examining studies published before the 2000s impractical.

A study which is highly relevant to the research context of the present study is Storch (2009), in which the impact of studying at an L2 medium university on L2 writing development was investigated. A total of 25 students were tasked with composing an argumentative essay of at least 300 words in a period of 55 minutes. Data collection took place at the beginning and end of the semester. The topic of the essays on both occasions was animal rights. A decrease was found in fluency (i.e., the number of words the students produced) from Time 1 to Time 2; however, the decrease was not statistically significant. Slight increases were observed in the dependent clause per clause (DC/C) and the clauses per T-unit (C/T) indices over time. Storch (2009) attributed the lack of improvement to the relatively short length of observation (12 weeks). Furthermore, Storch (2009) pointed out that the lack of improvement might be explained by the fact that the participants in her study were already at an advanced level of proficiency, and improvement for advanced students might be more difficult or might take a longer time to achieve.

In another study, Verspoor et al. (2012) holistically coded 437 texts written by Dutch learners of English as an L2 for proficiency levels, ranging from beginner to intermediate (Common European Framework of Reference [CEFR] A1.1 to B1.2). The authors hand-coded for 64 variables at the sentence, phrase, and word level. It was found that sentence length, all dependent clauses combined, and the use of present and past tense distinguished between levels of writing proficiency. Nevertheless, the specific constructions displayed nonlinear development, variability, and alternating associations among the different variables as expected from a dynamic usage-based perspective. In the case of participants with proficiency levels between A.1.1 and A.1.2 on the CEFR, the researchers found that mainly lexical changes appeared, while in the case of those with proficiency levels between A.1.2 and A2 on the CEFR it was predominantly syntactic changes which took place. However, those with proficiency levels between A2 and B1.1 showed both lexical and syntactic changes, while only lexical changes took place for those participants between B1.1 and B1.2 proficiency levels on the CEFR.

In another study conducted at an Australian university, Knoch et al. (2015) investigated 31 undergraduate students' L2 writing proficiency over a three-year degree study. The participants were given 30 minutes to compose an argumentative essay at the beginning and at the end of the study. The topic of the writing prompts was the same at Time 1 and Time 2. A statistically significant increase was found in the fluency measure (number of words) from Time 1 to Time 2. As far as syntactic complexity was concerned, an increase was found in the average clause length index, a stagnation was detected in the C/T index, and a decrease was found in the ratio of DC/C index. However, none of these changes in syntactic complexity were statistically significant. The authors explained the lack of statistically significant changes in syntactic complexity as a consequence of the relatively short length of the essays the students produced during the 30 minutes. They also claimed that "in short texts, the number of dependent clauses is likely to be quite finite" (p. 50). In addition, the researchers pointed out that the lack of improvement might be attributed

to the fact that the participants did not receive feedback on their essays. However, both Ferris (2003) and Leki (2006) claimed that feedback is possibly the most important factor in L2 writing improvement.

At a British university, Mazgutova and Kormos (2015) investigated how the syntactic features of L2 students' academic writing changed during a one-month long intensive English for Academic Purposes (EAP) program. The participants were grouped according to their level of proficiency, which included intermediate (N = 14) and upper-intermediate (N = 25). The students wrote an essay at the beginning and another at the end of the investigation. For the intermediate students' essays, increases were detected in the MLTU, dependent clause per T-unit (DC/T), modifiers per noun phrase, complex nominals (CN), and the syntactic structure similarity (STRUT) indices from Time 1 to Time 2. Out of these, the changes in the modifiers per noun phrase, the CN, and the STRUT indices were statistically significant. Interestingly, for the upper-intermediate students' essays decreases were found in the MLTU and the DC/T indices from Time 1 to 2. However, increases were detected in the modifiers per noun phrase, the CN, and STRUT indices from Time 1 to Time 2 in the upper-intermediate learners' essays. From these, only the change in the syntactic structure similarity index was statistically significant. Based on the results, the researchers concluded that learners could significantly improve their academic writing over a one-month-long EAP course.

In a more recent study, Polat et al. (2020) investigated growth patterns in the written syntactic complexity of 284 Turkish learners of English. Data were collected at three points in time, resulting in a final corpus of 852 written samples over three semesters. The authors investigated the indices which characterize elementary, pre-intermediate, and intermediate EFL learners' writing. In addition, the authors examined how syntactic complexity indices change over time as students progressed from elementary to pre-intermediate and intermediate levels. It was found that the writing of elementary and pre-intermediate learners was

characterized by phrasal coordination, while intermediate writers relied on constructions that were more complex and possessed greater variety.

Research Gap

As seen in the previous section, earlier studies on the longitudinal development of syntactic complexity and fluency in writing predominantly focused on Asian learners studying at Asian universities (Jiang et al., 2019; Polat et al., 2020), Asian learners studying in the UK (Mazgutova & Kormos, 2015), and Asian learners studying at Australian universities (Knoch et al., 2015; Storch, 2009). Consequently, our knowledge about the longitudinal development of syntactic complexity and fluency in writing in the Hungarian context is limited. To date, only four studies have investigated the longitudinal development of syntactic complexity among Hungarian learners. However, these studies were predominantly case studies (Wind, 2018, 2021; Wind & Harding, 2020) or L2 writing was not the primary focus of investigation (Wind & Zólyomi, 2022). Wind (2018) traced four learners' writing development over a nine-month period. The four learners were asked to compose International English Language Testing System (IELTS)-type argumentative essays. However, in contrast to the previous studies on L2 writing development, Wind (2018) adopted a time-series analysis: instead of using a pre-test/post-test design, written samples were collected 23 times over the course of nine months. Since it was unreasonable to run statistical analyses (e.g., the Wilcoxon signed-rank test) due to the small number of participants, developmental trajectories were plotted and inspected for trends. In addition, the data points were smoothed by applying 3-period moving averages. Upward trend lines were detected in the finite verb ratio (FVR) index for two learners, while downward trends were found in the FVR index for the other two learners. Upward trend lines were detected in the subordination index, measured by the DC/C, for three learners, while a downward trend was found in the DC/C index for one learner. This result indicated that three learners tended to use more dependent

clauses in their essays over time. As far as phrasal complexity (measured by the CN/C index) was concerned, an upward trend was detected in the data of only one learner. This finding implied that only one participant's data demonstrated the characteristics of academic writing (i.e., phrasal elaboration). In another study, Wind and Harding (2020) examined one participant's syntactic development, measured with the finite verb ratio, and found stagnation over time. The authors attributed the stagnation of syntactic complexity to the lack of self-regulatory strategies used by their participant. In another study, Wind (2021) traced six students' syntactic and lexical development over a seven-month period. The trend lines showed slightly upward trends for the syntactic and lexical complexity indices measured by the FVR and the average word length (AWL) indices, respectively. The researcher also measured the extent to which participants were engaged in self-reflective processes by conducting semi-structured interviews right after the students finished composing their essays. It was concluded that as the learners tended to become more engaged in self-reflective processes, slight improvements were detected in their L2 writing. However, it is important to note that was a multiple-case study, and thus generalizations cannot be drawn from it. Lastly, Wind and Zólyomi (2022) investigated the longitudinal development of self-assessment abilities along with linguistic complexity and found that learners composed shorter sentences after taking the advanced writing course as measured by the MLC index.

To conclude, numerous recent studies have investigated the development of syntactic complexity and fluency in L2 writing. However, previous quantitative studies have mainly focused on Asian leaners of English studying abroad (e.g., in the United Kingdom and Australia). Previous developmental studies focusing on Hungarian EFL learners adopted a multiple-case study research design or L2 writing was not the main focus of investigation; therefore, there is a clear need to explore the development of syntax and fluency by adopting a pre-test/post-test research design.

Research Questions

Although numerous studies have investigated the longitudinal development of syntactic complexity and fluency (Jiang et al., 2019; Knoch et al., 2015; Mazgutova & Kormos, 2015; Polat et al., 2020; Storch, 2009), no previous quantitative studies have focused on Hungarian EFL leaners. Therefore, this study aims to fill this gap by answering the following research questions:

> RQ1: How do fluency and syntactic complexity indices change in argumentative essays during a one-semester advanced writing course?
>
> RQ2: How do fluency and syntactic complexity indices change in the UTT and EAS students' argumentative essays during a one-semester advanced writing course?

Methodology

Participants

Altogether 63 EFL student participants took part in this study, 90% of whom were from a Hungarian L1 background, while the remaining 10% were from Chinese, Romanian, Spanish, and Kazakh L1 background (see Table 1). Fifty-seven percent of the participants were in the UTT program, while 39% of the students were in the EAS program. In addition, there were two Erasmus students who took the advanced writing (AW) course. The proficiency of the participants was around CEFR level C1, since the prerequisite for taking the AW course was the successful completion of the proficiency exam at the end of the first year of both the EAS and UTT programs. The proficiency exam at the university is equivalent to CEFR level C1.

Table 1

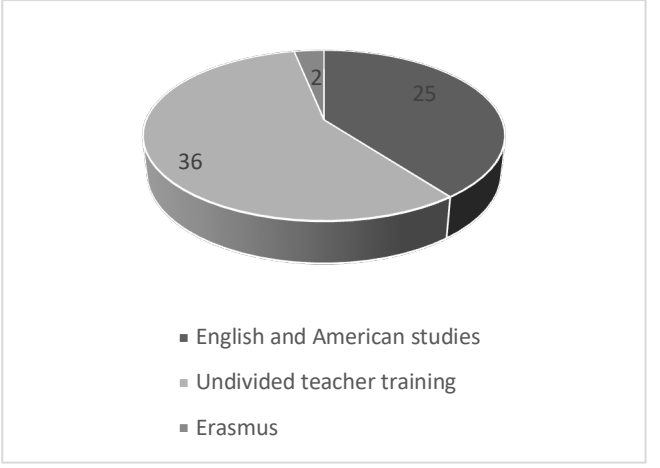
Participants in the AW Courses

Sex	Female	51
	Male	12
L1 background	Hungarian	57
	Chinese	3
	Romanian	1
	Spanish	1
	Kazakh	1

The participants were recruited from six advanced writing courses from a large university in Budapest. There were three teachers involved in this project: the author taught three AW courses, while two of his colleagues (a PhD student and a native speaker of English) taught the three other courses. The author taught the three AW course in three consecutive semesters, while his colleagues taught the courses in the spring term of 2020.

Figure 1

The Distribution of the Students Across the Programs



Advanced Writing Course

The advanced writing course is held by instructors of the Department of English Applied Linguistics (DEAL) at a large university in Hungary. The prerequisite to attend the AW course is the successful completion of the two-semester academic skills (AS) 1 and 2 courses in the first year of the undergraduate studies of the EAS and UTT students. During the AS 1 and 2 courses, students master the skills of paraphrasing, summarizing, and synthesizing (Tankó, 2019). At the end of the AS 1 course, the students are required to take the Academic Skills Test (AST), in which they are asked to write a guided summary. The result of the AST constitutes 40% of the students' final grade. At the end of the AS 2 course, the students are required to compose an argumentative essay in which they synthesize at least five different sources. As opposed to the AST, at the end of the AS 2 course the final grade is given by the instructor of the course solely based on the course work.

During the AW course, the instructors can design their own curriculum; therefore, there might be major differences in the methods that the teachers use. Nine sections of the course are usually held in the autumn term, while three are held in the spring term. The majority of the AW course instructors are native speakers.

Instruments

Each participant was required to compose two argumentative essays: one at the beginning and one at the end of the course. The students were asked to compose a *minimum* 200-word-long essay in 30 minutes. Topics for the writing prompts were chosen in relation to the general field of language education in order to control for topic difficulty, as it was presumed that EFL learners would hold an opinion on language pedagogy since they had experience in learning foreign languages. The two topics which were chosen are as follows:

> Topic A: A native language teacher is always better than a non-native one. To what extent do you agree? And why?
>
> Topic B: The older you get, the more difficult it is to learn a foreign language. To what extent do you agree? And why?

The order of the writing tasks was counterbalanced, meaning that half of the participants completed Task A at the beginning of the course and Topic B at the end. The other half of the participants started with Topic B and completed Topic A at the end of the course.

Data Analysis

The L2 Syntactic Complexity Analyzer (L2SCA) was used to calculate the syntactic complexity indices chosen for this study (Ai & Lu, 2013; Lu, 2010, 2011; Lu & Ai, 2015). Before running the software, the digitalized essays were corrected for misspellings and erroneous punctuation so that the computational program would be able to detect and identify lexical items. Table 2 shows the syntactic complexity indices used in this study.

Norris and Ortega (2009) claimed that it might be redundant to use two indices to measure the same construct, as in the case of using the DC/C and the DC/T indices to measure subordination. However, in this study the most important aim was to explore whether any of the syntactic complexity indices available in the L2SCA changed over time, meaning that such potential redundancies did not present an issue.

The number of words index was used to measure fluency in this study and is also calculated by the L2SCA. Although in this study the number of words required was set by the task instructions, it was presumed that the number of words the students produced in 30 minutes reflects their level of writing fluency.

Table 2

Syntactic Complexity Indices Calculated by the L2SCA

Index	Code
Mean length of sentence	MLS
Mean length of T-unit	MLT
Mean length of clause	MLC
Clause per sentence	C/S
Verb phrase per T-unit	VP/T
Clause per T-unit	C/T
Dependent clause per clause	DC/C
Dependent clause per T-unit	DC/T
T-unit per sentence	T/S
Complex T-unit ratio	CT/T
Coordinate phrase per T-unit	CP/C
Coordinate phrase per clause	CP/T
Complex nominal per T-unit	CN/T
Complex nominal per clause	CN/C

Statistical analyses were carried out using JASP Team version 0.16 (2021). First, the distribution of the data was tested by means of the Shapiro Wilk test, which revealed non-normal distribution for the syntactic complexity and fluency indices. Therefore, nonparametric tests were used for statistical inference. The Wilcoxon signed-rank test, a nonparametric equivalent to the paired sample t-test, was applied to examine the differences between Time 1 and Time 2. Effect size was calculated by the $r = z/\sqrt{N}$ formula. Absolute effect sizes of 0.1–0.29 were taken as indicating a small effect, 0.3 to 0.49 indicated a medium effect, and values greater than 0.5 were considered a large effect (Cohen, 1969).

Results

This section provides an overview of the findings in light of the research questions that guided this study. The descriptive statistics (i.e., mean and standard deviation) for all 14 syntactic complexity indices and one fluency

index are displayed in Table 3. The descriptive statistics show that the fluency index increased from Time 1 to Time 2. As far as the syntactic complexity indices are concerned, only three indices (the VP/T, the CN/T, and the CN/C) increased from Time 1 to Time 2. Two syntactic complexity indices, the DC/C and the T/S, remained the same over time, while nine syntactic complexity indices decreased from Time 1 to Time 2.

Table 3

Descriptive Statistics for the EAS and UTT Groups (Combined)

Index	Time 1		Time 2	
	Mean	SD	Mean	SD
nwords	238.48	54.67	260.56	66.65
MLS	22.76	6.22	22.30	5.37
MLT	19.61	5.16	19.18	4.06
MLC	10.08	1.76	9.95	1.19
C/S	2.28	0.59	2.24	0.46
VP/T	2.75	0.68	2.76	0.58
C/T	1.96	0.46	1.93	0.36
DC/C	0.44	0.11	0.44	0.10
DC/T	0.90	0.41	0.87	0.34
T/S	1.17	0.16	1.17	0.16
CT/T	0.60	0.18	0.59	0.15
CP/T	0.46	0.28	0.38	0.21
CP/C	0.24	0.13	0.20	0.12
CN/T	2.34	0.72	2.39	0.78
CN/C	1.21	0.30	1.24	0.32

Table 4 shows the descriptive statistics for the UTT and the EAS groups. It can be seen that the seven indices (the NW, C/T, DC/C, DC/T, C/T, CN/T, and CN/C) increased, while seven indices (the MLS, the MLT, the MLC, the C/S, the T/S, the CP/T, and the CP/C) decreased. One index (VP/T) stagnated in the UTT students' essays from Time 1 to Time 2. As far as the

EAS students were concerned, increases were detected in the case of six syntactic complexity indices (the NW, C/S, VP/T, T/S, CN/T, and CN/C), whereas nine indices decreased (the MLS, MLT, MLC, C/T, DC/C, DC/T, C/T, CP/T, and CP/C) from Time 1 to Time 2.

Table 4

Descriptive Statistics for the UTT and the EAS Groups

	UTT				EAS			
	Time 1		Time 2		Time 1		Time 2	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
nwords	239.5	60.17	250.5	51.64	236.4	49.27	269.32	41.76
MLS	21.41	5.64	20.78	3.68	24.79	6.81	24.60	7.09
MLT	18.41	5.01	18.39	3.88	21.16	5.42	20.34	4.53
MLC	9.81	2.01	9.79	1.23	10.53	1.32	10.20	1.12
C/S	2.20	0.51	2.14	0.35	2.39	0.70	2.41	0.60
VP/T	2.63	0.60	2.63	0.59	2.85	0.79	2.94	0.59
C/T	1.88	0.37	1.89	0.36	2.04	0.59	2.00	0.39
DC/C	0.42	0.10	0.43	0.10	0.46	0.10	0.45	0.09
DC/T	0.82	0.34	0.83	0.35	0.99	0.48	0.93	0.33
T/S	1.17	0.15	1.15	0.15	1.17	0.18	1.20	0.17
CT/T	0.56	0.17	0.59	0.15	0.63	0.19	0.60	0.16
CP/T	0.43	0.26	0.37	0.20	0.49	0.29	0.37	0.23
CP/C	0.23	0.14	0.20	0.12	0.24	0.11	0.19	0.12
CN/T	2.17	0.70	2.22	0.70	2.57	0.78	2.64	0.90
CN/C	1.16	0.33	1.18	0.28	1.28	0.26	1.32	0.37

For the combined groups, the results of the Wilcoxon signed-rank tests only showed statistically significant differences for the fluency index ($Z = - 2.76$, $p < 0.01$, $r = -0.35$). The effect size was medium for the fluency index. The significant difference in the number of words metric suggests an improvement in fluency. Interestingly, statistically significant

differences were not found in the syntactic complexity indices, indicating stagnation in development.

Similarly, as the UTT and the EAS groups are concerned, only the fluency index was statistically significant for the EAS group from Time 1 to Time 2 (Z = -2.58, p <0.01, r = -0.33). However, none of the syntactic complexity indices showed statistically significant changes from Time 1 to Time 2 for either group.

Discussion

This study found that undergraduate EFL students improved their fluency over one semester during an AW course. However, no improvements were detected in the longitudinal development of syntactic complexity in L2 writing.

The lack of improvement in syntactic complexity is not uncommon in the literature on L2 writing development. For example, Knoch et al. (2015) also found stagnation in the clauses per T-unit ratio index, while the ratio of dependent clauses to clauses index decreased over three years in ESL students' writing. Only a slight increase was detected in the average clause length in word index. The authors explained the lack of improvement in syntactic complexity as a result of the relatively short length of the writing task, claiming that the number of dependent clauses might be quite limited in short texts. In addition, their study found the mean value of the NW was 281.16 at the first measurement point, while the mean NW was 325.06 at the last one. Therefore, it might be presumed that the lack of improvement in syntactic complexity might be explained by the relatively short length of the written samples collected for this study. Knoch et al. (2015) also attributed the stagnation in development to the fact that their participants produced very little writing during their degree program. However, in the present study the undergraduates were asked to compose several other shorter texts while attending the AW course. It is also important to note that the participants in this study were

taught by three different teachers who required their students to produce different numbers of texts throughout the semester.

Another reason for the lack of improvement in syntactic complexity in this study might be that the learners did not receive feedback on the argumentative essays that they produced for this study. However, Ferris (2003) and Leki (2006) claimed that feedback on writing might be the most fundamental factor leading to improvement in L2 writing. Nevertheless, the teachers of the AW course provided feedback on the other writing assignments that their students produced during the semester. It might be presumed that the type of feedback that learners received on their written assignments might not be the most appropriate for the learners' needs.

In addition, Storch (2009) explained the lack of improvement in grammatical complexity, measured by the ratio of clauses per T-unit and the dependent clause ratio indices, by the relatively short length of the observation period (i.e., 12 weeks). It might be presumed that the participants in this study were similarly unable to measurably develop their syntax over the semester, which was the nearly same period as in Storch's (2009) study. According to Ortega (2003), grammatical complexity requires a minimum of a year to improve. In this study, the length of the observation was limited to 13 weeks.

Another possible reason for the stagnation of syntactic development might be that the participants in this study were already at an advanced level of proficiency. Consequently, improvement for these learners might be more difficult and take a longer time to achieve. In addition, literature on syntactic development suggests that lower proficiency learners develop in more areas of syntax than higher proficiency learners (Mazgutova & Kormos, 2015; Polat et al., 2020). Another reason for the stagnation of syntactic complexity might be related to possible issues with the task instruction. In this study, only the minimum text length was indicated (200 words). However, the maximum length was not displayed in the instruction. Therefore, the students might have focused on fluency (i.e., composing lengthier essays) and not on improving their syntactic complexity.

Conclusion

This study investigated the longitudinal development of fluency and syntactic complexity in university students' writing at a large university in Hungary. It was found that the university students' essays from Time 1 and Time 2 showed statistically significant improvements in fluency during a semester-long AW course. Furthermore, it was found that the data from the EAS students' essays showed statistically significant increases, while the data from the UTT students did not. One of the possible explanations for the stagnation of syntactic complexity in these cases might be the limited length of observation and the relatively high level of language proficiency of the learners in this study.

This study has several limitations. As mentioned above, the length of observation was narrow; one semester might not be enough for such students to show statistically significant improvements. Therefore, future studies carried out with the same population could lengthen the period of observation and trace the learners' writing development during their Academic Skills 1 and 2 courses as well as their AW courses. During these courses, students are instructed how to compose academic texts, and tracing learners' L2 writing development throughout these three semesters might show a different picture of development.

There are also some pedagogical implications of this study. AW course instructors should draw their students' attention to the syntactic features of academic writing. In other words, students should be instructed to rely less on clausal subordination and instead focus on increasing the level of phrasal embedding in their writing. Consequently, students will be able to better demonstrate the characteristic features of academic writing. Future studies should also set a maximum length in the task instructions in order to avoid a focus on fluency.

Note: The author is a member of the MTA-ELTE Foreign Language Teaching Research Group. This study was funded by the Scientific Foundations of Education Research Program of the Hungarian Academy of Sciences.

References

Abdel Latif, M. M. M. (2013). What do we mean by writing fluency and how can it be validly measured? *Applied Linguistics*, *34*(1), 99–105. https://doi.org/10.1093/applin/ams073

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 Publishing Company. https://doi.org/10.1075/scl.59.15ai

Baba, K. (2009). Aspects of lexical proficiency in writing summaries in a foreign language. *Journal of Second Language Writing*, *18*(3), 191–208. https://doi.org/10.1016/j.jslw.2009.05.003

Berman, R. A., & Nir-Sagiv, B. (2007). Comparing narrative and expository text construction across adolescence: a developmental paradox. *Discourse Processes*, *43*, 79–120. http://dx.doi.org/10.1080/01638530709336894

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

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, *45*(1), 5–35. http://dx.doi.org/10.5054/tq.2011.244483

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21–46). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.32.02bul

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. Academic Press.

Dahl, Ö. (2004). *The growth and maintenance of linguistic complexity*. John Benjamins Publishing Company. https://doi.org/10.1075/slcs.71

Fang, Z., Schleppegrell, M. J., & Cox, B. (2006). Understanding the language demands of schooling: Nouns in academic registers. *Journal of Literacy Research*, *38*(3), 247–273. http://dx.doi.org/10.1207/s15548430jlr3803_1

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

Halliday, M. A. K., & Martin, J. R. (1993/1996). *Writing science: Literacy and discursive power*. Falmer Press.

Henry, K. (1996). Early L2 writing development: A study of autobiographical essays by university-level students of Russian. *The Modern Language Journal*, *80*(3), 309–326. https://doi.org/10.1111/j.1540-4781.1996.tb01613.x

Hunt, K. W. (1965). Grammatical structures written at three grade levels. *NCTE Research Report* No. 3.

JASP Team (2021). *JASP* (Version 0.16) [Computer software]. https://jasp-stats.org/

Jiang, J., Bi, P., & Liu, H. (2019). Syntactic complexity development in the writings of EFL learners: Insights from a dependency syntactically-annotated corpus. *Journal of Second Language Writing*, *46*. Article 100666. https://doi.org/10.1016/j.jslw.2019.100666

Knoch, U. (2007). *Diagnostic writing assessment: The development and validation of a rating scale* [Doctoral dissertation, University of Auckland].

Leki, I. (2006). You cannot ignore: L2 graduate students' response to discipline-based written feedback. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 266–286). Cambridge University Press. https://doi.org/10.1017/CBO9781139524742.016

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://onlinelibrary.wiley.com/journal/15457249

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

Mazgutova, D., & Kormos, J. (2015). Syntactic and lexical development in an intensive English for Academic Purposes programme. *Journal of Second Language Writing*, *29*, 3–15. http://dx.doi.org/10.1016/j.jslw.2015.06.004

Miestamo, M., Sinnemäki, K., & Karlsson, F. (Eds.). (2008). *Language complexity: Typology, contact, change*. John Benjamins Publishing Company. https://doi.org/10.1075/slcs.94

Miller, K. S. (2000). Academic writers on-line: Investigating pausing in the production of text. *Language Teaching Research*, *4*(2), 123–148. https://doi.org/10.1177/136216880000400203

Nippold, M. A. (2004). Research on later language development: International perspectives. In R. Berman (Ed.), *Language development across childhood and adolescence* (pp. 1–8). John Benjamins Publishing Company. https://doi.org/10.1075/tilar.3.04nip

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, *30*(4), 555–578. https://doi.org/10.1093/applin/amp044

Polat, N., Mahalingappa, L., & Mancilla, R. L. (2020). Longitudinal growth trajectories of written syntactic complexity: The case of Turkish learners in an intensive English program. *Applied Linguistics*, *41*(5), 688–711. https://doi.org/10.1093/applin/amz034

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

Storch, N. (2009). The impact of studying in a second language (L2) medium university on the development of L2 writing. *Journal of Second Language Writing*, *18*(2), 103–118. https://doi.org/10.1016/j.jslw.2009.02.003

Valdés, G., Haro, P., & Arriarza, M. P. E. (1992). The development of writing abilities in a foreign language: Contributions toward a general theory of L2 writing. *The Modern Language Journal*, *76*(3), 333–352. https://doi.org/10.2307/330163

Verspoor, M., Schmid, M. S., & Xu, X. (2012). A dynamic usage based perspective on L2 writing. *Journal of Second Language Writing*, *21*(3), 239–263. https://doi.org/10.1016/j.jslw.2012.03.007

Wind, A. M. (2018). *Second language writing development from a Complex Dynamic Systems Theory perspective: A multiple case-study of Hungarian learners of English*. [Doctoral Dissertation, Lancaster University]. https://doi.org/10.17635/lancaster/thesis/519

Wind, A. M. (2021). Nonlinearity and inter- and intra-individual variability in the extent of engagement in self-reflection and its role in second language writing: A multiple-case study. *System*, *103*, Article 102672. https://doi.org/10.1016/j.system.2021.102672

Wind, A. M., & Harding, L. W. (2020). Attractor states in the development of linguistic complexity in second language writing and the role of self-regulation: A longitudinal case study. In W. Lowie, M. Michel, A. Rousse-Malpat, M. Keijzer, & R. Steinkrauss (Eds.), *Usage based dynamics in second language development* (pp. 130–154). Multilingual Matters. https://doi.org/10.21832/9781788925259-009

Wind, A. M., & Zólyomi, A. (2022). The longitudinal development of self-assessment and academic writing: An advanced writing programme. *Language Learning in Higher Education*, *12*(1), 185–207. https://doi.org/10.1515/cercles-2022-2046

5

Students' Opinion on the Usefulness of Netflix, Instagram, and TikTok in EFL Learning—A Pilot Study

Boglárka Spissich
spissich.boglarka@btk.elte.hu

Abstract

The present research explored Hungarian secondary school students' opinions on the usefulness of films and social media in English as a foreign language (EFL) learning. Nowadays teenagers spend a great amount of time on their phones using different mobile applications. Much of the content they see online is in English, which can help them improve their EFL skills. In this study, 101 secondary school students between the ages of 12–18 were asked about their habits using Netflix, Instagram, and TikTok using a questionnaire. The participants of the study were all from the same secondary grammar school located in a larger city in Hungary. Results showed that learners do not use the applications to explicitly improve their EFL skills, but they do notice language development. Teenagers highlighted vocabulary development as the greatest benefit to be gained from the three applications, followed by listening skills; however, they were doubtful about their role in grammar learning. When participants were analyzed as groups, it was found that those who had been learning English for more than five years noticed that their language skills improved when they were asked about the advantages of watching films in English. In the case of the social media applications, gender differences were also revealed: females found the applications to be more useful for EFL learning.

***Keywords*:** EFL learning, films, social media, implicit vs. explicit learning, motivation

DOI: 10.21862/ELTE.DEAL.2022.5.119

Students' Opinion on the Usefulness of Netflix, Instagram, and TikTok in EFL Learning—A Pilot Study

Nowadays technology and the Internet are part of our everyday life. Thousands of smartphone applications exist for different purposes, and younger generations are especially drawn to these apps and websites, using them on a daily basis. Each user can find something they are interested in, whether it is photography, foreign languages, or news. However, there are some applications which are used by the majority of teenage users, such as Netflix, Instagram, and TikTok (Iqbal, 2021). It is well-known that teenagers spend a considerable amount of time on their smartphones and computers surfing these online platforms. Although they primarily use these applications for entertainment and not for language learning, learners often need English to communicate with others or to understand online content (Arif, 2019). The aim of this study is to explore how useful students find these applications for the improvement of their English as a foreign language (EFL) skills.

Social media provides new perspectives for EFL learning and teaching (Yadav, 2021). Despite the fact that these applications provide a valuable opportunity for EFL teaching and learning, many language instructors believe that using social media in class is more distracting than beneficial (Allam & Elyas, 2016); thus, learners are often left alone to discover the benefits of such platforms in an out-of-class environment. It is important to note the different ways in which applications are used by students in order to see how learners can use them to enhance their language knowledge. The first platform examined in this study is Netflix, the most popular streaming platform whose subscribers make up 37% of Internet users (Panda, 2020). Users are not able to communicate with others through the platform, as the application is solely used to consume content such as movies and TV series. Watching films in the target language is mostly associated with receptive skills, but with guidance, films can be used to practice speaking skills as well (Khan, 2015). The other two applications, Instagram and TikTok, allow users to communicate with

each other in various ways. Instagram's main purpose is to share photos, but it is also possible to use the application to post short videos, comment, or send messages to others. Besides its primary functions, users can edit pictures by applying filters, share their location, tag their friends, and use hashtags to find content they like and make their posts more visible to other users (Handayani, 2016). The TikTok application, which gained popularity around the time of the COVID-19 pandemic, is one of the most popular platforms among teenagers today (Iqbal, 2021). Among its 98 million European users, 60% are between 16–24 years old. Since its primary function is to share short videos, TikTok is mostly associated with listening skills and vocabulary acquisition in regard to its language learning opportunities (Yang, 2020). Besides posting one's own videos, it is possible to share, like, and comment on these clips as well. Messaging is also possible, although first users need to follow the person they want to connect with.

The purpose of this study was to compare and contrast secondary school students' opinions about the usefulness of the three above-mentioned platforms: Netflix, Instagram, and TikTok. Participants in the study were asked about their habits related to the use of the applications, the reasons that they use them, and in what ways they think the given app can improve their EFL knowledge. An online questionnaire was distributed to 101 secondary school participants to find answers to these questions.

Literature Review

The following section focuses on the relevant theoretical and empirical studies to provide a comprehensive overview of the relevant concepts. Firstly, implicit and explicit EFL learning will be explored and discussed alongside the theory of noticing. This is followed by a review of the definitions and roles of intrinsic and extrinsic motivation. Finally, the use of films and social media in EFL learning is discussed in connection with listening skills, as well as vocabulary and grammar knowledge.

Implicit and Explicit EFL Learning

The difference between implicit and explicit learning can be defined by the learner's conscious efforts. Implicit or incidental language learning is "associated with more natural language learning, that is, outside formal classroom instruction" (Brutin et al., 2011, p. 759) during which "no rules are formulated" (DeKeyser, 1994, p. 188). Explicit learning, on the other hand, is defined as "a process during which participants acquire conscious (explicit) knowledge" (Rebuschat, 2015, p. 13). Watching films in a foreign language is often linked to implicit learning, as learners may have the opportunity to observe grammar rules and utilize features of spoken language through viewing films (Giampieri, 2018).

Although learners do not pay conscious attention to specific aspects of the language during implicit learning, they may notice different features and even improvement in their EFL skills. Schmidt (1990) claimed that language learning is not possible without awareness, but it is only through noticing certain aspects of the language that incidental learning can occur. Schmidt (2010) succinctly described the Noticing Hypothesis: "in the simplest terms, people learn about the things that they attend to and do not learn much about the things they do not attend to" (p. 722). Based on the hypothesis, it is possible that learners might implicitly learn language elements while paying attention to films or other types of video content. Navidinia et al. (2019) concluded their study by investigating the role of noticing in EFL speaking accuracy, claiming that "noticing had a positive effect on the EFL learners' grammatical, phonological and lexico-semantic accuracy" (p. 94).

Previous studies concluded that the most effective way to learn a language is the combination of explicit and implicit learning (Berry & Broadbent, 1988; Reber et al., 1980). According to Ellis (2015), "learners' language systematicity emerges from their history of interactions of implicit and explicit language learning" (p. 21). While explicit awareness is needed for the memorization of grammar rules (DeKeyser, 1994), vocabulary acquisition can happen implicitly through frequently viewing

or listening to content created in the target language (Webb, 2010). It can be concluded then that explicit learning might be more efficient in some aspects of EFL learning but engaging in out-of-class activities in the target language is also certainly useful for learners (Lai et al., 2015).

Intrinsic and Extrinsic Motivation in EFL Learning

"Motivation is one of the main determinants of second/foreign language learning achievement" (Dörnyei, 1994, p. 273). Motivation is often referred to in terms of two main types: extrinsic and intrinsic motivation. Extrinsic motivation is experienced when learners are motivated by someone or something else, such as when they are rewarded for learning. Intrinsic motivation, however, comes from within the learner; an intrinsically motivated language learner is likely to have an "interest in foreign languages, cultures and people" (p. 275) and "the desire for new stimuli and challenges" (p. 275).

Learners who choose to watch films or engage with social media in English often do so for the purpose of enjoyment. This intrinsic motivation, however, is closely related to the learner's language level (Spithill, 1980). On the one hand, at more advanced stages "the student finds the language intrinsically rewarding, wants to use it and is approaching cultural identification" (p. 74). A teacher's role at this stage is to guide learners and show them how to stay motivated and interested in the target language. Lower-level learners, on the other hand, may find authentic English to be too demanding, which may demotivate them. It is possible to encourage them, as with "proper motivational techniques, both fatigue and boredom can be reduced" (p. 72). Spithill suggested that with time and effort, extrinsic motivation turns into intrinsic motivation, but teachers always have a role in guiding and encouraging learners.

Csíkszentmihályi (1985) agreed that although intrinsic motivation is based on learners' interests, teachers can influence its development and can help learners achieve a flow state, which is described as: "action follows upon action according to an internal logic that seems to need no

conscious intervention by the actor" (p. 36). Not only motivational techniques but also feedback given by a teacher can influence how easily one reaches a state of flow. Teachers, according to the researcher, should promote the idea that learning can be an enjoyable activity by keeping the aims and requests of students in mind. Dörnyei (1994) also highlighted the importance of enjoyment in his 30-item list of recommendations for teachers on how to motivate students. He believed that including "a sociocultural component in the L2 syllabus" (p. 281), such as TV programs or videos, makes language learning more enjoyable, thus improving the EFL learning process.

Learning EFL Through Films and Social Media

The benefits of using target language films and videos in foreign language (FL) teaching was first discussed in the 1990s when a group of French learners were instructed with the help of videos (Secules et al., 1992). In the 2010s, researchers examined the effects of films used as textbooks (Hertel & Harrington, 2015) as well as films as supplementary teaching tools (Rouxel-Cubberly, 2014). These studies showed positive results regarding in-class video use. However, incorporating videos into language lessons can be time-consuming, so learners are more likely to engage in this activity in their free time. Although this engagement is mainly for entertainment, a study involving Chinese secondary school students found that watching videos in English is often more beneficial than explicit EFL practice (Lai et al., 2015).

Watching films and videos online is mostly associated with an improvement in listening skills. Learners believe that watching films is more beneficial for their listening skills than listening to audio tapes (Pamungkas & Adi, 2020). In a study conducted among Thai EFL students, TikTok was rated the second most useful application for listening skill development after Netflix (Parnrod et al., 2020). In Yang's (2020) research on TikTok as a learning tool, secondary school students expressed that they mostly used the application to improve their EFL listening skills.

Similar to TikTok, short videos can also be shared on Instagram, and it is advised that learners watch created by native English speakers to enhance their listening skills (Handayani, 2016).

Although Khan (2015) mentioned the need for a teacher to facilitate the use of videos in the classroom to develop speaking skills, students believe that the video sharing platform TikTok is the second most useful application for oral skills development (Yang, 2020). One explanation for this can be that because learners hear the native-speaker pronunciation of words, it makes them less anxious to use these words in their own speech (Alkathiri, 2019). Another aspect of speaking is vocabulary knowledge, and both learners and researchers agree that watching films and videos enhances learners' vocabulary (Rodgers, 2013; Yang, 2020); however, for this to occur EFL learners need to engage with media habitually in order to encounter words multiple times (Webb, 2010). The practicality of subtitles is supported from elementary (Koolstra & Beentjes, 1990) to university-level EFL learners (Gorijan, 2014). While the studies above promote the use of English audio with subtitles in the learners' native language, there is empirical evidence that the use of English subtitles has a positive effect on vocabulary development as well (Masrai & Milton, 2018).

Grammar learning, on the other hand, is mainly associated with in-class EFL teaching. Articles can be found on how to teach grammar with the help of Instagram (Handayani, 2016) or films (Kabooha, 2016), but as grammar rules are more easily learnt explicitly (Williams, 2005), this topic has not been widely researched. Moreover, EFL learners have demonstrated a somewhat negative attitude toward learning grammar via social media. TikTok was found to be useful in this area by only 44.9% of the participants in Yang's (2020) study, and Instagram was mentioned as an unreliable source for grammar "because slang or inappropriate English use is pretty common" (Gonulal, 2019, p. 317) on the platform.

Based on the above-mentioned studies, it can be concluded that the out-of-class use of films and social media is mainly associated with the improvement of vocabulary and listening skills. In order to gather

additional data and gain more information about how learners use the three previously mentioned applications, the present research aims to answer the following research questions:

1. How often do Hungarian secondary school students use Netflix, Instagram, and TikTok for explicit language learning?

2. What motivates Hungarian secondary school students to use Netflix, Instagram, and TikTok for EFL learning?

3. How useful do Hungarian secondary school students find Netflix, Instagram, and TikTok for implicit EFL learning concerning listening skills, vocabulary, and grammar knowledge?

Methods

The present research is a pilot study for a questionnaire aimed at examining secondary school students' language learning habits via social media. Questions were formed about three of the most popular applications used by today's learners: Netflix, Instagram, and TikTok. The questionnaire was designed to examine how learners view these platforms and social media forums as tools for learning and practicing EFL.

Participants and Setting

The participants in the study were 101 students from a Hungarian secondary grammar school. The distribution of the participants can be seen in Table 1. There were 60 females and 39 males participating in the research, and two students who preferred not stating their gender. Out of the 101 students, 80 studied English as their first foreign language (FL1), and the remaining of the students chose it as a second foreign language

(FL2). These 21 students learnt German as their first foreign language. More than half of the students (n = 67) had been learning EFL for more than five years, and another 24 for more than three years. About half of the participants (n = 56) were in seventh or eighth grade at the secondary school, which means that they were 12–14 years old, with another 28 participants in their ninth year of school. There were no participants older than 18.

Table 1

Distribution of Participants

Gender		En. FL1	En. FL2	7th-8th grade	9th grade	10th-12th grade	EFL 5 or more years	EFL less than 5 years
Female	60	47	13	36	18	6	43	17
Male	39	32	7	20	10	9	32	7
NA*	2	2	0	0	0	2	2	0
Total	101	81	20	56	28	17	77	24

*Prefer not to say

Instrument

An online questionnaire was used to measure the different constructs using a 5-point Likert scale which the participants used to indicate how much they agreed with the given statement. The same statements were repeated three times for the three different applications, and measured a total of eight main scales:

1. Explicit learning (11 items): measuring learners' intentional EFL learning through the given application. *Example: When I hear a new item of vocabulary in an English film, I look it up.*

2. Implicit learning (16 items): measuring learners' incidental EFL learning through the given application. *Example: I believe I can learn new items of vocabulary from watching films in English even if I don't pay direct attention to them.*
3. Noticing (6 items): measuring the extent to which learners realize they learn EFL while using the given application. *Example: Since watching films in English on Netflix, I have noticed that my listening skills have improved.*
4. Vocabulary (5 items): measuring how useful learners find the given platform for the implicit learning of new EFL vocabulary items. *Example: I use EFL vocabulary items that I heard in English films.*
5. Listening skills (4 items): measuring how useful learners find the given platform for the implicit practice of EFL listening skills. *Example: I believe if I watch films in English on Netflix, I will do better on an EFL listening task.*
6. Grammar knowledge (4 items): measuring how useful learners find the given platform for the implicit learning of new EFL grammar rules. *Example: I believe I can use grammar structures that I have only heard in English films before.*
7. Intrinsic motivation (3 items): measuring how intrinsically motivated learners are to use the given platform for EFL learning. *Example: I watch films on Netflix because I am interested in this activity.*
8. Extrinsic motivation (4 items): measuring how extrinsically motivated learners are to use the given platform for EFL learning. *Example: I watch films on Netflix because my parents encourage me to do so.*

Data Collection and Data Analysis

Due to the COVID-19 pandemic, an online version of the instrument was sent to the secondary school. With the help of the principal, English teachers forwarded the form to their students. The language of the questionnaire was Hungarian, the participants' native language, to help

ensure that there were no misunderstandings. The questionnaire used for the present study was divided into four main sections: biography questions followed by statements about Netflix, Instagram, and TikTok. After filling in the items about themselves and their EFL learning experiences, participants had to indicate with a simple yes or no answer if they had a Netflix account. If yes, they answered a number of statements about the application; otherwise, the Google Form automatically directed them to the next section, where the same procedure was followed. There was not a single participant who did not use at least one of the applications.

Data collected using the Google Form was downloaded into an Excel file. The Statistical Package for the Social Sciences 25 (SPSS) software was used for the analysis. After descriptive statistics were calculated for the participant and app data, each application was further examined. In addition to checking for gender and age differences, *t* tests were run to see how the number of years spent learning English influenced the participants' opinions on the usefulness of social media in EFL learning. One of the most significant grouping variables was the number of years spent studying English. Students were put into two different groups: those who had been learning English for less than five years and those with five years or more of learning experience.

Results and Discussion

In the following section, each platform will be separately examined in regard to the previously introduced scales, as some participants did not have accounts on all of three of them. The application used by the most participants was Instagram (n = 85), followed by Netflix (n = 75), and finally TikTok (n = 57), which is used by more than half of the students participating in this study despite being the least used out of the three applications. Different statistical tests were run which analyzed those who had accounts on the given application. Table 2 shows how frequently students opened and used these apps. In the case of Netflix, 41% of the

participants used it on a weekly basis, with 51% of the students using it more frequently than that. As for Instagram, 64% of those participants who reported having an account claimed to open the application more than once a day, and only 10% used it less frequently than once a day. The descriptive statistics for TikTok were similar to that of Instagram, as most students used it more than once a day (70%), and 10% used it less than once a day. In the analysis, the answers provided by all of these participants were included.

Table 2

Distribution of Frequency of Application Use

	Once a month	Once a week	Once a day	More than once a day	Number of students using the application
Netflix	5	32	26	14	77
Instagram	2	7	21	55	85
TikTok	0	6	11	40	57

Each of the eight previously mentioned scales was examined for all three platforms. Almost all scales showed relatively high Cronbach's alpha values, meaning that with one exception all of them were reliable (see Table 3). In regard to the motivation scales, reliability analysis showed that the items for extrinsic motivation did not exhibit internal reliability. For this reason, only the results for intrinsic motivation will be presented in the following section. Both general and application-specific conclusions were drawn from the results.

Netflix

Although the participants reported rarely using Netflix for explicit language learning ($M = 2.86$, $SD = 0.88$), the results show that many found

Table 3

Descriptive Statistics and Reliability Analysis of Scales

Scales		*Cr. Alpha*	*M*	*SD*
Netflix				
	Implicit learning	.86	4.11	.54
	Explicit learning	.89	2.86	.88
	Noticing	.81	4.06	.72
	Vocabulary	.78	4.26	.62
	Listening skills	.76	4.24	.66
	Grammar knowledge	.79	3.66	.92
	Intrinsic motivation	.79	4.52	.67
	*Extrinsic motivation**	*.41*	*1.31*	*.44*
Instagram				
	Implicit learning	.93	3.28	.91
	Explicit learning	.90	1.92	.79
	Noticing	.95	2.93	1.15
	Vocabulary	.87	3.51	1.09
	Listening skills	.82	2.73	.98
	Grammar knowledge	.89	3.04	1.13
	Intrinsic motivation	.84	4.43	.87
	*Extrinsic motivation**	*.48*	*1.38*	*.51*
TikTok				
	Implicit learning	.92	3.47	.89
	Explicit learning	.88	2.09	.77
	Noticing	.93	3.33	1.18
	Vocabulary	.80	3.98	.86
	Listening skills	.74	3.58	.80
	Grammar knowledge	.74	2.97	1.00
	Intrinsic motivation	.73	4.56	.67
	*Extrinsic motivation**	*.61*	*1.11*	*.35*

*Unreliable scales

it helpful for implicit language learning (M = 4.11, SD = 0.54). When looking at the skills separately, vocabulary was the skill that learners thought Netflix enhances the most (M = 4.26, SD = 0.62), followed by listening skills (M = 4.24, SD = 0.66). While the participants generally agreed that grammar was the aspect of the language that they were the least likely to learn from movies, they showed agreement regarding the possible benefits that watching Netflix has on the development of grammar knowledge (M = 3.66, SD = 0.92).

An independent samples t-test was conducted to examine differences between groups based on age, gender, and number of years learning English. No significant differences were found in the case of gender and age. As previously mentioned, significant differences were observed between students learning English for less than five years and for five years or more (Table 4). In addition to being more likely to consider Netflix as a great tool for implicit learning, students with more than five years of EFL experience reported noticing implicit English learning with the help of Netflix. Those with less than five years of experience, although they appear to believe that Netflix helped them in implicit learning, did not report noticing how much they usually learn. Both groups found Netflix to be beneficial for learning and practicing all of the mentioned skills, but more experienced learners recognized its importance in enhancing grammar knowledge to a greater extent compared to the other group.

As for motivation, 90% of students did not consider watching films in English as a task that they had to do; instead, they mainly reported engaging in this activity for entertainment purposes. The intrinsic motivation scale was found to have a relatively high mean value (M = 4.52, SD = 0.67), which means that students agreed that they used this platform because it made them feel relaxed and entertained. More than 90% of the participants did not use Netflix in class nor were they asked to watch films in English as a homework assignment.

Table 4

Comparison of the Means of Students' Opinions About Netflix Using Independent Samples T-test

	Less than 5 years		5 years or more		*t*	*p*
	M	*SD*	*M*	*SD*		
Implicit learning	3.89	0.54	4.18	0.54	-2.21	.031
Explicit learning	2.91	0.65	2.76	0.99	ns	ns
Noticing	3.88	0.57	4.27	0.65	-2.54	.013
Vocabulary	4.05	0.69	4.3	0.63	ns	ns
Listening skills	4.12	0.58	4.24	0.78	ns	ns
Grammar knowledge	3.31	1.01	3.85	0.82	-2.48	.015

Instagram

Descriptive statistics showed that, similarly to Netflix, the participants found Instagram to be a beneficial tool for implicit language learning ($M = 3.28$, $SD = 0.91$). Regarding the four language skills, participants believed that this application was mostly beneficial for building vocabulary ($M = 3.51$, $SD = 1.09$). As for listening ($M = 2.73$, $SD = 0.98$) and grammar improvement ($M = 3.04$, $SD = 1.13$), the participants did not seem to find this platform as useful as Netflix.

Although Instagram was used by 60 female and 39 male participants, there were no significant differences in how the two genders used the app; all of the participants reported posting and commenting about the same amount. However, there were a number of differences in how the two genders viewed Instagram as a tool for EFL learning (Table 5). Females found it to be a more useful tool for implicit language learning than males did; they were also keener to notice the potential benefits of Instagram on EFL learning, finding it more useful in connection with all of the skills examined in the study. A qualitative study would allow these differences to be investigated further.

Table 5

Comparison of the Means of Students' Opinions About Instagram Using Independent-Samples T-tests

	Female		Male		*t*	*p*
	M	*SD*	*M*	*SD*		
Implicit learning	3.47	0.93	2.87	0.86	2.92	.005
Explicit learning	1.97	0.87	1.96	0.70	ns	ns
Noticing	3.26	1.10	2.76	0.98	2.12	.038
Vocabulary	3.67	0.96	3.23	1.12	1.93	.057
Listening skills	3.36	1.22	2.74	1.15	2.26	.026
Grammar knowledge	3.23	1.09	2.44	0.94	3.35	.001

More than half of the students who claimed to have an Instagram account (55%) reported following at least one person who provides study materials and tips on how to learn English in their posts. High levels of intrinsic motivation were reported in connection with this application (M = 4.43, SD = 0.87); students appear to use Instagram because they find it interesting, engaging, and relaxing. Nobody reported being encouraged to use Instagram by parents or teachers.

TikTok

The teenage participants found TikTok to be beneficial for implicit EFL learning (M = 3.47, SD = 0.89). In regard to language skills, TikTok was also considered to be most useful for learning new vocabulary items (M = 3.98, SD = 0.86), as well as improving listening skills (M = 3.58, SD = 1.19) and grammar knowledge (M = 2.97, SD = 1.00).

When comparing male and female participants, TikTok was used for the same purposes: mainly for watching and liking videos as well as commenting on them. However, as in the case of Instagram, males and females had different views concerning the usefulness of the platform for

implicit EFL learning (Table 6). In general, females found the application to be more useful than males did, and this was the case for other skills as well, such as vocabulary learning and practicing grammar.

Table 6

Comparison of the Means of Students' Opinions About TikTok Using Independent Samples T-tests

	Female		Male		*t*	*p*
	M	*SD*	*M*	*SD*		
Implicit learning	3.8	0.83	3.16	0.67	2.71	.009
Explicit learning	2.16	0.83	2.01	0.6	ns	ns
Noticing	4.33	1.3	3.78	1.25	ns	ns
Vocabulary	4.17	0.85	3.59	0.82	2.29	.026
Listening skills	3.74	1.21	3.25	1.18	ns	ns
Grammar knowledge	3.32	1.02	2.56	0.89	2.6	.012

When examining how the number of years spent learning EFL influenced their use of the application, no significant differences were found. In addition, students reported rarely using TikTok for explicit language learning ($M = 2.09$, $SD = 0.77$). Based on their self-reports, learners' main reason for using the application was intrinsic motivation ($M = 4.56$, $SD = 0.67$), as they found the platform entertaining.

Conclusion

Based on the results, it can be concluded that using Netflix, Instagram, and TikTok in EFL teaching can present a number of potential learning opportunities. Learners found that the three online platforms and applications examined in this study are beneficial tools for EFL learning. Those students who have been learning EFL for at least five years find Netflix more useful than those who have been learning for less than five years. This might be due to the fact that Netflix is a streaming platform,

and for lower proficiency level students, watching films might be too challenging. This, however, can be changed with the help of a facilitator who can guide learners on what programs to watch and how. Regarding the two other social media applications, Instagram and TikTok, females found them more useful for EFL learning than males. Although male students do not appear to use these applications for EFL learning, with engaging tasks they too can be encouraged to discover how social media can be beneficial for their language learning endeavors.

However, students in the study reported that teachers do not see social media as something beneficial to EFL teaching. Although teenagers almost exclusively use these applications for entertainment, they also believe that they learn language implicitly in the process, which shows their effectiveness. Combining classroom EFL teaching with the use of social media can be challenging, but with careful planning and further research, it can be possible to determine what EFL skills social media has the potential to enhance. Moreover, a qualitative study could also be considered in order to investigate the factors behind the results. Students could be asked about their preferences and choices, and the reasons behind the observed gender differences could be further examined in this way. Another aspect of interest would be the comparison of the three apps, specifically which ones students think are the best suited for different skills and how parents and teachers view the benefits of these applications for EFL teaching and learning. Lastly, considering the reliability issue that arose in this pilot study, the extrinsic motivation scale needs to be improved in order to be examined in future research.

References

Alkathiri, L. A. (2019). Students' perspectives towards using YouTube in improving EFL learners' motivation to speak. *Journal of Education and Culture Studies, 3*(1), 12–30. http://dx.doi.org/10.22158/jecs.v3n1p12

Allam, M., & Elyas, T. (2016). Perceptions of using social media as an ELT tool among EFL teachers in the Saudi context. *English Language Teaching, 9*(7), 1–9. http://dx.doi.org/10.5539/elt.v9n7p1

Arif, T. Z. Z. A. (2019). The use of social media for English language learning: An exploratory study of EFL university students. *Metathesis: Journal of English Language Literature and Teaching, 3*(2), 224–233. https://doi.org/10.31002/metathesis.v3i2.1921

Berry, D. C., & Broadbent, D. E. (1988). Interactive tasks and the implicit-explicit distinction. *British Journal of Psychology, 79*, 251–272. https://doi.org/10.1111/j.2044-8295.1988.tb02286.x

Brutin, A., López, M. G., & Mesa, R. E. (2011). Incidental L2 vocabulary learning: An impracticable term? *TESOL Quarterly, 45*(4), 759–768. http:/doi.org/10.5054/tq.2011.268061

Csíkszentmihályi, M. (1985). *Beyond boredom and anxiety*. Jossey-Bass Publishers.

DeKeyser, R. (1994). Implicit and explicit learning of L2 grammar: A pilot study. *TESOL Quarterly, 28*(1), 188–194. https://doi.org/10.2307/3587210

Dörnyei, Z. (1994). Motivation and motivating in the foreign language classroom. *The Modern Language Journal, 78*(3), 273–284. https://doi.org/10.2307/330107

Ellis, N. (2015). Implicit and explicit language learning: Their dynamic interface and complexity. In P. Rebuschat (Ed.) *Implicit and explicit learning of languages* (pp. 3–23). John Benjamins Publishing Company.

Giampieri, P. (2018). Spoken language features (and anomalies) in films for ESL classes. *Language Learning in Higher Education, 8*(2), 399–425. http:/doi.org/10.1515/cercles-2018-0022

Gonulal, T. (2019). The use of Instagram as a mobile-assisted language learning tool. *Contemporary Educational Technology, 10*(3), 309–323. https://doi.org/10.30935/cet.590108

Gorijan, B. (2014). The effect of movie subtitling on incidental vocabulary learning among EFL learners. *International Journal of Asian Science, 4*(9), 1013–1026. https://archive.aessweb.com/index.php/5007/article/view/2695

Handayani, F. (2016). Instagram as a teaching tool? Really? *Proceedings of the Fourth International Seminar on English Language Teaching (IELTS-4), 4*(1), 320–327. https://scholar.google.co.id/citations?user=CUHOP6QAAAAJ&hl=en

Hertel, T. J., & Harrington, S. (2015). Promoting cultural and linguistic competence with documentary film in Spanish. *Hispania 98*(3), 549–569. https://doi.org/10.1353/hpn.2015.0094

Iqbal, M. (2021, September 28). *TikTok revenue and usage statistics.* Business of Apps. https://www.businessofapps.com/data/tik-tok-statistics/

Kabooha, R. H. (2016) Using movies in EFL classrooms: A study conducted at the English Language Institute (ELI), King Abdul-Aziz University. *English Language Teaching, 9*(3), 248–257. http://dx.doi.org/10.5539/elt.v9n3p248

Khan, A. (2015). Using films in the ESL classroom to improve communication skills of non-native learners. *ELT Voices – International Journal for Teachers of English, 5*(4), 46–52. https://www.academia.edu/15220670/Using_films_in_the_ESL_classroom_to_improve_communication_skills_of_non-native_learners

Koolstra, C. M., & Beentjes, J. W. J. (1999). Children's vocabulary acquisition in a foreign language through watching subtitled television programs at home. *Educational Technology Research and Development, 47*(1), 51–60. https://doi.org/10.1007/BF02299476

Lai, C., Zhu, W., & Gong, G. (2015). Understanding the quality of out-of-class English learning. *TESOL Quarterly, 49*(2), 278–308. https://doi.org/10.1002/tesq.171

Masrai, A., & Milton, J. (2018). The role of informal learning activities in improving L2 lexical access and acquisition in L1 Arabic speakers learning EFL. *The Language Learning Journal, 34*(1), 1–11. https://doi.org/10.1080/09571736.2018.1520655

Navidinia, H., Mobraki, M., & Malezkadeh, F. (2019). Investigating the effect of noticing on EFL students' speaking accuracy. *International Journal of Instruction, 12*(1), 83–98. https://doi.org/10.29333/iji.2019.1216a

Pamungkas, I. B. A., & Adi, S. S. (2020). Students' perception about improving English listening skills using movies among the vocational high school students. *ERUDIO Journal of Educational Innovation, 7*(2), 128–138. https://doi.org/10.18551/erudio.7-2.5

Panda, N. (2020). Teaching English language skills of b. tech students using Netflix as a tool. *Mukt Shabd Journal, 9*(5), 2563–2568. http://doi.org/10.13140/RG.2.2.28833.02406

Parnrod, U., Sinraksa, D., Nung-alee, A., Yacharat, A., Junnuan, C., Laptawewong, T., & Yeera, T. (2020). Using M-learning to enhance EFL students' English language skills. *International Journal of Advanced Research in Technology and Innovation, 2*(4), 29–36. http://myjms.mohe.gov.my/index.php/ijarti

Reber, A. S., Kassin S. M., Lewic, S., & Cantor, G. (1980). On the relationship between implicit and explicit modes in the learning of a complex rule structure. *Journal of Experimental Psychology, 6*(5), 492–502. http://doi.org/10.1037/0278-7393.6.5.492

Rebuschat, R. (2015). Implicit and explicit learning of languages. In R. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. xiii–xxii). John Benjamins Publishing Company.

Rodgers, M. P. H. (2013). *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and captions* [Unpublished Doctoral dissertation]. Victoria University of Wellington. http://hdl.handle.net/10063/2870

Rouxel-Cubberley, N. (2014). The film trailer project: French films as textbooks. *The French Review, 88*(1), 117–133. https://www.jstor.org/stable/24547698

Schmidt, R. W. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan, & I. Walker (Eds.), *Proceedings of CLaSIC 2010, Singapore, December 2-4* (pp. 721–737). National University of Singapore, Centre for Language Studies.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Secules, T., Herron, C., & Tomasello, M. (1992). The effect of video context on foreign language learning. *The Modern Language Journal, 76*(4), 480–490. https://doi.org/10.2307/330049

Spithill, A. C. (1980). Motivation and language teaching. *Hispania, 63*(1), 72–76. https://doi.org/10.2307/340815

Webb, S. (2010). A corpus driven study of the potential for vocabulary learning through watching movies. *International Journal of Corpus Linguistics, 14*(4), 497–519. https://doi.org/ 10.1075/ijcl.15.4.03web

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition, 27*(2), 269–304. http://doi.org/10.1017/S0272263105050138

Yang, H. (2020). Secondary-school students' perspectives of utilizing Tik Tok for English learning in and beyond the EFL classroom. *2020 3rd International Conference on Education Technology and Social Science,* 162–183. http://clausiuspress.com/conferences/AETP/ETSS%202020/G7605.pdf

6

Individual Differences Revisited

Éva Illés
illes.eva@btk.elte.hu

Abstract

The present theoretical study aims to provide a critical overview of the main trends and developments in the conceptualization and research of individual differences (ID) in second language acquisition (SLA). Similarly to approaches in natural sciences, in SLA, too, there has been a shift from the perception of IDs as relatively stable, learner-internal and monolithic notions to a view of IDs comprising complex dynamic systems that interact with each other and their environment. It is argued that a dynamic conceptualization of IDs better reflects the reality of the classroom, and it keys in with one of the main goals of education, which is developing learners' skills, abilities, and dispositions by ways and means that cater for learners' individual needs. A move towards the classroom can be seen in the implementation of research findings in language pedagogy. Whereas earlier the flow of information was from researchers to teachers, recently there have been attempts to utilize teachers' experiential knowledge and feed it back to the shaping of theory, thus forging a more balanced and fruitful relationship between practitioners and researchers.

Keywords: Second Language Acquisition; state vs. trait; aptitude; motivation; complex dynamic systems; situated, classroom-based approach

DOI: 10.21862/ELTE.DEAL.2022.6.141

Individual Differences Revisited

Genius is one percent inspiration
and 99 percent perspiration.
(Thomas Edison)

The title of this paper reflects the steadily increasing concern with Individual differences (ID) in both second language studies and language pedagogy. Dörnyei and Ryan (2015) attributed this growing interest to the widespread learning of foreign languages, resulting from population mobility and mass education. In the narrower context of language teaching, especially English language teaching (ELT), the sharper focus on IDs may also stem from the importance attached to learner-centredness in Communicative Language Teaching (CLT), which has led to the acknowledgement of the fact that learners differ in many ways and have distinct needs and learning paths (Xiaoju, 1990). In effect, learner-centredness and the concern with IDs in CLT were transferred to the practice of teaching in the form of basic principles very early on, such as the ones outlined in an undeservedly forgotten book, *Practical Techniques for language teaching*, by Lewis and Hill (1992). These, still highly relevant principles include, among others, the following advice: "Teach the students, not the book" (p. 8), "Involve students in the learning process" (p. 9), and "Vary what you do, and how you do it" (p. 13).

Research into individual differences gained momentum with studies mushrooming and leading to developments that seem to replicate how science has progressed in various disciplines, including humanities. A case in point for the latter is English as a lingua franca (ELF) research which started out with the traditional view of ELF as a possible codifiable variety (Jenkins et al., 2001; Seidlhofer, 2001) that can be captured with the help of corpora. However, further research revealed that given its spread and the number and diversity of its speakers, ELF cannot be perceived as the stable construct of variety and should rather be conceptualized in pragmatic terms as a particular type of communication characterized by

dynamicity and variability (Illés, 2020; Jenkins, 2015). Similarly, the first phase of ID research aimed to "identify those learner characteristics that have the most significant effect on learning outcomes and then to analyze the specific effects of particular characteristics" (Dörnyei & Ryan, 2015, p. 5). The initial perception of IDs as stable learner characteristics has then been replaced by the view of IDs as changing, evolving constructs, the examination of which requires a Complex Dynamic Systems Theory (CDST) approach that can more aptly reveal the true nature of IDs.

Research into IDs has been instigated by the need to make language teaching more effective through a better understanding of the key participant, the learner. Therefore, the connection between research and practice has been of particular importance. Researchers have been investigating issues arising in the classroom and then reconnecting with the classroom through the exploration of the implications of their findings for the practice of teaching. However, there have been dissenting voices regarding the cooperation of researchers and practitioners. Medgyes (2017), for instance, claimed that the relationship between academics and teachers is not a happy and mutually beneficial one, and that there are reasons why advancements in science filter through to the classroom slowly, if at all. One such reason is the assumed hierarchical relationship between researchers and teachers, where researchers occupy the upper echelons (Illés, 2012, 2016)—except in the case of action research, where the researcher and the teacher is the same person. Another one is the fact that when theory is applied, the descriptive purpose and modus operandi of scientific inquiry becomes prescriptive as soon as its advancements find their way into language pedagogy.

This article intends to address the two problems outlined above: the development of the conceptualization and research of IDs, and the ways ID research has been applied in language teaching. In both cases, the focus is on some of the contentious issues which provide food for thought for researchers and teachers. As a corollary, questions concerning the relationship between research and the practice of language teaching will also be addressed.

ID Conceptualization and Research

Issues in Conceptualization

As Dörnyei and Ryan (2015) pointed out, the first step was the identification of those learner features that contribute to the effectiveness and outcome of language learning. Individual differences thus traditionally include variables such as intelligence, aptitude, motivation, age, learning styles, and learning strategies. With time, other individual differences, such as social identity, gender, ethnicity (Li et al., 2022), and emotions (Csizér & Albert, 2021; Dörnyei & Ryan, 2015) have been added and deemed relevant to language teaching and learning. In order to provide a wider framework, Kormos and Sáfár (2008) categorized IDs into three groups: affective (e.g., motivation, language learning anxiety), cognitive (e.g., intelligence, foreign language aptitude), and personality-related (e.g., extraversion, emotional stability). Some IDs, such as motivation, have attracted considerable scholarly interest, while others, like identity and creativity, have come to the fore recently (Csizér & Albert, 2021).

Whereas there is general consensus about the identification of IDs, difficulties and differences arise when it comes to the definition of ID variables. A case in point is aptitude, which looms large in both lay and professional discourse. Dörnyei (2005) criticized the convenient but atheoretical conception that "language aptitude is what language aptitude tests measure" (p. 35), which may entail disputable post-hoc conceptualizations of the underlying construct (Zólyomi, 2022). In line with this, Carroll's (1990) identification of the four components of aptitude—phonic coding ability, inductive language learning ability, grammatical sensitivity, and rote learning ability—was based on a previously developed Modern Language Aptitude Test (MLAT) test. Interestingly, despite the fact that designing an instrument preceding the conceptualization of language aptitude is and can be seen as a weakness, it has not stopped researchers from using and validating various aptitude

tests (Zólyomi, 2022). Such an approach then poses the commonsensical question of how it is possible to measure a previously undefined construct. In more technical parlance, post-hoc conceptualization raises the issue of construct validity (Zólyomi, 2022), whose definition stresses the sound understanding of the concept to be assessed: "In this type of validity agreement is sought on the 'operationalized' forms of a construct, clarifying what we mean when we use this construct (Cohen et al., 2007, p. 138).

Other delineations of aptitude, such as the one below seem less circular than the one discussed above:

> Aptitude as a concept corresponds to the notion that in approaching a particular learning task or program, the individual may be thought of as possessing some current state of capability of learning that task—if the individual is motivated, and has the opportunity of doing so. That capability is presumed to depend on some combination of *more or less enduring characteristics* [emphasis added] of the individual. (Carroll, 1981, p. 84)

The perception of aptitude above reflects what Dörnyei and Ryan (2015) called the classic ID paradigm, the view of ID variables as "distinctly definable," "relatively stable," and "learner-internal (p. 6) psychological constructs. This is in contrast with the recent conceptualization of IDs as complex, interacting, and socially interdependent attributes which change in relation to both context and time. These opposing paradigms, by and large, represent the trait and state dichotomy in ID research. While traits are dispositions which remain mostly unchanged in an individual's life over time and comprise general tendencies in a person's behavior, states are ephemeral, and display temporal and situational variation (Csizér & Albert, 2021). Csizér and Albert (2021) likened the trait and state dichotomy to the nature vs. nurture conundrum which, as will be seen below, gains particular relevance in the context of teaching.

The investigation of anxiety as a trait is represented by the Foreign Language Classroom Anxiety Scale (Horwitz et al., 1986) used to identify anxious foreign language learners. The instrument allows researchers not only to categorize learners as high- or low-anxious, but it can also reveal those aspects of the foreign language learning situation which give rise to particular levels of anxiety. This view also results in generalizations, such as the one below:

> ... people who fear negative evaluation rarely initiate conversation and interact minimally. Language students who experience this anxiety tend to sit passively in the classroom, withdraw from activities that could increase their language skills, and may even avoid class entirely. (Gregersen & Horwitz, 2002, pp. 562–563)

However, there seems to be an inherent contradiction in claims like the one above, which is that the very term of *individual* differences runs counter to categorizing and attributing specific characteristics to learners belonging to a particular research-identified group. In addition, IDs perceived as permanent personal characteristics possessed by everybody and differing only in degree does not solve the question of "how to conceive of general laws or categories for describing human individuality that at the same time do justice to the full array of human uniqueness" (Dörnyei & Ryan, 2015, p. 3). In pedagogy, the compartmentalization of learners and the assumption of stable learner dispositions defeat the object since the overall aim of education is and should be making a difference by developing learners' skills and abilities in ways which suit their individual needs. In other words, teachers need to believe in the power of nurture, or their efforts lack purpose and direction.

Csizér and Albert (2021) solved the trait vs. state conundrum by reconciling the two. They argued that the dichotomy should be seen as trait and state perspectives which "inform language learning processes in their own unique way" (p. 343). This is a valid point but with the caveat that the two perspectives inform the *investigation* of IDs—by means of

employing either cross-sectional design to identify patterns and trends or adopting approaches that can capture the complexity of IDs as they occur in the actuality of the language learning process. The emphasis on investigation is meant to highlight the fact that of the two perspectives, it is the dynamic conceptualization of IDs that prevails in language classrooms. The reason for this is that research and teaching practice differ considerably. Whereas research aims to identify trends and patterns, and utilizes tested scientific means of research, the classroom, despite the necessary planning, abounds in unpredictable, one-off episodes that force teachers to make split-second decisions. The varying contexts require different types of knowledge. While researchers' knowledge is abstract, impersonal, and theoretical in nature, teachers possess more practical, personal, and context-specific knowledge (McIntyre, 2005). In crude terms, the researcher's task is to find order in the chaos classrooms present, whereas the teacher's job is to navigate and exploit the complex everyday challenges of teaching for their students' benefit.

In research, the perception of IDs as relatively stable, monolithic, and context-independent human characteristics was challenged early on. Referring to motivation, Ellis and Larsen-Freeman (2006) claimed that it is "less a trait than fluid play, an ever changing one that emerges from the processes of interaction of many agents, internal and external, in the ever-changing complex world of the learners" (p. 563). Dörnyei (2009) followed suit and debunked what he called the prevailing "individual differences myth" (p. 188). He argued that IDs should be seen as complex dynamic systems which are neither stable, nor monolithic or free from the influence of the environment. Rather, IDs are multicomponential with different parts interacting with each other and their environment at a particular point in time but also over time. Like other complex dynamic systems, IDs are sensitive to initial conditions, and even small changes can result in disproportionate modifications both between and within IDs. This is something teachers intuitively know, that is, how a smile or a well-placed or well-timed question or feedback can make a huge difference in the motivation of a student in a particular lesson or, maybe, beyond.

In addition, the dynamic perspective stresses the impact of the environment, which, in the case of education, may include that of the school as well as the wider context. Thus, a student may be anxious not only because of a test or an exam but because of troubles in the family or the effects of problems pertaining to them in society. These and other factors and issues are what teachers need to know about and address on an individual basis when applying IDs in their classroom practice.

Rather than compartmentalizing and labelling them, Dörnyei (2009) presented IDs as attractor states in a weblike changing system, which comprise times of relative stability—as pointed out by Wind and Harding (2020, pp. 135–136):

> Sometimes, complex systems show a great deal of variability and change over time. However, there are certain periods of inherent stability when only stronger external forces might cause a change in the system. A great deal of variability might be the harbinger of change in a complex system, while little intra-variability can be a sign of system stability. States of less variability are called 'attractor states'.

These relatively stable phases may undergo change as a consequence of a disturbance or more forceful push from the outside (Dörnyei, 2009). For example, a praise at the right time and context can push a learner out of their current state of uninterestedness and engage them in what is going on in class. If that push is strong enough and the state of interest remains, the system reorganizes itself and arrives at a new equilibrium and attractor state, resulting in continued interest on the part of the learner. In the case of a group, such an attractor state can occur when, as the result of the interplay of many internal and external influences, the class stabilizes "into a coherent group and a discernible pattern of behaviour emerge[s]" (Hiver, 2015, p. 21). Dörnyei (2009) saw the identification of potential attractors as a possible direction for future research but warned that while

investigating them, researchers should not fall back on the traditional search for linear cause and effect relations between IDs.

Issues in Research

Even though IDs perceived as complex dynamic systems represent the reality of classroom experience, the approach presents problems not so much in terms of conceptualization but, rather, regarding how to research such interdependent, interacting, and dynamic systems and subsystems.

Traditional research is based on trait-like conception of IDs, which assumes, among others, single causes, linear causality, and clearly-definable categories and labels (Schumann, 2015). The result is snapshots of different IDs—pictures frozen in time and space; lacking change, variation, or messy 'noises'. In this paradigm, it is often a quantitative approach that is taken to the investigation of IDs. Dörnyei (2009) argued that "traditional quantitative statistical methods of data analysis are inadequate to be used with dynamic systems because these procedures are based on group averages and thus iron out idiosyncratic details" (p. 107). The suggestion is the use of a qualitative or mixed methods approach, longitudinal research, and modelling as in natural sciences.

There have been attempts to develop methods that can be employed to capture IDs in their entirety and inherent dynamicity. One such endeavor is Csizér and Piniel's (2015) longitudinal study of a group of 21 students in an academic writing seminar at a Hungarian university. The theoretical background comprises Dörnyei's (2009) tripartite macrosystem consisting of cognitive, affective, and motivational IDs factors, and the L2 Motivational Self System. The researchers used both quantitative (questionnaires) and qualitative (short essay) data sources. Csizér and Piniel (2015) adopted a "simplistic-complexity approach, where micro-social phenomena are under scrutiny with the possibility of drawing macro-social implications" (p. 165). The research has yielded findings about the fluctuations, albeit small ones, in the participants' motivation, anxiety, and self-efficacy. The relative stability of most factors

under scrutiny has been explained by the fact that the course was compulsory and the participating students were motivated and had well-developed ideal L2 selves. Overall, the study managed to provide a relatively accurate picture of the students' dispositions over a 14-week period of time. It, however, was unable to throw light either on the role and influence many other factors may have had on the participants' dispositions, or on the finer details or interplay of interdependent ID variables. More importantly, however, even though the study no doubt contributes to the development of suitable methods for complexity research in humanities, it has very little to offer to teachers, including the academic instructors of the particular course at the Hungarian university.

To realize a better connectedness of ID research with classroom practice in primary and secondary education, an alternative approach proposed by Henry et al. (2019) may prove to be fruitful. Rather than starting with the micro-social factors as in the study by Csizér and Piniel (2015), in order to gain insights for the classroom, the macro-social context was examined with primary focus on learners' English language use and simultaneous acquisition outside of school in Sweden. The locally specific and relevant question researchers aimed to answer was how students could be motivated to achieve higher levels of proficiency when English became an integral part of their everyday life. The issue then was not what features of motivation were prevalent in the Swedish context but what motivational methods, tasks, and materials resulted in increased motivation and active engagement in the good practice of selected school teachers. The interest in the classroom was reflected in the choice of theories which were judged more pertinent to the specific Swedish context than the most widely applied L2 motivational Self System or Gardner's Socio Educational Model (Henry et al., 2019). Given the aim of gaining insights from the classroom, the theories included flow, engagement, willingness to communicate, and directed motivational currents (Henry et al., 2019). The reason for not drawing on the most widely employed theories was as follows:

> … while most of these theories are important in providing generalizable understandings of motivational dispositions, and insights into the role of various cognitive and affective factors that influence an individual's learning goals and the effort invested in the learning process, they have surprisingly little to say about what goes on in language classrooms, or about the social interactions within which motivation can emerge. (Henry, 2019, p. 44)

Therefore, while in research the question of what happens regarding motivation in the process of language acquisition is addressed, in Henry et al. (2019) the issue is how motivation can be increased and what theoretical explanations underlie the success of particular practices in a specific locality.

Second Language Acquisition (SLA)

A further and rarely raised issue is SLA, the field of inquiry within which IDs are investigated. The term implies that the learner acquired their first language, which is then followed by the learning of a second language which—as extensive research shows—is predominantly English. The focus on English and on how it is learnt as a second language after the learners' mother tongue reflects a monolingual, Anglo-Saxon bias which has recently been challenged by ELF (e.g., Jenkins, 2015; Seidlhofer, 2011, 2017) and Third Language Acquisition (TLA) research. Given the diversity of speakers' linguacultural backgrounds, ELF is, by definition, a multilingual context of use: "ELF is bound to co-exist with other languages; it forms part of individual's bi- or multilingual repertoires" (Seidlhofer, 2011, p. 68). Apart from ELF, there is also a political reason why multilingualism has been brought into focus. As a consequence of EU targets, according to which EU citizens should speak two languages apart from their mother tongue, learning a third language and consequently multilingualism is on the rise in Europe. This, in return, calls for research in third language acquisition (TLA) which, researchers argue, is

significantly different from SLA (for more detail, see Gutiérrez Eugenio, 2018). The current conceptualization of ID variables as complex dynamic systems, therefore, should be related to the individual differences of learners acquiring not only two but additional languages as well.

IDs in Language Pedagogy

As has been noted above, ID is one of the areas within applied linguistics which is directly connected to classroom practice. In what follows, three approaches to the implementation of ID research are outlined.

One such approach entails researchers creating the connection between the two contexts by making suggestions for teachers on the basis of their findings. A well-known example is the ten commandments, a list of macrostrategies for motivating language learners by Dörnyei and Csizér (1998), which include, among others, "Present the tasks properly," "Make the language classes interesting," and "Familiarize learners with the target language culture" (p. 215). Even though their wording is cautious and includes the caveat that "[t]hese motivational strategies were intended to be broad recommendations rather than prescriptive rules that every teacher must observe in order to motivate their students" (Dörnyei & Csizér, 1998, p. 209), the fact remains that researchers propose guidelines for teachers who work in a context of which researchers do not have much experience. This matters because, as has been pointed out above, teaching and research differ considerably. While research necessarily implies distance from one's object of inquiry as well as the identification of patterns and regularities, the classroom is about the here and now, comprising a medley of predictable and unforeseeable events, and minute and changing details that require online and instantaneous decisions on the part of teachers (Illés, 2012; McIntyre, 2005). The guidelines are also problematic in the sense that teachers know, for example, that making their language classes interesting (Commandment No. 6, p. 215) motivates students. What they find challenging is how to make their classes interesting on a day-to-day, minute-by-minute basis.

Henry et al. (2019) referred to this as a problematic blind spot in research in that research fails to reveal how motivation ebbs and flows in the interaction between the classroom participants. This gap has, in fact, resulted in the call for situated approaches which have been developed since the publication of the ten commandments more than 20 years ago. The dominant use of English as a global lingua franca, on the other hand, has made the "familiarize learners with the target language culture" commandment obsolete since the number of cultures in which English is used is hard, if not impossible, to decide. Similarly, what counts as proper presentation of a task varies and depends on many factors, such as the students, the nature of the task, the time when the task is being presented, etc.

Another problem regarding the guidelines has been highlighted by the authors themselves:

> One weakness of this list was that it was not based on systematic research; rather, it was the result of a synthesis of personal experience and a semi-formal survey amongst two groups of graduate students and a group of international teachers on a British Council summer course (Dörnyei & Csizér, 1998, p. 209).

Even when recommendations are based on more reliable and systematic inquiry, they are still generated using findings of research with a relatively small number of participants representing a particular context at a particular time. This being the case, any claim to universal applicability is highly questionable. In fact, the results of empirical studies based on Dörnyei's (2001) more comprehensive and detailed taxonomy of motivational strategies consisting of 102 micro- and 32 macrostrategies indicated that few macrostrategies apply widely, and that there is not much agreement about the importance attached to various macrostrategies in different contexts (Lamb, 2017).

A second approach to creating a link between ID research and practice is providing a critical summary of relevant studies, be they

theoretical or empirical, to enable teachers to make informed decisions when applying research in their particular practice. This approach engages teachers as active agents who should be able to appraise and reappraise what has been offered to them by researchers, rather than being treated as sheer "consumers of research" (Widdowson, 1984, p. 90).

The publications taking such an approach include the book *How languages are learned* by Lightbown and Spada, which was first published in 1994 and is currently in its fifth edition (2021). One of the strengths of the books is that the discussion of how languages are learned is based on the findings of relevant and cutting-edge research at the time of publication. The overall aim seems to be equipping teaching professionals with the knowledge that enables them to critically evaluate theory in relation to their classroom practice.

In what follows, the question of how the changes in the conceptualization and research methodology of IDs are reflected in different editions of *How languages are learned* is addressed. This section focuses on the second (1999), third (2006) and fifth (2021) editions of the book.

In each edition, there is one chapter dedicated to IDs. In all three editions, ID variables are outlined separately with the list of ID variables remaining unchanged over time, except for one interesting omission: while intelligence features in the second and third editions, it was left out in the latest, 2021 one. Possible explanations may include the relative lack of research in this area, the complexity of IQ, and the contentious nature of IQ tests which assess "only a limited range of abilities" (Lightbown & Spada, 2006, p. 57). With regards to aptitude, the discussion becomes longer and more detailed over the years. Although there are overlaps in the three editions when it comes to the studies included, the way aptitude is captured reflects some of the changes in the way IDs have been perceived in ID research. Initially (1999, 2006) aptitude is described in relation to the components that feature in aptitude tests such as MLAT or Pimsleur Language Aptitude Battery (PLAB). The conclusion in 1999 and 2006 is that teachers should vary their teaching activities in order to better

cater for learners' aptitude profiles. The argument is that if students are in learning environments which are compatible with their aptitude profiles, they reach higher levels of attainment (2006). Although in the 2021 edition the section finishes with a 2012 study that highlights the fact that talent combines in successful language learners with hard work, overall, aptitude is still seen as a trait, and it is teaching approaches which are expected to adjust to learners' aptitude profiles rather than focus on the development of learners' abilities.

In all three editions, the section on motivation and attitudes consists of two parts, one containing a brief outline of the most significant theories, while the other one is concerned with motivation in classroom settings, with the latter providing fairly specific suggestions about how to increase the level of motivation for students. Interestingly, while the second and third editions finish the section with the caveat that cultural and age differences are likely to affect effectiveness in various teaching/learning contexts, the book published in 2021 mentions this in passing, seemingly attaching less importance to it. This edition, however, already makes reference to research influenced by Complex Dynamic System Theory, of which there is a brief outline in a subsequent section of the book.

In the 1999 edition, learning styles are outlined under the heading *Learner preferences,* but the term learning styles features in the more recent editions (2006, 2021). Interestingly, doubts about learning styles and their applicability in teaching practice arise only in the most recent edition: "Research has not found strong evidence for the value of 'teaching to' learning styles, or even clear evidence that they exist in any neurologically meaningful way" (2021, pp. 94–95). A similarly cautious and balanced view prevails in the discussion of age and language instruction: "Clearly the age in which instruction begins is not the only variable that determines success in the L2 classroom, and an early start is no guarantee of long-term success" (2021, p. 103). Overall, updating the book entailed not only the inclusion of new developments in ID research but also the changes, doubts and questions that have been raised with regard to the conceptualization of IDs in SLA research.

An approach similar to the *How language are learned* book has been taken by articles in the *ELT Journal*, which often aim to challenge established ways of thinking in ELT. Such articles are the ones, for example, that debunk long-standing myths about age (Abello-Contesse, 2008), aptitude (Mercer, 2012) and learning styles (Lethaby & Harries, 2015). In order to raise teachers' awareness, the authors juxtapose deeply-held beliefs—such as the younger the learner, the better the outcomes, or that catering for sensory learning styles enhances learning—with findings of scientific investigations. In the article on learning styles (Lethaby & Harries, 2015), for example, the comparison reveals that in teacher training myths are often perpetuated rather than being subjected to critical appraisal and resulting in a change in thinking about them. In the quoted articles, the authors stressed the complexity of the nature and interaction of IDs, and emphasized that the success of learning depends more on locally and individually appropriate learning conditions than on what are assumed to be innate abilities. The argument is that abilities should be seen as "dynamic, varied, and multidimensional" (Mercer, 2012, p. 28), which cannot be captured as fixed unitary notions as is often the case with aptitude or intelligence in ELT. Apart from highlighting the discrepancy between beliefs commonly held by ELT professionals, the highly relevant conclusion Mercer (2012) puts forward is that learners and teachers should adopt a growth mindset and "believe in the capacity of all of our learners to continually develop and further expand their language learning abilities" (p. 28). This message, in fact, confirms one of the basic tenets of pedagogy: If teachers do not believe in being able to make a difference, their efforts are deemed futile from the very beginning.

The third possible way of linking up research with practice is represented by Henry et al.'s study (2019) on L2 motivational practices in Sweden. By adopting a situated, classroom-based approach, the research aimed to fill the blind spot mentioned above in reference to the ten commandments by Dörnyei and Csizér (1998). According to Henry (2019), the need for research to be situated in the classroom arose because of the (necessary) limitations of quantitative, qualitative, and mixed methods

research, which cannot account for how motivation develops and fluctuates between teachers and learners in the actuality of classroom teaching. In addition, research methods relying on secondary information rather than the primary source of the classroom fail to draw on "teachers' experiential knowledge" which "far outweighs the results so far produced by L2 motivation research" (p. 37). The study by Henry et al. (2019), therefore, broke away from the tradition of investigating motivation by developing theoretical postulations for the design of instruments that are tested with research participants. Here, research took place in the classrooms of a selected group of teachers in Sweden, and the focus was on what these teachers did to motivate their students and what methods, tasks and activities proved motivating.

Of the three ways of linking up research with practice, it is this last one which seems to create the most direct connection between the two very different domains and, therefore, offers the most benefit to those for whom research is conducted in the first place: teachers working in particular educational contexts. However, the situated approach engages teachers not only as beneficiaries of research but also as active researchers and theory developers, on a par with their non-teaching counterparts. The study conducted in Sweden offers an excellent example of the cooperation between researchers and practitioners in service of public education and the improvement of pedagogical practices, in particular.

Conclusions

Drawing on the relevant literature, the paper aimed to provide a critical overview of the development in the conceptualization and research of IDs. The study of IDs follows the path also taken by natural sciences: from investigating phenomena in reference to their constituents and assuming a linear relationship between them to the conception of IDs as complex dynamic systems. Whereas the former conceptualization can be researched using traditional methods, for example with the adoption of a quantitative approach, the latter poses considerable problems in that it

requires new ways to grasp the interconnectedness, dynamism, and non-linearity within and between IDs as well as their environment. Developing a complex dynamic systems approach is all the more important as the lived experience of IDs in the actuality of the classroom is far from the idealizations and generalizations traditional ID research presents.

Similarly, in language pedagogy there have been different approaches to the application of research findings in practice. Initially, the flow of information was from researchers to practitioners where researchers advised their teaching colleagues on the pedagogical relevance and applicability of their investigations. Another approach to the implementation of developments in research has been presenting a critical overview, and letting the teachers decide how they apply the knowledge thus gained for making informed decisions in their practice. A third way of relating ID research to the classroom is realized through the cooperation of teachers and researchers, working on a par and contributing in equal measure to the improvement of local practices. It seems that a complex dynamic systems conceptualization coupled with insights from both research and the classroom can produce results that benefit those who are the target beneficiaries of research, that is, the person at the chalkface.

Note: This study was funded by the Scientific Foundations of Education Research Program of the Hungarian Academy of Sciences.

References

Abello-Contesse, C. (2008). Age and the critical period hypothesis. *ELT Journal*, *63*(2), 170–172. https://doi.org/10.1093/elt/ccn072

Carroll, J. B. (1981). Twenty-five years of research in foreign language aptitude. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 83–118). Newbury House.

Carroll, J. B. (1990). Cognitive abilities in foreign language aptitude: Then and now. In T. S. Parry & C. W. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 11–29). Prentice Hall.

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

Csizér, K., & Albert, Á. (2021). Trait and state perspectives of individual differences research. In T. Gregersen & S. Mercer (Eds.), *The Routledge handbook of the psychology of language learning and teaching* (pp. 337–347). Routledge.

Csizér, K., & Piniel, K. (2015). Changes in motivation, anxiety and self-efficacy during the course of an academic writing seminar. In Dörnyei, Z., MacIntyre, P. D., & Henry, A. (Eds.), *Motivational dynamics in language learning* (pp. 164–194). Multilingual Matters.

Dörnyei, Z. (2001). *Motivational strategies in the language classroom*. Cambridge University Press.

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

Dörnyei, Z. (2009). *The psychology of second language acquisition*. Oxford University Press.

Dörnyei, Z., & Csizér, K. (1998). Ten commandments for motivating language learners: Results of an empirical study. *Language Teaching Research*, *2*(3), 203–229.

Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited*. Routledge.

Ellis, N., & Larsen-Freeman, D. (2006). Language emergence: Implications for applied linguistics–Introduction to the special issue. *Applied Linguistics*, *27*(4), 538–589. https://doi.org/10.1093/applin/aml028

Gregersen, T., & Horwitz, E. K. (2002). Language learning and perfectionism: Anxious and non-anxious language learners' reactions to their own oral performance. *The Modern Language Journal*, *86*(4), 562–570. https://doi.org/10.1111/1540-4781.00161

Gutiérrez Eugenio, E. (2018). *L3 teachers' beliefs about multilingualism in Europe.* Unpublished doctoral dissertation. Eötvös Loránd University.

Henry, A. (2019). Young people and English. In A. Henry, P. Sundovist, & C. Thorsen (Eds.), *Motivational practice: Insights from the classroom* (pp. 23–41). Studentlitteratur.

Henry, A., Sundqvist, P., & Thorsen, C. (2019). *Motivational practice: Insights from the classroom*. Studentlitteratur.

Hiver, P. (2015). Attractor states. In Dörnyei, Z., MacIntyre, P. D., & Henry, A. (Eds.), *Motivational dynamics in language learning* (pp. 20–28). Multilingual Matters.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, *70*, 125–132.

Illés, É. (2012). Applied linguistics: Who is it for? In É. Illés & T. Eitler (Eds.), *Studies in applied linguistics: In honour of Edit H. Kontra* (pp. 65–73). ELTE, BTK.

Illés, É. (2016). Issues in ELF-aware teacher education. *Journal of English as a Lingua Franca*, *5*(1), 135–145. https://doi.org/10.1515/jelf-2016-0006

Illés, É. (2020). *Understanding context in language use and teaching: An ELF perspective*. Routledge.

Jenkins, J. (2015). Repositioning English and multilingualism in English as a lingua franca. *Englishes in Practice*, *2*(3), 49–85. https://doi.org/10.1515/eip2015-0003

Jenkins, J., Modiano, M., & Seidlhofer, B. (2001). Euro-English. *English Today*, *17*(4), 13–19. https://doi.org/10.1017/S0266078401004023

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition*, *11*(2), 261–271. https://doi.org/10.1017/S1366728908003416

Lamb, M. (2017). The motivational dimension of language teaching. *Language Teaching*, *50*(3), 301–346. https://doi.org/10.1017/S0261444817000088

Lethaby, C., & Harries, P. (2015). Learning styles and teacher training: Are we perpetuating neuromyths? *ELT Journal*, *70*(1), 16–27. https://doi.org/10.1093/elt/ccv051

Lewis, M., & Hill, J. (1992). *Practical techniques for language teaching* (4th ed.). Language Teaching Publications.

Li, S., Hiver, P., & Papi, M. (2022). (Eds.). *The Routledge handbook of second language acquisition and individual differences*. Routledge.

Lightbown, P. M., & Spada, N. (1999). *How languages are learnt* (2nd ed.). Oxford University Press.

Lightbown, P. M., & Spada, N. (2006). *How languages are learnt* (3rd ed.). Oxford University Press.

Lightbown, P. M., & Spada, N. (2021). *How languages are learnt* (5th ed.). Oxford University Press.

McIntyre, D. (2005). Bridging the gap between research and practice. *Cambridge Journal of Education*, *35*(3), 357–382. https://doi.org/10.1080/03057640500319065

Medgyes, P. (2017). The (ir)relevance of academic research for the language teacher. *ELT Journal*, *71*(4), 491–498. https://doi.org/10.1093/elt/ccx034

Mercer, S. (2012). Dispelling the myth of the natural-born linguist. *ELT Journal*, *66*(10), 22–29. https://doi.org/10.1093/elt/ccr022

Schumann, J. H. (2015). Foreword. In Dörnyei, Z., MacIntyre, P. D., & Henry, A. (Eds.), *Motivational dynamics in language learning* (pp. xv–xix). Multilingual Matters.

Seidlhofer, B. (2001). Closing a conceptual gap: The case for a description of English as a lingua franca. *International Journal of Applied Linguistics*, *11*(2), 133–158. https://doi.org/10.1111/1473-4192.00011

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

Seidlhofer, B. (2017). English as a lingua franca and multilingualism. In Cenoz, J. et al. (Eds.), *Language awareness and multilingualism* (pp. 391–404). Springer.

Widdowson, H. G. (1984). The incentive value of theory in teacher education. *ELT Journal*, *38*(2), 86–90.

Wind, A. M., & Harding, L. W. (2020). Attractor states in the development of linguistic complexity in second language writing and the role of self-regulation: A longitudinal case study. In W. Lowie, M. Michel, A. Rousse-Malpat, M. Keijzer, & R. Steinkrauss (Eds.), *Usage-based dynamics in second language development* (pp. 130–154). Multilingual Matters. https://doi.org/10.21832/9781788925259-009

Xiaoju, L. (1990). In defence of the communicative approach. In R. Rossner & R. Bolitho (Eds.), *Currents of change in English language teaching* (pp. 59–73). Oxford University Press.

Zólyomi, A. (2022). Investigating language aptitude in the success of adult second language acquisition. *Journal of Adult Learning, Knowledge and Innovation*, *5*(1), 1–11. https://doi.org/10.1556/2059.2020.00001

7

Teacher Trainees' Self-Efficacy Beliefs in Light of Their Perceived Language Aptitude and Explicit—Implicit Language Learning Behavior

Anna Zólyomi
zolyomi.anna@btk.elte.hu

Abstract

The causes leading to early teacher attrition may be numerous; however, studies have shown that low self-efficacy beliefs as well as too high initial self-efficacy beliefs may result in teachers leaving the profession. Self-efficacy beliefs are rarely investigated along with individual differences, and even if they are, these endeavors are limited to studying the relationship between self-efficacy beliefs and proficiency. This exploratory study aims to examine the factor structure and reliability of an instrument that is intended to measure self-efficacy beliefs, perceived language aptitude, and explicit–implicit learning behavior to be used in the future for studying teachers' self-efficacy beliefs. In addition, it also investigates the relationship between self-efficacy beliefs and perceived language aptitude. Thirdly, participants' explicit–implicit behavioral profiles are studied to assess whether there are any significant intergroup differences concerning self-efficacy. To this end, 62 teacher trainees filled in an online questionnaire, and the data was subjected to exploratory factor analysis, Pearson correlations, and cluster analysis. The results show a factor structure of nine scales that can reliably measure the proposed constructs; perceived language aptitude appears to account for 35% of variance in self-efficacy beliefs; and based on the teacher trainees' profiles, those who employ both explicit and implicit learning behaviors have higher self-efficacy beliefs. The results also lend support to the single-system approach of explicit–implicit learning, meaning that these processes are

DOI: 10.21862/ELTE.DEAL.2022.7.163

not completely independent. Pedagogical implications that point to the role of the interplay of explicit and implicit learning are also discussed.

Keywords: self-efficacy beliefs, perceived language aptitude, explicit–implicit learning, pre-service teachers, exploratory factor analysis

Teacher Trainees' Self-efficacy Beliefs in Light of Their Perceived Language Aptitude and Explicit—Implicit Language Learning Habits

To address the theme of this volume about challenges and opportunities, I would like to start by identifying the problem of early teacher attrition. Unfortunately, many teachers leave the profession after spending a couple of years or even one single year in the practicum because they suddenly face a reality shock (Pendergast et al., 2011). The final decision of dropping out, naturally, may be influenced by various factors; nevertheless, studies have shown that increasing teachers' self-efficacy might lead to a lower drop-out rate, which is exactly why it is important to develop teachers' self-efficacy at the outset (Pendergast et al., 2011; Yada et al., 2021). However, the line of challenges is further endorsed by the notion that extremely high initial self-efficacy beliefs may also lead to experiencing reality shock (Pendergast et al., 2011); therefore, instead of having to push a sense of self-efficacy beyond a threshold, it would be pivotal to maintain it at an optimal level.

Recently, in the Hungarian context, Csizér et al. (2021) have drawn attention to the fact that self-efficacy beliefs are rarely investigated alongside individual differences. Many studies have investigated the association between self-efficacy beliefs and perceived proficiency (e.g., Hoang & Wyatt, 2021); however, to my knowledge, none of these studies have examined the relationship between self-efficacy beliefs and perceived language aptitude. Nevertheless, investigating language aptitude is of utmost importance because it has shown great predictive power in language proficiency and achievement (Li, 2016). Recent studies have started to investigate the interplay of language aptitude and explicit–implicit encoding processes referred to as cognitive aptitudes for explicit and implicit learning (Granena, 2020; Granena & Yilmaz, 2019). Consequently, this study is addressing a research niche concerning the interrelationship between self-efficacy beliefs, language aptitude, and explicit–implicit learning.

In his recent systematic review, Hoang (2018) argued that most studies investigating teacher self-efficacy predominantly involve quantitative data collection procedures, and the participants are primarily practicing teachers and are mainly from Middle Eastern countries. Although following the tradition of quantitative data collection, to address Hoang's call, this study intends to explore teacher trainees' self-efficacy beliefs in the Hungarian context. Specifically, this exploratory study embarks on discovering the possible constructs that influence the self-efficacy beliefs of teachers who have not started their practicum yet. Even though this study alone is not an attempt to offer a ready-made solution to the issue of teacher attrition, it may shed light upon the possible relationships between self-efficacy, language aptitude, and explicit–implicit learning, and thus it may highlight the opportunities we have in the mission to keep our teachers in the profession. Therefore, the aim of this study is 1) to explore the underlying latent dimensions appearing in the questionnaire that intends to measure self-efficacy beliefs, perceived language aptitude, and explicit–implicit learning behavior, 2) to examine the relationship between teacher trainees' self-efficacy beliefs and perceived language aptitude, and 3) to discover the characteristics of teacher trainees who differ in their explicit–implicit learning behavior. By answering the first research question, I intend to develop an instrument that can reliably measure the proposed constructs so that it can be used and replicated in upcoming studies on self-efficacy beliefs, perceived language aptitude, and explicit–implicit learning behavior. While answering the second and the third research questions, I hope that I can discover whether perceived language aptitude is important in the formulation of self-efficacy beliefs and whether those participants who differ in their learning habits have different levels of self-efficacy. With the results of this study, I intend to inform fellow researchers, practicing teachers, and teacher trainers as well as teacher trainees as I would like to make recommendations for improving teacher training.

Theoretical Background

Teachers' beliefs are greatly influenced by their experiences (Borg, 2003), and three concepts need to be discussed when focusing on teacher trainees' beliefs and experiences. The notion of *apprenticeship of observation* (Lortie, 1975) postulates that earlier experiences have an impact on teachers' practices as they usually acquire and employ those characteristics that they witnessed from their teachers. This means that positive experiences may form certain characteristics of behavior in teachers-to-be where these experiences function as models or ideals. Its opposite notion, the *anti-apprenticeship of observation* (Moodie, 2016), posits that negative experiences trigger a behavior to avoid doing something as a teacher. This means that negative experiences may function as counterexamples. Lastly, the *apprenticeship of learning* (Pendergast et al., 2011) incorporates all the experiences teacher trainees have during their years of studying in teacher training. Thus, the experiences mentioned above weigh considerably in formulating their beliefs, which subsequently has an impact on the practices they will be using in their classes (Borg, 2003). These beliefs include not only their general beliefs (e.g., beliefs about learning the English language, see for instance Smid & Zólyomi, 2021) but their self-related beliefs including self-efficacy beliefs.

The theoretical framework for self-efficacy beliefs originates from Bandura's (1986) Social Cognitive Theory. In his theory, he posits that self-efficacy beliefs entail an individual's perception about their own ability to perform a task or an activity. Historically, language pedagogy and applied linguistics follow the theories in psychology and transfer these theories to language learning. Around the turn of the millennium (Seligman & Csíkszentmihályi, 2000), a paradigm shift happened with focus turning from negative experiences (e.g., foreign language anxiety) to positive experiences (e.g., enjoyment, self-efficacy beliefs). This also meant the change of the social-psychological period to the cognitive-situated period (Mills, 2014). Therefore, self-efficacy beliefs, being cognitively situated, gained prominence after this shift of focus (Mills, 2014).

This study focuses on the self-efficacy beliefs of future teachers specifically, which denote "the beliefs teachers hold about their own perceived ability in undertaking certain teaching tasks" (Pendergast et al., 2011, p. 47). Bandura (1997) claimed that teachers with high self-efficacy perceive all students, including those who may be struggling, "teachable through extra effort and appropriate techniques […] through effective teaching" (p. 240), while teachers with low self-efficacy perceive that they have small impact on learners' cognitive progress—if at all. This, apparently, is in line with Lou and Noels' (2017) foreign language mindset theory originating from Dweck's (2006) domain-general mindset theory. According to Lou and Noels (2017), there are two camps of learners; the *essentialists*, who have a fixed mindset and believe that skills cannot be developed (entity theory) and *constructivists*, who have a growth mindset and believe that skills are malleable (incremental theory). As such, those students who are convinced of the mutability of, for example, their language aptitude (Singleton, 2017) tend to have a growth mindset towards language learning. In contrast, a fixed mindset "leads people to interpret setbacks as a reflection of their underlying incompetence and to show defensive or ineffective self-regulatory strategies in the face of threat" (Dweck & Molden, 2017, p. 137). Along with this theory, those who do not have a fixed mindset towards their skills are more willing to invest effort in language learning, perceive their language aptitude as mutable, and may have higher self-efficacy beliefs. Mindsets also shape teachers' philosophy of implementing differentiated instruction (Tomlinson, 1999). For example, Coubergs et al. (2017) have developed the DI-Quest questionnaire to measure teachers' perceptions about differentiated instruction (DI), and using exploratory and confirmatory factor analysis, they established that mindset, specifically, fixed mindset, is one of the components affecting DI philosophy.

Many studies emphasized that teachers display a sense of low self-efficacy with regard to adapting education to individual students' needs, that is, implementing DI in Hungary (e.g., Öveges & Csizér, 2018; Smid & Zólyomi, 2021; Zólyomi, 2022) as well as internationally (Suprayogi et al.,

2017). With the emerging need to adapt instruction that caters for learners' individual differences with the purpose of creating inclusive classrooms, it becomes increasingly important to employ effective DI techniques. The above-mentioned studies also highlight that teachers, besides finding the implementation of DI challenging, also perceive DI as pivotal in language learning success.

Measuring self-efficacy beliefs is quite challenging owing to the fact that it is a complex construct depending on the context, the situation; moreover, we can also differentiate general and specific self-efficacy beliefs (Faez & Valeo, 2012, Hoang & Wyatt, 2021). Despite the fact that measuring teacher self-efficacy is indeed a "thorny issue" (Tschannen-Moran & Woolfolk Hoy, 2001, p. 794), there were many attempts to design scales measuring teacher self-efficacy, for example, the Teacher Sense of Efficacy Scale (Tschannen-Moran & Woolfolk Hoy, 2001), the Norwegian Teacher Self-efficacy Scale (NTSES; Skaalvik & Skaalvik, 2007), and the Teacher Efficacy for Inclusive Practice Scale (Sharma et al., 2011). Through exploratory factor analysis on the NTSES, Skaalvik and Skaalvik (2007) found six dimensions of teacher self-efficacy, which are instruction, adapting education to individual students' needs, motivating students, keeping discipline, cooperating with colleagues and parents, and coping with changes and challenges. These factors all involve a certain flexibility that is required from the teacher, and previous studies have shown that raising awareness of the importance of adaptability is key for teachers who are at the beginning of their practice (Cansiz & Cansiz, 2019; Faez & Valeo, 2012). In the Hungarian context, Kóródi et al. (2020) employed the NTSES to analyze teachers' self-efficacy beliefs amid the COVID-19 pandemic, and their factor analysis showed slightly different results related to both structure and content. No unique self-developed scale has yet been designed in the Hungarian context to measure teachers' self-efficacy beliefs.

Faez et al. (2019), in their recent meta-analysis argued that self-efficacy beliefs and proficiency show a moderate relationship (r = .37) based on the results of 19 empirical studies, leading to the interpretation

that 13% of the variance is explained by language proficiency. This means that although there seems to be an association between the two constructs, they do not show great overlap, so self-efficacy beliefs are not influenced solely by proficiency. In order to explore what leads to higher self-efficacy beliefs, perceived language aptitude is considered in this study because previous research endeavors have shown that among all of the individual variables, this cognitive construct has the greatest impact on language attainment (Li, 2016). For illustration, Granena and Yilmaz (2019) emphasized that language aptitude can explain 25% of the variance in language learning success.

Bandura (1997) claimed that the perception and interpretation of emotions may weigh considerably more than the actual intensity of emotions. Thinking along the lines of Bandura (1997), it is not necessarily the score on the language aptitude test that foretells the level of success an individual is capable of achieving but the self-perceived level of aptitude and the importance they attribute to this construct. These cognitive processes may weigh markedly more than a single number indicating a score on a test. This suggests that it may be beneficial to develop a growth mindset in pre-service teachers during their teacher training (during *apprenticeship of learning*). Duckworth (2016), in undertaking the investigation of the notion of *grit*, marked the key point of recent language aptitude research claiming that "if we overemphasize talent, we underemphasize everything else" (p. 38); therefore, if pre-service teachers attribute excessive importance to language aptitude, they are probably less willing to consciously put effort in their progress.

In recent studies, language aptitude is investigated along with explicit–implicit learning leading to what we call cognitive aptitudes or explicit and implicit language aptitude (Granena, 2020; Granena & Yilmaz, 2019). Explicit learning denotes a conscious encoding, an analytical and effortful process taking more time, where learners focus on the form and verbalize rules, while implicit learning is a faster process: learners use a holistic approach effortlessly and unconsciously, focus is on the meaning,

and learners use their instincts rather than rules (Ellis, 2005; Granena, 2020).

Those instruments which do not measure both explicit and implicit learning processes presuppose that these two processes are completely orthogonal (i.e., independent from each other). This notion is alluded to as the *polarity fallacy* (Reber, 1993), highlighting that these processes may have a certain overlap. Granena and Yilmaz (2019) also argued that this dual-system approach was favored only before we had advancements in knowledge based on recent studies. Therefore, the concept of *explicit–implicit learning* will be used in this study instead of *explicit/implicit learning* or *explicit vs. implicit learning* to denote that these processes are not completely independent from each other. Thus, "the current understanding of aptitude is that of a multicomponential construct encompassing abilities from both the implicit and explicit cognitive domains" (Granena & Yilmaz, 2019, p. 242).

To date, little is known about the interrelationship between self-efficacy beliefs, perceived language aptitude, and explicit–implicit learning. In light of the proposed aims and the theoretical background, the present study intends to answer the following questions:

1. What are the underlying dimensions of the questionnaire compiled to measure pre-service teachers' self-efficacy beliefs, perceptions of language aptitude, and explicit–implicit learning?

2. What is the relationship between pre-service teachers' self-efficacy beliefs and perceived language aptitude?

3. What are the characteristics of teacher trainees who differ in their explicit–implicit learning behavior?

Research Methods

Participants

Seventy participants filled out the questionnaire. They were recruited via convenience sampling, that is, based on easy accessibility, and purposive sampling, meaning that they had to be teacher trainees studying English, who were still in the phase of the *apprenticeship of learning*. Six of them indicated that they would not be teachers of English; therefore, their data was disregarded so that they would not form a different subpopulation. Two additional participants who reported that they were qualified teachers already and were not studying anymore were also removed. Thus, the final sample included 62 teacher trainees who were enrolled in a variety of majors besides English, the three most popular ones being Hungarian (n = 14), History (n = 13), and German (n = 10), and there were two students in this sample who were in the single-subject teacher training program. There were 47 females and 15 males, which is representative of the gender ratio in the profession, their age ranged from 18 to 43 (M = 22.31, SD = 3.57), and generally, they started learning languages at the age of 8 (SD = 2.15). Based on their self-reported data, with regard to their proficiency, 50% of them have obtained an advanced level language exam certificate in English (C1 based on the CEFR; Council of Europe, 2001), but 4.8% did not obtain any yet. With regard to the teacher training context in Hungary, teacher education until 2021 lasted four + one years for primary school teachers and five + one years for secondary school, but based on the new Governmental Decree, this is unified to five years [Decree no. 538/2021. (IX. 15.)], which is issued from September 2022. Thus, the participants in this study belong to the previous system. There were 13 first-year students, 23 third-year students, 11 fourth-year students, 14 fifth-year students, and one sixth-year student. Fifty-five percent of them had a job besides their university studies; of this, 23% were related to language teaching.

The Instrument

This study employed a cross-sectional questionnaire design encompassing scales related to self-efficacy beliefs, language aptitude, pedagogical mindsets, and the behavioral scales from the explicit–implicit learning habits survey (EXIS; Zólyomi, 2021). Besides the background questionnaire with 10 questions, the questionnaire item pool contained 12 scales with 67 items, which was reduced during factor analysis; consequently, the final instrument contained nine scales with 33 items arranged on a 5-point Likert scale ranging from 1 (*not at all true*) to 5 (*completely true*). The instrument was administered in Hungarian, the mother tongue of the participants, and was later translated by the author (see the Appendix). It took approximately 10 minutes to complete it. The final scales are presented below:

> The items for the future-oriented teacher self-efficacy beliefs scale were adapted from Hoang and Wyatt (2021) and Faez and Valeo (2012), and those for the perceived importance of differentiated instruction scale are based on Zólyomi (2022):
>
> 1. Self-efficacy (four items, $\alpha = .737$): the extent to which teacher trainees perceive that they can successfully develop their students' skills in the future. Sample item: I am confident that I will be able to respond appropriately to students' questions related to the classroom material.
>
> 2. Perceived importance of differentiated instruction (four items, $\alpha = .765$): the extent to which participants find paying attention to students' individual differences important for language learning success. Sample item: I think it is important to pay attention to individual differences in the classroom, such as the different preferred ways of learning.

The language aptitude scales are based on Zólyomi (2020, 2021):

1. Perceived language aptitude (five items, α = .905): the extent to which participants perceive themselves as talented in language learning. Sample item: I have no particular difficulties when it comes to language learning.

2. Perceived importance of language aptitude (four items, α = .827): the extent to which participants find language aptitude important in being successful. Sample item: I think that in order to be successful in language learning, it is important to be able to learn languages quickly compared to our peers.

The effort invested in learning the English language scale is based on Csizér (2020) with slight modifications, and teachers' pedagogical mindsets was adapted from Dweck's (2006) theory as well as from Bandura (1997) and Tomlinson and Imbeau (2010):

1. Effort (four items, α = .757): the extent to which learners are willing to put effort into learning English. Sample item: I am willing to make an effort to learn English.

2. Fixed pedagogical mindset (four items, α = .736): the extent to which participants believe that skills are static and unmalleable. Sample item: I will not consider constructive negative feedback during my teaching career because it does not serve my development.

The scales about language use and language learning behavior were adapted with modifications from the Explicit–implicit learning habits survey (EXIS; Zólyomi, 2021):

1. Implicit language use (four items, α = .840): the behavioral dimension of unconscious English language usage. This scale was intended to measure the extent to which participants report using the language implicitly. Sample item: When I speak English, I use verb tenses instinctively, rather than consciously thought through.

2. Explicit language use (two items, α = .772): the behavioral dimension of conscious English language use. This scale was intended to measure the extent to which participants report using the language explicitly. Sample item: When I speak English, I could subsequently justify why I used that particular verb tense.

3. Explicit vocabulary learning (two items, α = .632): the behavioral dimension of conscious language learning. This scale was intended to measure the extent to which participants report using explicit learning habits. Sample item: When I watch series/movies in English, I write down unknown phrases for learning purposes.

Data Collection and Analysis

After creating the item pool based on Zólyomi (2021), I obtained feedback from an expert and two pieces of peer feedback (from third-year PhD students, who are also English as a foreign language teachers) on the instrument. After refining the instrument, the concurrent think-aloud protocol (Dörnyei, 2007; Willis et al., 1991) was applied with three members of the target population, who shared their opinions regarding the understandability of the items and gave feedback on the instrument from all of the aspects they found relevant. This face-to-face session was recorded for later analysis and lasted 31 minutes of which filling out the questionnaire took between 10–15 minutes. After modifications based on

the think-alouds, which included wording, item randomization, and understandability issues, the questionnaire was ready to be administered for a larger sample. The questionnaire data of the think-alouds were disregarded in later analysis as major modifications were made to the instrument after it. The online data collection lasted for two weeks at the end of 2021. I used Google Forms mainly because of the practical build of the tool: namely, its function of requiring an input to each item before the answers can be submitted, and thus missing responses were not a concern.

The Statistical Package for Social Sciences (SPSS) version 28 was used to analyze the data. After screening and preparing the data for analysis, reliability analyses were run in several steps: exploratory factor analysis (EFA) with varimax orthogonal rotation based on eigenvalues greater than one was computed thematically as a data reduction technique in order to discover the underlying latent dimensions in the dataset. To finalize the scales, the following rules of thumb (Field, 2018; Székelyi & Barna, 2002) were followed: 1) the Kaiser-Meyer-Olkin (*KMO*) measure of sampling adequacy had to be above .50 for factorability; 2) the result of the Bartlett's test of sphericity had to be statistically significant ($p < .05$); 3) the extraction values in the communalities for each item had to be above .25 but below .90; 4) the factor loadings for each item had to be above .50; 5) the rotation sums of squared loadings had to be ≥ 1.0; 6) the X^2 metric of the goodness of fit had to avoid reaching statistical significance ($p > .05$); 7) and finally, in the post-analysis phase, the Cronbach's alpha (α) for each factor had to be above .60. During the factor analysis, items were deleted one by one based on cross-loadings and low or too high extraction communalities. Once the scales were finalized, combined clustering was applied to create groups based on the behavioral scales of the EXIS. First, I computed hierarchical clustering to check the dendrogram, which is a plot showing the arrangement of the clusters, and I fed that solution into the K-Means algorithm, the validity of which was further ensured by post-hoc tests.

Results and Discussion

The Factor Structure of the Questionnaire

As highlighted by Hoekstra et al. (2018) and subsequently in Piniel and Zólyomi's (2022) meta-analysis, relying solely on the Cronbach's α internal consistency measure when developing scales is problematic due to the fact that it is not able to establish unidimensionality. Therefore, exploratory factor analysis was conducted thematically to answer the first research question (presented in Tables 1–4). In all instances, maximum likelihood extraction was used with varimax orthogonal rotation, following Székelyi and Barna's (2002) guidelines. Originally, the self-efficacy scale included 11 items of which seven had to be excluded based on the results of the factor analysis. There was no need to delete any items from the original pool for the perceived importance of differentiated instruction scale; therefore, the final scale contained four items. Based on the final solution, the $KMO = .760$ was acceptable, the result of the Bartlett's test of sphericity was statistically significant ($X^2(28) = 123.63$, $p < .001$), and the goodness of fit ($X^2(13) = 4.82$, $p = .979$) was also acceptable. The final solution resulted in two factors that could measure the two constructs ($\alpha > .70$ in both scales). The rotated factor matrix for these two scales can be seen in Table 1.

Regarding the scales related to language aptitude, one item had to be deleted from the perceived language aptitude scale and two from the perceived importance of language aptitude scale. The *KMO* was acceptable, the Bartlett's test of sphericity was significant ($KMO = .821$; $X^2(36) = 325.25$, $p < .001$), and the final solution indicated a good fit ($X^2(19) = 26.64$, $p = .113$). The final factor structure contains two scales that are able to measure the proposed constructs reliably ($\alpha > .80$ for both scales), see Table 2.

Table 1

Rotated Factor Matrix of Pre-Service Teachers' Perceived Importance of Differentiated Instruction and Future-Oriented Self-Efficacy Beliefs

Items	Factors		h^2
	1 – perceived importance of DI	2 – self-efficacy	
v24. I think it is important to pay attention to individual differences in the classroom, such as the different preferred ways of learning.	**.777**	.108	.616
v27. It is important to tailor education for language learners to be successful.	**.675**	.068	.460
v26. I think it is important to take individual characteristics into account in class while teaching.	**.660**	–.001	.436
v25. When I will be teaching, I will pay special attention to the fact that not all students are the same/identical.	**.579**	.023	.336
v03. I am confident that I will be able to develop students' oral skills.	.127	**.719**	.533
v10. I am confident that I will be able to develop students' pronunciation.	–.004	**.691**	.478
v07. I am confident that I will be able to respond appropriately to students' questions related to the classroom material.	.286	**.658**	.515
v01. I am confident that my general language skills are sufficient to be a language teacher.	–.109	**.587**	.356
Rotation sums of squared loadings	1.940	1.789	–
Average communality	–	–	.466
Cronbach's α	.765	.737	–

Note. Major factor loadings for each item coefficients above .50 are highlighted in bold. Extraction method: maximum likelihood; Rotation method: varimax with Kaiser normalization. Rotation converged in three iterations. DI = Differentiated instruction; h^2 = communalities.

Table 2

Rotated Factor Matrix of Pre-Service Teachers' Perceived Language Aptitude and Perceived Importance of Language Aptitude

Items	Factors		h^2
	1 – perceived LA	2 – perceived importance of LA	
v14. I have no particular difficulties when it comes to language learning.	**.840**	.087	.713
v12. I can learn languages fast compared to my peers.	**.832**	.070	.696
v17. I have good general language abilities.	**.831**	.281	.770
v16. I can claim that I have a good sense for languages since I easily overcome obstacles in language learning.	**.792**	.347	.748
v15. I think I have good general language skills.	**.675**	.194	.493
v21. I think that in order to be successful in language learning, it is important that we have good general language skills.	.123	**.851**	.739
v23. I think that in order to be successful in language learning, it is important to have good general language abilities.	.078	**.833**	.700
v22. I think that in order to be successful in language learning, it is important to have a good sense for languages so that we easily overcome obstacles in language learning.	.219	**.696**	.533
v18. I think that in order to be successful in language learning, it is important to be able to learn languages quickly compared to our peers.	.238	**.547**	.356
Rotation sums of squared loadings	3.296	2.451	–
Average communality	–	–	.639
Cronbach's α	.905	.827	–

Note. Major factor loadings for each item coefficients above .50 are highlighted in bold. Extraction method: maximum likelihood; Rotation method: varimax with Kaiser normalization. Rotation converged in three iterations. LA = language aptitude; h^2 = communalities.

The item pool originally contained items separately for fixed and growth mindsets ($k = 8$ for each); however, during factor analysis, items pertaining to growth mindset did not produce consistent results. In fact, as indicated in Table 3, the two scales had to be reduced to one factor, and the final factor structure for the fixed pedagogical mindset scale contains one item from the growth mindset scale, which loaded highly and negatively on the factor. This item, after establishing the final factor structure, was reverse-coded in further analyses. This decision is similar to Coubergs et al.'s (2017) EFA on the DI-Quest instrument, where they found fixed mindsets emerging from the data. Only one item had to be deleted from the effort invested in learning English scale; thus, the final scale included five items for effort and four items for fixed pedagogical mindset with acceptable factorability indices ($KMO = .732$, $X^2(28) = 132.90$, $p < .001$) and a good fit ($X^2(13) = 10.10$, $p = .686$). The Cronbach's α internal consistency measure was above .70 for both scales, so the items can be considered to be reliable.

The factor analysis of the scales measuring explicit–implicit language learning and language use adapted from the EXIS (Zólyomi, 2021) indicated that there were only three underlying dimensions of these variables (see Table 4). Altogether 10 items were deleted during factor analysis. Interestingly, one item originally from the explicit language use scale loaded highly and negatively on the implicit language use scale; therefore, this item was reverse coded in further analyses, and the number of factors was reduced from four to three. The *KMO* indicated good factorability and Bartlett's test of sphericity was statistically significant ($KMO = .633$, $X^2(28) = 186.87$, $p < .001$). The final factor solution displays a good fit ($X^2(7) = 13.91$, $p = .053$) with the Cronbach's α being above .60 for all three scales. This factor solution is slightly different from that of Zólyomi (2021), but the causes of this difference might be twofold. First, Zólyomi (2021) examined a general adult language learner population, which may lead to a higher variance in the results; second, Zólyomi (2021) used principal components analysis earlier on the EXIS, which may not account for covariance (Tabachnick & Fidell, 2018).

To sum up the underlying factor structure of the complete questionnaire, it seems that the finalized scales can measure the proposed constructs reliably. The English translation of the final instrument can be found in the Appendix along with the items that were deleted during reliability analyses (indicated with italics).

Table 3

Rotated Factor Matrix of Pre-Service Teachers' Effort Invested in Learning English and Fixed Pedagogical Mindset

Items	Factors		h^2
	1 – effort	2 – fixed pedagogical mindset	
v66. Learning English is a very important thing in my life.	**.814**	.003	.663
v67. I am willing to make an effort to learn English.	**.771**	–.180	.626
v63. I do my best to learn English well.	**.576**	–.180	.365
v64. It is important for me to learn English well.	**.519**	–.006	.270
v40. I consider a successful teacher colleague a threat to myself.	–.181	**.692**	.512
v36. I would not be happy for pedagogical challenges in my teaching career.	–.072	**.675**	.461
*v31. I will be open to constructive criticism during my teaching career because they serve my development.	.156	**–.657**	.456
v39. I will not consider constructive negative feedback during my teaching career because it does not serve my development.	.049	**.630**	.399
Rotation sums of squared loadings	1.923	1.828	–
Average communality	–	–	.469
Cronbach's α	.757	.736	–

Note. Major factor loadings for each item coefficients above .50 are highlighted in bold. Extraction method: maximum likelihood; Rotation method: varimax with Kaiser normalization. Rotation converged in three iterations. h^2 = communalities.

* reverse-coded when creating the final scales

Table 4

Rotated Factor Matrix of Pre-Service Teachers' Implicit and Explicit Behaviors

Items	Factors			h^2
	1 – IMP language use	2 – EXP language use	3 – EXP vocabulary learning	
v45. When I speak English, the words come naturally, so I do not have to consciously think about what words to use.	**.894**	.078	–.100	.815
v48. When I speak English, the sentences always come automatically, I do not have to think much.	**.771**	.329	–.156	.728
*v52. If I do not think in advance about exactly what words I want to use when I speak English, it may happen that the words do not come naturally.	**–.691**	–.114	.199	.530
v47. When I speak English, I use verb tenses instinctively, so not consciously thought through.	**.686**	–.236	.048	.529
v51. When I speak English, I know exactly what verb tense I am using.	.118	**.840**	.090	.728
v50. When I speak English, I could subsequently justify why I used that particular verb tense.	.008	**.729**	.151	.554
v59. When I read in English, I usually make a glossary for unfamiliar words.	–.103	.164	**.857**	.772
v61. When I watch series/movies in English, I write down unknown phrases for learning purposes.	–.084	.055	**.521**	.281
Rotation sums of squared loadings	2.374	1.450	1.113	–
Average communality	–	–	–	.617
Cronbach's α	.840	.772	.632	–

Note. Major factor loadings for each item coefficients above .50 are highlighted in bold. Extraction method: maximum likelihood; Rotation method: varimax with Kaiser normalization. Rotation converged in three iterations. IMP = Implicit, EXP = Explicit; h^2 = communalities.

* reverse-coded when creating the final scales

The Relationship Between Self-Efficacy Beliefs and Perceived Language Aptitude

In an attempt to answer the second research question, Pearson product-moment correlation coefficients were calculated. In the present study, the relationship between self-efficacy beliefs and perceived language aptitude was large—$r(60) = .591$, $p < .001$—based on Howitt and Cramer's guidelines (2017). Their causal relationship was checked by regression analysis, which showed that perceived language aptitude explains 35% (R^2) of the variance of self-efficacy beliefs in the teacher trainee population. Additionally, these two constructs showed the largest correlation between all the constructs involved in the present study.

In their recent meta-analysis, Faez et al. (2019) reported an overall moderate relationship ($r = .37$) between self-efficacy beliefs and proficiency, leading to the interpretation that 13% of the variance was explained by language proficiency; thus, 87% of the variance was still left unexplained; therefore, the researchers concluded that self-efficacy subsumes much more than proficiency alone. Granena and Yilmaz (2019) stated that language aptitude explains 25% of the variance in foreign language attainment and proficiency. In light of these results, the 35% accounted for by perceived language aptitude in self-efficacy beliefs is noteworthy.

Explicit–Implicit Behavioral Profiles

Combined cluster analysis was used to answer the third research question. The reason for choosing the three behavioral scales from EXIS (Zólyomi, 2021) as clustering scales was twofold. Firstly, these scales showed the largest standard deviations in the dataset; secondly, the purpose of the cluster analysis was to detect if there were any significant differences in self-efficacy beliefs in teacher trainees who employed different learning techniques. As can be seen in Table 5, almost half of the participants belong to the group that demonstrates implicit as well as explicit language

use. The fact that there seem to be tendencies and not clear-cut mutually exclusive categories points to the dynamic, single-system approach of explicit–implicit learning as suggested by Granena and Yilmaz (2019). Even though explicit and implicit learning appear to be different constructs based on the exploratory factor analysis, concluding from the cluster analysis, there is some extent of overlap in these processes. The second group shows dominant explicit language use and explicit vocabulary learning tendencies, whereas the third group, constituting 16% of the whole sample, displays a strong inclination for implicit language use.

Table 5

The Final Cluster Centers With the EXIS Behavioral Clustering Scales

Clustering Scales	Cluster centers in the three clusters		
	1	2	3
n (%)	30 (48)	22 (35)	10 (16)
Group labels	Implicit-explicit language users	Explicit dominant	Implicit language users
Implicit language use	4.43	2.73	4.18
Explicit language use	4.42	3.75	2.55
Explicit vocabulary learning	3.10	3.61	1.75

Table 6 presents the results of the analysis of variance (ANOVA) controlling for the cluster memberships and indicates that there are significant intergroup differences between seven scales. The members of the three groups tend to agree that DI is important in language learning success, which is in line with the results of previous studies (Öveges & Csizér, 2018; Smid & Zólyomi, 2021; Suprayogi et al., 2017; Zólyomi, 2022). Similarly, all three groups consider language aptitude important, but apparently they do not attribute an overly decisive role to it. This is favorable since if they do not attribute language aptitude to have a key part in language learning success, they may exert more effort in their

language learning (Duckworth, 2016). The participants who displayed implicit as well as explicit language use have the highest perceived language aptitude. They also put more effort into learning English than the other groups, and most importantly, this group shows the highest self-efficacy beliefs. This enables us to conclude that those teacher trainees who employ both learning habits are at an advantage compared to those who do not. The third group (implicit language users) has statistically significantly higher means on the fixed mindset scale, which is quite straightforward since they tend to rely mainly on one learning habit.

Table 6

The One–Way ANOVA of Each Cluster on the Scales

Scales	Groups			*F*	*df*	η^2	Post-hoc comparison
	1	2	3				
IMP language use	4.43	2.73	4.18	52.89***	61	.642	2 < 3 , 1
EXP language use	4.42	3.75	2.55	23.07***		.439	3 < 2 < 1
EXP vocabulary learning	3.10	3.61	1.75	13.01***		.306	3 < 1 , 2
Perceived LA	4.20	3.46	3.32	6.94**		.190	3 , 2 < 1
Effort	4.77	4.44	4.23	7.84***		.210	3 , 2 < 1
Importance of DI	4.60	4.58	4.50	0.18		.006	3 , 2 , 1
Self-efficacy beliefs	4.28	3.72	3.68	6.00**		.169	3 , 2 < 1
Fixed mindset	1.52	1.63	2.15	4.18*		.124	1 , 2 < 3
Perceived importance of LA	3.48	3.59	3.28	0.40		.013	3 , 1 , 2

Note. Post–hoc test used: Duncan. IMP = Implicit; EXP = Explicit; LA = Language aptitude; < = significant difference; , = lack of significant difference.

* $p < .05$, ** $p < .01$, *** $p < .001$.

Conclusion

The first aim of this study was to explore the factor structure of an instrument designed to measure teachers' self-efficacy beliefs, perceived language aptitude, and explicit–implicit learning. The exploratory factor analysis revealed altogether nine factors, and the scales proved to measure the proposed constructs reliably. As there was no instrument developed yet to measure these constructs together in the Hungarian context, this may be useful for fellow researchers who intend to investigate these constructs in upcoming studies.

The second aim of this research endeavor was to analyze the relationship between self-efficacy beliefs and perceived language aptitude, which, besides being the strongest correlation amongst the scales, demonstrated that 35% of the variance in self-efficacy beliefs could be explained by perceived language aptitude; thus, as compared to previous studies, perceived language aptitude seems to have more predictive power than self-perceived proficiency. As this is a novel result, it may be beneficial to investigate self-efficacy with perceived language aptitude in future studies as it may be more important than self-perceived proficiency.

The third aim of the study was to analyze the profiles and characteristics of teacher trainees who differ in their explicit–implicit learning behavior. Most importantly, the results lend support to the single-system approach to explicit–implicit learning (Granena & Yilmaz, 2019), which means that these two processes may not be completely independent. In addition, those who employ both learning habits tend to show higher self-efficacy beliefs; therefore, it is of utmost importance to reconsider the role of explicit–implicit learning in teacher education. Among the pedagogical implications of this study, it is important to highlight the role of perceived language aptitude and explicit–implicit learning processes during pre-service teachers' *apprenticeship of learning* (Pendergast et al., 2011). Therefore, tailoring teaching methods for flexible differentiated instruction addressing both explicit and implicit learning

may enhance teacher trainees' self-efficacy beliefs, which, in turn, may lead to a lower rate of early attrition.

The limitations of this study include the sampling procedure, namely, the size of the sample as well as the sampling method. The results are also limited to the Hungarian context. There are, of course, aspects that either could not be addressed or are beyond the scope of this study. Further studies could address the development of self-efficacy beliefs in the Hungarian context by collecting and analyzing longitudinal data as required by Complex Dynamic Systems Theory. Future studies could also perform confirmatory factor analysis on the proposed scales.

Note: Supported by the ÚNKP-21-3 New National Excellence Program of the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation Fund.

References

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman and Company.

Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching*, *36*(2), 81–109. https://doi.org/10.1017/S0261444803001903

Cansiz, M., & Cansiz, N. (2019). How do sources of self-efficacy predict preservice teachers' beliefs related to constructivist and traditional approaches to teaching and learning? *SAGE Open*, *9*(4), 1–8.

Coubergs, C., Struyven, K., Vanthournout, G., & Engels, N. (2017). Measuring teachers' perceptions about differentiated instruction: The DI–Quest instrument and model. *Studies in Educational Evaluation, 53*, 41–54. https://doi.org/10.1016/j.stueduc.2017.02.004

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Council of Europe. https://rm.coe.int/16802fc1bf

Csizér, K. (2020). *Second language motivation research in a European context: The case of Hungary*. Springer.

Csizér, K., Albert, Á., & Piniel, K. (2021). The interrelationship of language learning autonomy, self-efficacy, motivation and emotions: The investigation of Hungarian secondary school students. In M. Pawlak (Ed.), *Investigating individual learner differences in second language learning* (pp. 1– 21). Springer.

Decree no. 538/2021. (IX. 15.) on the amendment of the Governmental Decree no. 283/2012 (X. 4.) on the system of teacher training, the order of specialization and the list of teaching professions. *Magyar Közlöny*, *170*, 8015–8022.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative and mixed methodologies*. Oxford University Press.

Duckworth, A. (2016). *Grit: The power of passion and perseverance*. Scribner.

Dweck, C. S. (2006). *Mindset: The new psychology of success*. Random House.

Dweck, C. S., & Molden, D. C. (2017). Mindsets: Their impact on competence motivation and acquisition. In A. J. Elliot, C. S. Dweck, & D. S. Yeager (Eds.), *Handbook of competence and motivation: Theory and application* (2nd ed.; pp. 135–154). The Guilford Press.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, *27*(2), 141–172. https://doi.org/10.1017/S0272263105050096

Faez, F., & Valeo, A. (2012). Teacher education: Novice teachers' perceptions of their preparedness and efficacy in the classroom. *TESOL Quarterly*, *46*(3), 450–471. https://doi.org/10.1002/tesq.37

Faez, F., Karas, M., & Uchihara, T. (2019). Connecting language proficiency to teaching ability: A meta-analysis. *Language Teaching Research*, *25*(5), 754–777. https://doi.org/10.1177/1362168819868667

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). Sage.

Granena, G. (2020). *Implicit language aptitude*. Cambridge University Press.

Granena, G., & Yilmaz, Y. (2019). Cognitive aptitudes for explicit and implicit learning. In Z. E. Wen, P. Skehan, A. Biedroń, S. Li, & R. L. Sparks (Eds.), *Language aptitude: Advancing theory, testing, research and practice* (pp. 238–256.) Routledge.

Hoang, T. (2018). Teacher self-efficacy research in English as a foreign language contexts: A systematic review. *The Journal of Asia TEFL*, *15*(4), 976–990. http://dx.doi.org/10.18823/asiatefl.2018.15.4.6.976

Hoang, T., & Wyatt, M. (2021). Exploring the self-efficacy beliefs of Vietnamese pre-service teachers of English as a foreign language. *System*, *96*, Article 102422. https://doi.org/10.1016/j.system.2020.102422

Hoekstra, R., Vugteveen, J., Warrens, M. J., Kruyen, P. M. (2018). An empirical analysis of alleged misunderstandings of coefficient alpha. *International Journal of Social Research Methodology*, *22*(4), 351–364. https://doi.org/10.1080/13645579.2018.1547523

Howitt, D., Cramer, D. (2017). *Understanding statistics in psychology with SPSS* (7th ed.). Pearson.

Kóródi, K., Jagodics, B., & Szabó, É. (2020). Az észlelt tanári énhatékonyságot befolyásoló tényezők vizsgálata a kényszerű digitális oktatás időszakában [Investigating the constructs influencing teachers' self-efficacy beliefs during emergency digital teaching]. *Iskolakultúra*, *30*(10), 38–52. https://doi.org/10.14232/ISKKULT.2020.10.38

Li, S. (2016). The construct validity of language aptitude. *Studies in Second Language Acquisition*, *38*, 801–842. https://doi.org/10.1017/S027226311500042X

Lortie, D. C. (1975). *Schoolteacher: A sociological study*. University of Chicago Press.

Lou, N. M., & Noels, K. A. (2017). Measuring language mindsets and modeling their relations with goal orientations and emotional and behavioral responses in failure situations. *The Modern Language Journal*, *101*(1), 1–30. https://doi.org/10.1111/modl.12380

Mills, N. (2014). Self-efficacy in second language acquisition. In S. Mercer & M. Williams (Eds.), *Multiple perspectives on the self in SLA* (pp. 6–22). Multilingual Matters.

Moodie, I. (2016). The anti-apprenticeship of observation: How negative prior language learning experience influences English language teachers' beliefs and practices. *System*, *60*, 29–41. http://dx.doi.org/10.1016/j.system.2016.05.011

Öveges, E., & Csizér, K. (Eds.). (2018). *Vizsgálat a köznevelésben folyó idegennyelv-oktatás kereteiről és hatékonyságáról: Kutatási jelentés* [Investigation of the context and efficiency of foreign language teaching in public education: Research report]. Oktatási Hivatal.

Pendergast, D., Garvis, S., & Keogh, J. (2011). Pre-service student-teacher self-efficacy beliefs: An insight into the making of teachers. *Australian Journal of Teacher Education*, *36*(12), 46–57. http://dx.doi.org/10.14221/ajte.2011v36n12.6

Piniel, K., & Zólyomi, A. (2022). Gender differences in foreign language classroom anxiety: Results of a meta-analysis. *Studies in Second Language Learning and Teaching*, *12*(2), 173–203. https://doi.org/10.14746/ssllt.2022.12.2.2

Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious.* Oxford University Press.

Seligman, M. E. P., & Csíkszentmihályi, M. (2000). Positive psychology: An introduction. *American Psychologist*, *55*(1), 5–14. https://doi.org/10.1037/0003-066X.55.1.5

Sharma, U., Loreman, T., & Forlin, C. (2011). Measuring teacher efficacy to implement inclusive practices. *Journal of Research in Special Educational Needs*, *12*(1), 12–21. https://doi.org/10.1111/j.1471-3802.2011.01200.x

Singleton, D. (2017). Language aptitude: Desirable trait or acquirable attribute? *Studies in Second Language Learning and Teaching*, 7(1), 89–103.

Skaalvik, E. M., & Skaalvik, S. (2007). Dimensions of teacher self-efficacy and relations with strain factors, perceived collective teacher efficacy, and teacher burnout. *Journal of Educational Psychology*, *99*(3), 611–625.

Smid, D., & Zólyomi, A. (2021). An interview study on Hungarian high school English teachers' beliefs about learning English. In G. Tankó & K. Csizér (Eds.), *DEAL 2021: Current explorations in English applied linguistics* (pp. 139– 162). Eötvös Loránd University.

Suprayogi, M. N., Valcke, M., & Godwin, R. (2017). Teachers and their implementation of differentiated instruction in the classroom. *Teaching and Teacher Education*, *67*, 291–301. https://doi.org/10.1016/j.tate.2017.06.020

Székelyi, M., & Barna, I. (2002). *Túlélőkészlet az SPSS-hez: Többváltozós elemzési technikákról társadalomkutatók számára* [Survival kit for SPSS: On multivariate statistical analyses for researchers of social sciences]. Typotex Kiadó.

Tabachnick, B. G., & Fidell, L. S. (2018). *Using multivariate statistics* (7th ed.). Pearson.

Tomlinson, C. A. (1999). *The differentiated classroom: Responding to the needs of all learners.* Association for Supervision and Curriculum Development.

Tomlinson, C. A., & Imbeau, M. B. (2010). *Learning and managing a differentiated classroom.* Association for Supervision and Curriculum Development.

Tschannen-Moran, M., & Woolfolk Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, *17*(7), 783–805.

Willis, G. B., Royston, P., & Bercini, D. (1991). The use of verbal report methods in the development and testing of survey questionnaires. *Applied Cognitive Psychology*, *5*, 251–267.

Yada, A., Björn, P. M., Savolainen, P., Kyttala, M., Aro, M., & Savoliainen, H. (2021). Pre-service teachers' self-efficacy in implementing inclusive practices and resilience in Finland. *Teaching and Teacher Education*, *105*, Article 103398. https://doi.org/10.1016/j.tate.2021.103398

Zólyomi, A. (2020). Advanced Hungarian language learners' beliefs and perceptions of language aptitude. In C. Kálmán (Ed.), *DEAL 2020: A snapshot of diversity in applied linguistics* (pp. 253–280). http://www.eltereader.hu/media/2020/12/Kálmán-Csaba_WEB.pdf

Zólyomi, A. (2021). Focusing on the explicit and implicit language learning-related beliefs and the habits of Hungarian adult English as a foreign language learners. *Working Papers in Language Pedagogy*, *16*, 82–108.

Zólyomi, A. (2022). Exploring Hungarian secondary school English teachers' beliefs about differentiated instruction. *Language Teaching Research*. Online first. https://doi.org/10.1177/13621688221114780

Appendix
The English Translation of the Complete Questionnaire

(The items deleted during exploratory factor analysis are italicized.)

(Future-oriented) Self-efficacy beliefs

1. I am confident that my general language skills are sufficient to be a language teacher.
2. *I am confident that I will be able to prepare a class to meet the outcome requirements, such as completing the high school final exam or the language exam.*
3. I am confident that I will be able to develop students' oral skills.
4. *I am confident that I will be able to develop students' writing skills.*
5. *I am confident that I will be able to develop students' reading comprehension skills.*
6. *I am confident that I will be able to develop students' listening comprehension skills.*
7. I am confident that I will be able to respond appropriately to students' questions related to the classroom material.
8. *I am confident that I will be able to make proper use of the ICT tools available in the classroom (e.g., smartboard).*
9. *I am confident that I will be able to help students learn grammar.*
10. I am confident that I will be able to develop students' pronunciation.
11. *I am confident that I will be able to put together effective lesson plans.*

Perceived importance of differentiated instruction

1. I think it is important to pay attention to individual differences in the classroom, such as the different preferred ways of learning.
2. When I will be teaching, I will pay special attention to the fact that not all students are the same/identical.
3. I think it is important to take individual characteristics into account in class while teaching.

4. It is important to tailor education for language learners to be successful.

Perceived language aptitude

1. I can learn languages fast compared to my peers.
2. *I can learn languages easily compared to my peers.*
3. I have no particular difficulties when it comes to language learning.
4. I think I have good general language skills.
5. I can claim that I have a good sense for languages since I easily overcome obstacles in language learning.
6. I have good general language abilities.

Perceived importance of language aptitude

1. I think that in order to be successful in language learning, it is important to be able to learn languages quickly compared to our peers.
2. *I think that in order to be successful in language learning, it is important to be able to learn languages easily compared to our peers.*
3. *I think that in order to be successful in language learning, it is important to be able not to have any particular difficulties, for example, not to get stuck while learning a grammatical formula.*
4. I think that in order to be successful in language learning, it is important that we have good general language skills.
5. I think that in order to be successful in language learning, it is important to have a good sense for languages so that we easily overcome obstacles in language learning.
6. I think that in order to be successful in language learning, it is important to have good general language abilities.

Effort invested in learning English

1. I do my best to learn English well.
2. It is important for me to learn English well.
3. *I am determined to learn English well.*
4. Learning English is a very important thing in my life.

5. I am willing to make an effort to learn English.

(Fixed) Pedagogical mindset

1. *It would be good if there were pedagogical challenges at times when I have to teach.*
2. *I will persevere in teaching despite possible failures.*
3. *I can achieve my teaching-related goals if I work hard.*
4. I will be open to constructive criticism during my teaching career because it serves my development. (reverse-coded item)
5. *I consider a successful teacher colleague an example to follow.*
6. *I need to constantly acquire new knowledge in order to achieve my teaching-related goals.*
7. *Anyone can be taught with extra effort and with appropriate teaching methods.*
8. *I will be able to motivate anyone during my teaching career.*
9. I would not be happy for pedagogical challenges in my teaching career.
10. *If obstacles arise in my teaching career, I may opt to give up the profession.*
11. *Making huge efforts to realize my teaching goals is futile.*
12. I will not consider constructive negative feedback during my teaching career because it does not serve my development.
13. I consider a successful teacher colleague a threat to myself.
14. *I think my existing knowledge will be enough to achieve my teaching goals, I do not think I will need new knowledge.*
15. *Extra effort and good teaching methods aside, there are students who cannot be taught.*
16. *If students are fundamentally not motivated, then as a teacher I can no longer do anything to change that.*

Implicit language use:

1. *When I speak English, I pay attention to the meaning of what is being said instead of insisting on the correct use of verb tenses.*

2. When I speak English, the words come naturally, so I do not have to consciously think about what words to use.
3. *When I speak English, I do not think about the grammar rules in my sentences.*
4. When I speak English, I use verb tenses instinctively, rather than consciously thought through.
5. When I speak English, the sentences always come automatically, I do not have to think much.
6. If I do not think in advance about exactly what words I want to use when I speak English, it may happen that the words do not come naturally. (reverse-coded)

Explicit language use:

1. *When I speak English, I always plan ahead for exactly what the sentence will be.*
2. When I speak English, I could subsequently justify why I used that particular verb tense.
3. When I speak English, I know exactly what verb tense I am using.
4. *When I use English, I always try to consciously incorporate newly learned words into my sentences.*

Explicit vocabulary learning:

1. When I read in English, I usually write a glossary of unfamiliar words.
2. *I always write a glossary of new words when I read a book in English.*
3. When I watch series/movies in English, I write down unknown phrases for learning purposes.
4. *I usually speak in English outside class and seminars on a weekly basis only to practice the language.*

8

Flow Experiences During Speaking Tasks in the Hungarian English as a Foreign Language High School Classroom

Katalin Piniel
brozik-piniel.katalin@btk.elte.hu

Fanni Ritecz
fanniritecz@gmail.com

Abstract

Positive psychology in education (Shernoff & Csíkszentmihályi, 2009) investigates what works well in the foreign language classroom in general, and what task types may lead to flow experiences (Piniel & Albert, 2019). To add to the discourse on the role of flow in language learning, the current study investigated task-specific flow related to speaking tasks in the Hungarian English as a Foreign Language high-school context with a conceptual and partial replication of Czimmermann and Piniel's (2016) research on task-specific flow. The research design closely built on the methods used in the initial study. We elicited data from 75 learners using the task-specific flow questionnaire on the same variety of task work modes (individual, pairwork, and groupwork) as in the initial study, but with a sole focus on oral tasks. Our results lend further support to the claim made by Czimmermann and Piniel (2016) that a large proportion of the learners are likely to experience flow in the classroom while completing language learning tasks. We also found empirical evidence underpinning the conceptual difference between the flow and anti-flow constructs, and a distinction between the various flow constituents. Unlike in the initial study, here we found the individual task to be associated with higher levels of flow in general; however, in terms of flow

DOI: 10.21862/ELTE.DEAL.2022.8.197

constituents, participants felt relatively low levels of control in all three tasks and found the pairwork mode to be associated with the highest level of challenge–skills balance.

Keywords: flow experiences, language learning, speaking tasks, replication study

Flow Experiences During Speaking Tasks in the Hungarian English as a Foreign Language High School Classroom

In the past few years, positive psychology (PP) in second language acquisition (SLA) has started to gain ground as a means to view and investigate foreign language learning with a special focus on the key actors: namely learners, teachers, and institutions (MacIntyre & Mercer, 2014). Positive psychology advocates a non-deficit approach to looking at human beings, which in the language learning context means that the focus is on what works well and what elements constitute positive language learning experiences that ultimately lead to successful foreign language attainment. One such positive type of experience that seems to be also relevant in formal educational contexts is the concept of *flow* (Shernoff & Csíkszentmihályi, 2009). This immersive state of concentration has been shown to appear in connection with various school subjects and to promote learning. Furthermore, it has also been suggested that flow may be relevant in language learning (Piniel & Albert, 2019).

In an edited volume, MacIntyre et al. (2016) paved the way for scholars to look at language learning through a PP lens. In one of the chapters, Czimmermann and Piniel (2016) investigated task-specific flow experiences in order to find further evidence for flow in advanced English as a Foreign Language (EFL) university classrooms. However, since then, few researchers have scrutinized flow experiences in foreign language classrooms. In order to find further support for the claims focusing on what task characteristics may be associated more with flow experiences in EFL lessons and to add to the discourse on the role of flow in language learning, we opted to follow Czimmermann and Piniel's (2016) study and conduct a partial replication in the Hungarian high school context. Overall, we can say that "when we carry out a replication study, our aim is to design, report, and compare a piece of research that repeats a previous study in some way" (McManus, in press, p. 5). With this aim, we set out to execute a replication of Czimmermann and Piniel (2016).

The paper follows the structural guidelines outlined by Porte and McManus (2019), and after providing a short background on flow theory, it summarizes the initial study that served as the foundation of the replication. Before moving on to the methods section, we also provide a brief overview of replication studies in general. The methods section describes the details of the current study in light of Czimmermann and Piniel's (2016) work. Finally, our results are presented and compared to those of the initial study.

Background

The notion of investigating flow experiences is strongly linked to psychological research on the creative processes of artists (Csíkszentmihályi, 1997) as they seem to be immersed in certain activities. Csíkszentmihályi et al. (2005) came to characterize these types of autotelic (i.e., rewarding in themselves) and intrinsically motivating experiences with the term *flow*, which they defined as "a subjective state that people report when they are completely involved in something to the point of forgetting time, fatigue, and everything else but the activity itself" (Csíkszentmihályi et al., 2005, p. 600). Since flow is a highly intensive as well as productive state that seems to be conducive to an individual's development and flourishment, the investigation of such momentary experiences is also one of the key issues within positive psychology research (Seligman & Csíkszentmihályi, 2000).

Not only artists but also people in their everyday lives seem to experience flow in various activities in which they engage (Csíkszentmihályi, 1997), even when it comes to learning in formal/institutional contexts such as schools (Shernoff & Csíkszentmihályi, 2009). For some time now, flow in academic contexts has been on the agenda for educational psychology research as well, and more recently, scholars have also begun investigating language learners' flow experiences (e.g., Dewaele et al., 2022; Dewaele & MacIntyre, 2022; Egbert, 2003; Hong et al., 2017; Ibrahim & Al-Hoorie, 2019; Li et al., 2019;

Liu et al., 2021, 2022) in hope of delineating engaging activities in and outside the classroom that can ultimately foster successful language attainment.

However, certain characteristics of flow make it more meaningful to investigate it on a task rather than a general classroom or course level. As described above, *flow* is a momentary experience characterized by intense levels of concentration, feelings of control, and a loss of self-consciousness, as well as a distorted sense of time, where the person finds the activity itself rewarding irrespective of the outcome (Nakamura & Csíkszentmihályi, 2002). Flow is considered to be also a dynamic state that changes from one moment to the next, which would make scrutinizing it on a task level reasonable.

The most important condition that needs to be met for flow to occur is a balance between the perceived skills of the learner and the perceived challenges of the task (Csíkszentmihályi et al., 2005). When this balance is upset, Csíkszentmihályi et al. (2005) talk about anti-flow experiences, which include anxiety, boredom, and apathy. More specifically, when the challenge is perceived to be more difficult than the person's perceived skills, anxiety arises; when the task is too easy, boredom sets in; and when neither the challenges of the task or the skills of the learner are high, the authors talk about apathy.

Researchers investigating flow experiences during tasks (e.g., Egbert, 2003; Franciosi, 2011) found it convenient to adopt a Task-based Language Teaching (TBLT) perspective, where the key unit of instruction and research was the task itself. A task is generally considered an activity that is meaningful and goal-oriented, engages learners' interest, and is related to real life activities where learners use the target language for solving a problem or to share experiences (Willis, 1996). Willis (1996) also highlighted what a task is not: For example, it does not entail language exercises where the target phrases are given or even role-plays that are pre-scripted, that is, a task does not provide a complete script to students. Ellis (2009) further provided four criteria regarding task-based language teaching, which include the notions that (1) focus is on meaning;

(2) in the task a 'gap' should be presented where learners, for example, need to communicate information or express their opinion about an issue; (3) task completion depends on learners' language knowledge; and (4) the outcome of performing a task is more essential than language use. In other words, language is the means not an end in TBLT; nonetheless, learners need to be able to perform the selected task, be it a simple everyday activity or a discussion of more complex issues using the target language (Long, 2014).

Some researchers in their work have focused on classroom flow experiences at the task-level and have come to similar conclusions on the role of skills-challenge balance and the level of control. Egbert (2003), for example, in her seminal article investigated optimal experiences in a Spanish as a foreign language classroom with the help of seven computer-mediated language learning tasks. Her results showed that flow did appear in the language classroom, especially when there was a balance between the challenge posed by the task and learners' skills, when the task was interesting, and when it allowed learners to exercise control. Furthermore, in face-to-face contexts of Japanese EFL learners, Aubrey (2017a, 2017b) used oral tasks related to inter-cultural experiences as a basis for looking into the optimal experiences of learners. The author found that intercultural contact, where control and sense of accomplishment played key roles, was conducive to flow. In another study on learning tasks, Cho (2018) in the Korean context investigated task modality, complexity, and their link to flow. The results of this study suggest that written tasks were more supportive of optimal experiences perhaps due to the fact that participants reported higher levels of the challenge–skill balance and higher level of competence with higher sense of control across different complexity conditions regarding the written tasks.

Initial Study

In order to further our understanding of flow experiences in the Hungarian instructed language learning context, we devised a replication study based on Czimmermann and Piniel's (2016) research, which involved advanced language learners in the Hungarian EFL context and focused on task-specific flow experiences as well as anti-flow (anxiety, apathy, boredom). In the initial study, altogether 85 first-year advanced learners of English studying in an English language and literature BA program at a major university in Hungary took part in performing a narrative domino task in one of three task modes (individual, pairwork, or group work). Data collection took place during EFL lessons in the presence of one of the researchers, and after the tasks students were asked to respond to a classroom flow and a task-specific flow questionnaire. The data analysis involved descriptive statistical analyses as well as cluster analysis, correlation analysis, and ANOVA. These were also used in the replication study in order to be able to meaningfully compare the results of the two studies. Overall, findings in the initial study suggested that flow appeared in the EFL classroom, especially where participants felt in control and their skill level matched the task's challenge irrespective of task modes.

Replication Studies

The reason we opted for a replication study was largely influenced by the fact that applied linguistics journals have recently started to appreciate the importance of conducting and publishing replications as evidenced by calls for papers specifically eliciting such reports (e.g., Studies in Second Language Acquisition, Language Teaching). The argument behind encouraging replications is that such studies are seen as "[a]dding to the validity of published research and help the field move forward in terms of providing a better understanding of already published results" (McManus, in press). As such, scholars view replications as a way of

consolidating evidence for the claims put forward in a discipline (McManus, in press).

There are various types of replication studies based on the extent to which the steps in the initial study are followed. According to Porte and McManus (2019), replication studies can be placed on a continuum from exact or what Marsden et al. (2018) call direct replications, to "modified repetitions" (Porte & McManus, 2019, p. 7) of the initial study. There seems to be no consensus in the nomenclature for the latter; therefore, a variety of labels appear in the literature from partial (American Psychological Association, 2022; Marsden et al., 2018), approximate or conceptual replications (Porte & McManus, 2019) to extension studies (focusing on the limitations of the initial study; McManus, in press). Even concerning the same label such as partial replications, we can find various definitions: While Marsden et al. (2018) talk about one variable being changed in comparison to the initial study, the American Psychological Association's (2022) Dictionary of Psychology defines partial replication as "replication of an empirical study in which only a subset of the study's design and methodology are repeated" (American Psychological Association, 2022).

In view of the above, our paper presents a study that falls in the category of conceptual replications whose aim is to establish "claims about theories and concepts" (McManus, in press, p. 7), in our case concerning flow in the language classroom. As such, the resemblance is confined to particular aspects of the initial study. For example, "[a] conceptual replication might ask the same question as some previous study, but it investigates this question in a different way" (McManus, in press, p. 7) as is the current case, and it also relies heavily on–although it does not copy closely–the methodology of the initial study by testing the flow-inducing nature of three different types of tasks during an EFL lesson. Since our work does not include a focus on general classroom flow as in the initial study, our research also qualifies as a partial replication as defined by the American Psychological Association's (2022) Dictionary of Psychology as cited above. Throughout our replication, we also tried to address some of the limitations of the initial study.

In order to ensure quality in terms of replication research, Marsden et al. (2018) suggested certain criteria to follow. First of all, the initial study should ideally be a significant one in the field. With respect to its focus on flow experiences in the classroom, Czimmermann and Piniel's (2016) work seems to be an important piece due to the fact that, generally speaking, there are only a few studies on flow experiences in language learning, and it appeared in a pioneering volume on positive psychology in language learning, which has since been considered as one of the first major publications on SLA and positive psychology (Dewaele et al., 2019). The next criteria Marsden et al. (2018) listed is that "research needs to be replicated to inform theory, method, or practice" (Marsden et al., 2018, p. 328), which the current study intends to do since it focuses on investigating an under-researched area of flow theory's applicability to the formal language learning context. Third, "researchers themselves [should] provide theoretical and methodological justifications in the rationales sections of their replication studies" (Marsden et al., 2018, p. 329). Some scholars have added additional criteria to those already listed. Namely, Porte and McManus (2019) suggested that researchers conducting replication studies should "seek out more evidence for what is presented as an outcome to any research" (p. 12), which in the present case means that we would like to follow up on Czimmermann and Piniel's (2016) findings on what kind of tasks tend to be more flow inducing in the language classroom. Furthermore, Al-Hoorie et al. (2021) formulated result interpretability, theoretical maturity, and terminological precision as criteria ensuring quality in replication research. Considering the above, transparency and comparability are key; therefore, replication research should follow the design and analysis of the initial study, and as is true for any research, it should be clear in terms of reporting the study (McManus, in press; Marsden et al., 2018).

Taking the above into account, the current study focuses on investigating task-specific flow related to speaking tasks in the EFL classroom by executing a conceptual but at the same time partial replication of Czimmermann and Piniel's (2016) research on task-specific

flow, where the "aim is to examine the same underlying theory as the original" (Porte & McManus, 2019, p. 84): more specifically, flow theory (Csíkszentmihályi, 1997) along with its applicability in the language classroom by focusing on speaking tasks in particular. Based on the above, and drawing on Czimmermann and Piniel (2016), we sought answers to the following research questions:

1. Do Hungarian EFL high school students experience flow during speaking tasks in various work modes?

2. How are task-specific flow and its components related to anti-flow experiences in speaking tasks?

3. Do task variations (individual, pairwork, or group work) contribute to students' task-specific flow experiences?

These questions are close adaptations of the task-specific research questions posed by Czimmermann and Piniel (2016) with a deliberate focus on flow at the task level.

Methods

In terms of its design, the present work fits on the replication study continuum (Marsden et al., 2018). It closely builds on the methods used in the initial study, considering the instrument (the questionnaire used to elicit information about flow experiences during classroom tasks) and data collection procedures, which include a variety of task work modes. To allow for the comparability of the results, thus fulfilling the main purpose of replication in research in terms of gathering more evidence to substantiate the initial study's findings, the data analysis procedures of the initial study were also followed closely. Since this was not an exact replication, we made several changes to the design. These involved the participants and the outcome measures, in hope of providing the

researchers with more information on "the application, relevance, or generalizability of the underlying theory or hypotheses of the original study" (Porte & McManus, 2019, p. 83). In the subsequent sections, details as to the similarities between the original and the replication study's data collection and analysis follow.

Participants

In Czimmermann and Piniel's (2016) study, the sample consisted of first-year BA students majoring in English studies at a Hungarian university, while in the present research, the population in focus were Hungarian secondary school learners: more specifically, those in their 12th year, finishing their studies. What motivated the change in the participants was our aim to investigate the presence of flow in language classrooms where the learners were not necessarily as highly motivated as English language majors and this way alleviate the possibility of gathering highly skewed data.

Compared to the 85 first-year English major BA student participants in the initial study, in this replication the participants were 78 Hungarian EFL learners completing their final year in secondary schools. The gender distribution in the two studies was similar but not identical: Czimmermann and Piniel's (2016) sample consisted of 60 females and 25 males, with an average age of 19.80 (SD = 1.90), while the current study had a gender distribution of 43 females and 35 males, with a mean age of 17.87 (SD = 0.34). In our replication, most of the participants (83%) learnt English as their first foreign language—a small percent of them (17%) chose English as their second foreign language. Their level of English proficiency was around B2, and they started learning English in grade 4 in elementary school. In the initial study, the average time spent learning English was 9.20 years (SD = 3.40), which means that the participants in that study also began learning English around grade 4 and their level of proficiency was at the B2 level, since that was an entry requirement to the English program where they studied.

Instruments

For the data collection in the present study, we used two types of tools: oral tasks as well as a questionnaire measuring learners' task-specific flow experiences (the same survey as Czimmermann and Piniel, 2016, used). In the initial study, the task the learners were asked to perform as individual, pairwork, or group work was a spoken or written narrative based on a picture domino activity (Wright et al., 2006). In our replication study, we focused on language learners' speaking skills and used oral tasks in individual, pairwork, and group work modes. The reason for the change in the task modality was the relative frequency of speaking activities as opposed to writing tasks in the Hungarian secondary foreign language classroom (Árva, 2012).

Tasks

The speaking tasks were designed based on Willis' (1996) criteria for an activity to be labelled as a task (i.e., they were focused on meaning, goal-oriented, related to real-life activities, and engaged learners' interest) as well as following the requisites formulated by Egbert (2003) and reiterated in the initial study that a task must meet in order to foster flow experiences (i.e., clear goals, appropriate challenges, sense of control, interest, feedback, and chance of focus; Egbert, 2003). These criteria were also checked against the tasks by an expert.

As for the particular language production tasks themselves, all the tasks were completed in English. The individual task in this replication study was very similar to the picture domino task found in Czimmermann and Piniel's (2016) study: The participants had to create a story in English based on a set of pictures. The researchers used not only cards but an online version of the game "Story dice" (https://davebirss.com/storydice/-9dice.html) to generate pictures. The second activity was a role-play activity in which participants were asked to act out a situation in pairs in English about travelling abroad, as part of the Erasmus+ Program. The third task was a debate in which participants were asked to address in

groups the possibility of abolishing homework (for a description of these tasks, see the Appendix). The third task was also carried out in English. Two learners from the target population helped in piloting the tasks. The results of the trials demonstrated that the tasks were easy to understand, adequate to the expected language skills, and mostly interesting for the participants; hence, they were suitable for generating flow experiences.

Questionnaire

In executing our replication, we used the same survey instrument to assess the presence of flow during these tasks as in the initial study. This meant that after completing the tasks, participants filled out the task-specific flow questionnaire in their mother tongue (Hungarian) from Czimmermann and Piniel (2016), which was originally based on Egbert's (2003) Perceptions Survey. The instrument tapped into the flow experiences at task-level, as well as anti-flow experiences. In terms of flow, the constructs that were measured included interest, control, attention, and challenge—skills balance. As for the anti-flow experiences, two scales measured boredom and apathy, respectively. In the initial study, a separate instrument measuring learners' state anxiety was not included here, as the main focus of the study was flow rather than anti-flow. To ensure the quality of the data collected, the questionnaire was tried out using the think aloud method with the help of two potential participants for the current study. The reliability of the scales was also assessed using Cronbach's alpha coefficients. With the aim of enhancing the consistency of the scales, the number of items had to be reduced by one for the challenge—skills balance and one for the task-specific control scale, while compared to the initial study, the items measuring attention were kept intact. After these changes, the consistency of the measures was found to be at similarly acceptable levels (Pallant, 2001) as in the initial study (see Table 1).

Data Collection Procedures

After contacting and gaining the consent of secondary school headmasters, English language teachers, the students, and their parents were informed of the research aims and procedures. Parental consent was also sought for the underage participants.

Three groups of students participated, and each one was assigned a different speaking task in terms of the individual, pairwork, or group work mode. The three groups of 26 learners completed the tasks and filled in the task-flow questionnaire during regular class time. Anonymity was ensured throughout the project. The first group completed the individual task of the narrative based on pictures appearing on the dice. The task was explained, students were given a few seconds to think and to digest the symbols, and then they were instructed to start when they thought they were ready. They all told their stories to the researcher individually. The second group did the role-play task. Explanation about the present research and about their task was given, and then they were asked to act out the situation in pairs. The third group of participants were asked to take part in a debate. The class was divided into four groups, two arguing for the debate motion and two against it. All of the tasks were completed in English. After the tasks, the learners filled out the questionnaire in their mother tongue, that is, Hungarian.

Data Analysis Procedures

Questionnaire data was recorded on a spreadsheet using the statistical software IBM SPSS statistics (Version 28). The responses of Likert-type items were recorded using numerical values (1 = strongly disagree, 2 = disagree, 3 = neither agree nor disagree, 4 = agree, 5 = strongly agree). The negative items were reversed. The items were aggregated into scales measuring the various flow and anti-flow constructs enumerated above and were used for further descriptive and inferential statistical

procedures, including cluster analysis, correlations, and ANOVA. These matched the procedures of the initial study for reasons of comparability.

Table 1

Descriptive Statistics for Each Scale and Their Corresponding Reliability Coefficients in Czimmermann and Piniel (2016) and the Present Study

Study	Scales	*M*	*SD*	Cronbach's α	*k*
Czimmermann & Piniel (2016)	Task-specific flow	3.56	0.64	.86	15
	Task-specific challenge–skills balance	3.56	0.72	.65	4
	Task-specific interest	3.18	0.99	.85	4
	Task-specific attention	3.86	0.92	.72	3
	Task-specific control	2.78	0.67	.64	4
	Task-specific boredom	1.78	0.70	.73	4
	Task-specific apathy	1.48	0.65	.71	4
Present study	Task-specific flow	3.26	0.60	.80	14
	Task-specific challenge–skills balance	3.83	0.85	.64	3
	Task-specific interest	2.93	0.87	.76	4
	Task-specific attention	3.60	0.85	.71	4
	Task-specific control	2.70	0.87	.63	3
	Task-specific boredom	2.06	0.83	.77	4
	Task-specific apathy	1.76	0.71	.61	4

Results and Discussion

Based on Porte and McManus (2019), the presentation of the results follows the structure in the original study; that is, the findings are summarized according to the research questions. The first research question targeted the general flow experiences during speaking tasks in various modes in Hungarian EFL high school lessons. Following the guidelines in the initial study, participants whose mean score on the task-flow measure was higher than 3 were considered to have experienced flow

during a task. Based on this, 79% (n = 59) of the participants reported to have been in the zone while completing one of the tasks, compared to the 71% in the initial study.

To investigate the natural distribution of students considering task-specific flow, clusters were formed using the same cluster analysis procedure as in Czimmermann and Piniel (2016). In line with the initial study's results, the largest number of cluster members appeared in the average flow group. That, taken together with the above-average group, suggests that most students found the tasks in both studies to be flow enhancing (see Table 2). The effect sizes as indicated by η^2 = .824 and ω^2 = .69 suggest a large effect (Kirk, 1996) given the sample size and the number of clusters or groups (Barnette & McLean, 2002). However, Norouzian and Plonsky (2018) advised researchers to interpret effect sizes with caution, as the sample size here is not very large and can inflate the effect size; moreover, the authors suggested that it would be more meaningful to compare the effect size to that of previous research, in the present case the initial study, which unfortunately was not available.

Table 2

Distribution of Students in Classroom Task-Specific (TS) Flow Clusters

Study	Flow Cluster	*M*	*SD*	*n*
Czimmermann & Piniel (2016)	Below-average TS	2.35	0.26	18
	Average TS	3.30	0.23	45
	Above-average TS	4.11	0.21	22
Present study	Below-average TS	2.31	0.33	11
	Average TS flow	3.10	0.22	41
	Above-average TS	3.93	0.27	26

Note: The scale means are based on the participants' responses on five-point Likert scales.

We also checked what mode the cluster members with average or above average flow belonged to, an aspect not looked into in the initial study. Here we found that in the above-average flow cohort, 15 (57.7%)

participants were from the individual mode; in the average flow cohort, 16 (39%) participants were from the group work mode; and 16 (39%) participants from the paired mode. This suggests that in our sample, individual tasks were more conducive to higher levels of flow, suggesting that a higher level of immersion in the task could be achieved where learners have a chance to work alone. This is in line with the notion that flow is ultimately an individual subjective experience (Csíkszentmihályi, 1997), but interestingly, as our results suggest, it is not limited to tasks where learners work alone, and perhaps flow also spreads among students completing the tasks in pairs or in groups (Piniel & Albert, 2019).

The mean values of the task-specific scales of flow and anti-flow experiences (see Table 1) also show the extent to which learners felt immersed in the language learning tasks. More specifically, the means of the task-specific scales in Table 1 indicate that in both the initial and the present study, the averages of *boredom* (M_{inital} = 1.78, $SD_{initial}$ = 0.70; $M_{present}$ = 2.06, $SD_{present}$ =0.83) and *apathy* (M_{inital} = 1.48, $SD_{initial}$ = 0.65; $M_{present}$ = 1.76, $SD_{present}$ = 0.71) are relatively lower than that of task-specific flow (M_{inital} = 3.56, $SD_{initial}$ = 0.64; $M_{present}$ = 3.26, $SD_{present}$ = 0.60). In terms of the task-specific flow constituents (see Table 1), while *attention* had the highest mean in the initial study, in the present study, it was the *challenge-skills balance* that had the highest average. These results are in line with what Dewaele and MacIntyre (2021) found in the Spanish as a foreign language classroom. That is, although anti-flow experiences are present, flow seems to be generally more prevalent. Specifically, the authors found that the balance of skills and task challenges seems to be decisive when it comes to experiencing flow during authentic speaking tasks such as debates (Dewaele & MacIntyre, 2021).

The second research question focused on task-specific flow and its components and their relationships to anti-flow experiences in the three speaking tasks. The initial study only focused on the link between the components and anti-flow constructs but not among the flow constituents. However, in the present study, we though it more insightful to investigate the link among the components of task-specific flow as well as their

relationship with anti-flow experiences. We found that among the constituents of flow, there is only a moderate level of significant correlation (see Table 3) between r = .24 and r = .53 (Plonsky & Oswald, 2014), providing evidence for divergent validity substantiating the notion that these constructs contribute to optimal experiences relatively separately. However, all of these, especially *interest* and *attention*, have a strong (Plonsky & Oswald, 2014) link to the overall aggregate of task-specific flow. What is more, they all have significant negative relationships with anti-flow constructs, with boredom depicting a strong inverse connection to attention as indicated by the large effect size of $r_{initial}$ = –.70 and $r_{present}$ = –.74 (Plonsky & Oswald, 2014) in both studies. The strong negative correlation among flow components and anti-flow constructs, especially between attention, interest, and boredom, is unsurprising, as what catches a learner's attention and what they find interesting is most probably going to support proactive engagement with the task and not going to foster boredom.

Next, we looked at the different task modes of the speaking activities more closely to answer the third research question on whether task variations (individual, pairwork, or group work) can be linked to students' task-specific flow experiences. From Table 4, it appears that the individual speaking task generated the highest levels of flow (as well as the subcomponents of *interest*, *control*, and *attention*) and the lowest in terms of *boredom*. This is in line with the notion also stated above that flow tends to be an individual experience, albeit it may be contagious and spread from one individual to the other (Bakker, 2005) in the classroom. Unfortunately, the initial study did not provide the averages for task flow concerning the various work modes, so comparison was not possible at this level of analysis.

In order to check whether the differences among the flow experiences in the three different task modes were significantly different, similarly to the initial study, we used one-way ANOVA. While Czimmermann and Piniel (2016) did not report any significant results, in the present case we found differences in terms of flow and all of its

measured components and the various modes regarding the speaking tasks: flow [$F(2, 75) = 4.19$, $p = .019$, $\eta 2 = .10$]; interest [$F(2, 75) = 3.93$, $p = .024$, $\eta 2 = .095$]; control [$F(2, 75) = 10.43$, $p < .001$, $\eta 2 = .218$]; attention [$F(2, 75) = 7.03$, $p = .002$, $\eta 2 = .158$]; challenge-skill balance [$F(2, 75) = 3.94$, $p = .024$, $\eta 2 = .095$]. The differences in task flow, interest, and challenge-skills balance were characterized by medium effect sizes, while the differences in mode concerning control and attention were associated with large effect sizes using η2 as indicators (Barnett & McLean, 2002).

The Tukey post hoc tests revealed that most significant differences concern the individual mode with tasks performed in other modes. This means that regarding the overall *task flow* experience, there is a significant difference between the individual and the pairwork mode ($p = .017$) as well as between the individual and groupwork mode ($p = .020$). As for the constituents of flow, there was a significant difference found between the individual and the pairwork mode concerning *interest* in the task ($p = .030$).

Table 3

Correlations Between Task-Specific Flow and its Components, and Anti-Flow Experiences

Study		Flow	Challenge—skills b.[1]	Interest	Attention	Control	Boredom	Apathy
Czimmermann & Piniel (2016)	Flow	ND	ND	ND	ND	ND	–.59**	-.53**
	Challenge—skills b.		ND	ND	ND	ND	–.28*	-.45**
	Interest			ND	ND	ND	–.61**	-.49**
	Attention				ND	ND	–.70**	-.54**
	Control					ND	–.23*	ND
	Boredom						ND	ND
	Apathy							ND
Present study	Flow	1.00	.48**	.83**	.78**	.62**	-.55**	-.52**
	Challenge—skills b.		1.00	.24*				-.24*
	Interest			1.00	.53**	.41**	-.45**	-.50**
	Attention				1.00	.32**	-.74**	-.42**
	Control					1.00	-.25*	
	Boredom						1.00	.50**
	Apathy							1.00

Notes: [1]Challenge—skills balance; * $p < .05$; ** $p < .01$; ND = no data

Table 4

The Mean and Standard Deviation Values of the Different Scales According to Task Mode

Task mode	Flow		Challenge—skills balance		Interest		Attention		Control		Boredom		Apathy	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Individual	3.53	0.61	3.55	0.72	3.17	1.02	4.01	0.66	3.27	0.94	1.53	0.50	1.70	0.74
Pair	3.13	0.57	4.12	0.94	2.56	0.66	3.44	0.83	2.44	0.79	2.40	0.85	1.86	0.84
Group	3.13	0.55	3.75	0.79	3.05	0.79	3.30	0.85	2.38	0.57	2.24	0.85	1.71	0.52

Regarding *control*, significant differences were found between the individual and the pairwork mode ($p = .001$) as well as the individual and the groupwork mode ($p < .001$). The relatively low mean for *control* for this sample was quite surprising. It seems that it is not very easy for tasks to allow participants to provide the freedom to make relevant decisions about how the speaking task unfolds. Concerning *attention*, there were significant differences between individual and pairwork ($p = .015$) as well as the individual and the groupwork modes ($p = .002$; the latter two links were already foreshadowed by the larger effect sizes associated with the ANOVA results.) In all the cases listed above, the averages for the individual task were higher. Finally, there was also a significant difference found between the individual and the pairwork task in terms of challenge–skills balance ($p = .020$), but here role-play had the higher value.

The reason for these differences appearing here but not in Czimmermann and Piniel (2016) may be attributed to the fact that here we looked at tasks in one modality (oral tasks) in different working modes rather than contrasting tasks of various modality. On the subject of anti-flow experiences, *boredom* and *apathy* depicted the lowest mean values during the individual task, which is good news and suggests that these tasks generally promoted flow rather than anti-flow. Due to the fact that the initial study did not report effect sizes, however, no statistically meaningful comparison could be made between the results obtained here and the initial study.

Conclusion

The present study aimed to investigate the task-level flow experiences of Hungarian high school learners by executing a replication of Czimmermann and Piniel's (2016) research. More specifically, with the participation of 75 language learners, we found that speaking tasks of various modes were capable of fostering flow, and task-specific flow components were negatively linked to anti-flow experiences in the speaking tasks. In both the present and the initial study, a similar proportion of learners reported being immersed in the task. In addition, here, according to our data, the individual task seemed to foster flow more

than pairwork or groupwork. In terms of anti-flow experiences, participants in both our replication and the initial study reported lower levels of anti-flow, which is a positive outcome. Similarly to Czimmermann and Piniel (2016), we found a strong negative relationship between flow and anti-flow components; what is more, in our study, there were also moderate links between flow constituents, which provided further evidence for divergent validity regarding the componential structure of flow. Finally, probably due to the variations in tasks, the differences in flow experiences among the modes were found to be statistically significant here but not in the initial study. This could be attributed to the fact that Czimmermann and Piniel (2016) worked with written and spoken narratives in different modes, while in our study, we included different topics (and different types of spoken discourse as output) all in the same (spoken) modality. It is interesting that overall, only investigating the skill of speaking in our study, participants reported relatively low levels of control and found the pairwork mode to be associated with the highest level of challenge and skills balance, suggesting that learners may feel more restricted and faced with more challenges when negotiating the outcome of speaking tasks where they also need to accommodate their partners.

Our research is not without limitations. First of all, we found it logical and more worthwhile to broaden our knowledge of flow theory's applicability in the language classroom by focusing on a variety of tasks pertaining to only one language skill: speaking. Consequently, the initial study was modified on several points, which at times made the direct comparison of the results somewhat difficult. Here we only controlled for the mode and the modality of the tasks but not their content (either in terms of topic or in terms of type of discourse). On the other hand, limitations of our replication can also be attributed to the fact that the initial study did not report certain results of data analysis which would have allowed for direct comparison of some of the results.

For future research, stemming from the limitations above, we would echo the call of a current methodological movement in applied linguistics

(Marsden & Plonsky, 2018) which urges transparency and rigor in reporting research. Additionally, in agreement with Marsden et al. (2018), we recommend the more widespread use of replication studies also within the area of language learner psychology in order for the field to be able to draw more substantiated conclusions concerning the phenomena under scrutiny.

Notes:
We have no conflict of interest to disclose.
The first author was supported by the National Research Development and Innovation Office in Hungary (NKFI-6-K-129149).

References

Al-Hoorie, A. H., Hiver, P., Larsen-Freeman, D., & Lowie, W. (2021). From replication to substantiation: A complexity theory perspective. *Language Teaching*, 1–16. https://doi.org/10.1017/S0261444821000409

American Psychological Association. (2022). *Partial replication.* https://dictionary.apa.org/partial-replication

Aubrey, S. (2017a). Inter-cultural contact and flow in a task-based Japanese EFL classroom. *Language Teaching Research*, *21*(6), 717–734. https://doi.org/10.1177/1362168816683560

Aubrey, S. (2017b). Measuring flow in the EFL classroom: Learners' perceptions of inter-and intra-cultural task-based interactions. *TESOL Quarterly*, *51*(3), 661–692. https://doi.org/10.1002/tesq.387

Árva, V. (2012). *Teaching English writing skills at Hungarian secondary schools.* Lap Lambert.

Bakker, A. B. (2005). Flow among music teachers and their students: The crossover of peak experiences. *Journal of Vocational Behavior*, *66*(1), 26–44. https://doi.org/10.1016/j.jvb.2003.11.001

Barnette, J., & McLean, J. (2002, April). *Shedding light on the eta-squared and omega-squared relationships with the standardized effect size* [Conference presentation]. Annual Meeting of the American Educational Research Association, New Orleans, LA, United States.

Cho, M. (2018). Task complexity and modality: Exploring learners' experience from the perspective of flow. *The Modern Language Journal*, *102*(1), 162–180. https://doi.org/10.1111/modl.12460

Csíkszentmihályi, M. (1997). *Flow. Az áramlat. A tökéletes élmény pszichológiája*. [*Flow: The psychology of optimal experience*]. Akadémiai Kiadó.

Csíkszentmihályi, M., Abuhamdeh, S., & Nakamura, J. (2005). Flow. In A. Elliot & C. Dweck (Eds.), *Handbook of competence and motivation* (pp. 598–623). The Guilford Press.

Czimmermann, É., & Piniel, K. (2016). Advanced language learners' experiences of flow in the Hungarian EFL classroom. In P. D. MacIntyre, T. Gregersen, & S. Mercer (Eds.), *Positive psychology in SLA* (pp. 193–214). Multilingual Matters.

Dewaele, J.-M., Albakistani, A., & Kamal Ahmed, I. (2022). Is flow possible in the emergency remote teaching foreign language classroom? *Education Sciences*, *12*(7), Article 444. https://doi.org/10.3390/educsci12070444

Dewaele, J. -M., Chen, X., Padilla, A. M., & Lake, J. (2019). The flowering of positive psychology in foreign language teaching and acquisition research. *Frontiers in Psychology*, *10*, Article 2128. https://doi.org/10.3389/fpsyg.2019.02128

Dewaele, J. -M., & MacIntyre, P. D. (2022). "You can't start a fire without a spark". Enjoyment, anxiety, and the emergence of flow in foreign language classrooms. *Applied Linguistics Review*. https://doi.org/10.1515/applirev-2021-0123

Dewaele, J.-M., & MacIntyre, P. D. (2021). El flujo en el aula de español como lengua extranjera [Flow in the Spanish foreign language classroom]. In M. del Carmen Méndez & J. Andoni Duñabeitia (Eds.), *Factores Cognitivos y Afectivos en la Enseñanza de Español como Lengua Extranjera* [Cognitive and Affective Factors in the Teaching of Spanish as a Foreign Language]. Routledge.

Egbert, J. (2003). A study of flow theory in the foreign language classroom. *The Modern Language Journal*, *87*(4), 499–518. https://doi.org/10.1111/1540-4781.00204

Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics, 19*(3), 221–246. http://dx.doi.org/10.1111/j.1473-4192.2009.00231.x

Franciosi, S. J. (2011). A comparison of computer game and language-learning task design using flow theory. *CALL-EJ*, *12*(1), 11–25.

Hong, J. C., Hwang, M. Y., Tai, K. H., & Lin, P. H. (2017). Intrinsic motivation of Chinese learning in predicting online learning self-efficacy and flow experience relevant to students' learning progress. *Computer Assisted Language Learning*, *30*(6), 552–574. https://doi.org/10.1080/09588221.2017.1329215

Ibrahim, Z., & Al-Hoorie, A. H. (2019). Shared, sustained flow: Triggering motivation with collaborative projects. *ELT Journal*, *73*(1), 51–60. https://doi.org/10.1093/elt/ccy025

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, *56*(5), 746–759. https://doi.org/10.1177/0013164496056005002

Li, R., Meng, Z., Tian, M., Zhang, Z., & Xiao, W. (2019). Modelling Chinese EFL learners' flow experiences in digital game-based vocabulary learning: The roles of learner and contextual factors. *Computer Assisted Language Learning, 34*(4), 483–505. https://doi.org/10.1080/09588221.2019.1619585

Liu, H., & Song, X. (2021). Exploring "flow" in young Chinese EFL learners' online activities. *System*, *96*, Article 102425. https://doi.org/10.1016/j.system.2020.102425

Liu P., Zhang, Y., & Liu, D. (2022). Flow experience in foreign language writing: Its effect on students' writing process and writing performance. *Frontiers in Psychology*, *13*, Article 952044. https://doi.org/10.3389/fpsyg.2022.952044

Long, M. (2014). *Second language acquisition and task-based language teaching*. John Wiley & Sons.

MacIntyre, P. D., & Mercer, S. (2014). Introducing positive psychology to SLA. *Studies in Second Language Learning and Teaching*, *4*(2), 153–172. https://doi.org/10.14746/ssllt.2014.4.2.2

MacIntyre, P. D., Gregersen, T., & Mercer, S. (Eds.) (2016). *Positive psychology in SLA*. Multilingual Matters.

Marsden, E., Morgan-Short, K., Thompson, S., & Abugaber, D. (2018). Replication in second language research: Narrative and systematic reviews and recommendations for the field. *Language Learning*, *68*(2), 321–391. https://doi.org/10.1111/lang.12286

Marsden, E. J., & Plonsky, L. (2018). Data, open science, and methodological reform in second language acquisition research. In A. Gudmestad & A. Edmonds (Eds.), *Critical reflections on data in second language acquisition.* John Benjamins Publishing Company.

McManus, K. (in press). Replication and open science in applied linguistics research. In L. Plonsky (Ed.), *Open science in applied linguistics*. John Benjamins Publishing Company.

Nakamura, J., & Csíkszentmihályi, M. (2002). The concept of flow. In C. R. Snyder & S. J. Lopez (Eds.), *Handbook of positive psychology* (pp. 89–105). Oxford University Press.

Norouzian, R., & Plonsky, L. (2018). Eta- and partial eta-squared in L2 research: A cautionary review and guide to more appropriate usage. *Second Language Research*, *34*(2), 257–271. https://doi.org/10.1177/0267658316684904

Pallant, J. (2001). *SPSS survival manual: A step by step guide to data analysis using SPSS for Windows.* Buckingham Open University Press.

Piniel, K., & Albert, Á. (2019). Motivation and flow. In M. Lamb, K. Csizér, A. Henry, & S. Ryan (Eds.), *The Palgrave handbook of motivation for language learning* (pp. 579–597). Palgrave Macmillan.

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

Porte, G., & McManus, K. (2019). *Doing replication research in applied linguistics*. Routledge.

Seligman, M. E. P., & Csíkszentmihályi, M. (2000). Positive psychology: An introduction. *American Psychologist 55*(2), 5–14. https://doi.org/10.1037/0003-066X.55.1.5

Shernoff, D., & Csíkszentmihályi, M. (2009). Flow in schools. Cultivating engaged learners and optimal learning environments. In R. Gilman, E. Huebner, & M. Furlong (Eds.), *Handbook of positive psychology in schools* (pp. 131–145). Taylor & Francis.

Willis, J. (1996). A flexible framework for task-based learning. In J. R. Willis, J. Willis, & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 52–62). Macmillan Education Australia.

Wright, A., Betteridge, D., & Buckby, M. (2006). *Games for language learning*. Cambridge University Press.

Appendix
Task descriptions

Story dice

Mode:	individual
Instructions:	Your job is to create a story using the given prompts. You can either keep the order of the pictures or you re-arrange the order of the dice. You don't have to take the images literally; they can be used metaphorically.
Prompts:	Story dice found at https://davebirss.com/storydice/9dice.html

Role-play

Mode:	pairwork
Instructions:	Your job is to act out the following situation in pairs: "Student A would like to travel abroad, as part of the Erasmus+ Program. Her/His task is to find as much information about the opportunity and apply to participate in the program. Student B is the teacher who is in charge of the Erasmus+ Program at the school. Her/His task is to provide information and help Student A with the application."
Prompts:	available places to travel to, financial issues, accommodation, official documents, deadline for the application

Debate

Mode:	Groupwork The class of 26 was divided into 4 groups: two were labelled Group A and two Group B.
Instructions:	Your job is to discuss and share your opinion on the following matter: "Homework should be banned." Group A has to collect arguments for, whereas group B has to collect arguments against this statement.

9

English Majors' Flow Experiences in L2 Writing: An Interview Study

Duha Alsayed-Ahmad
alsayedahmad@student.elte.hu

Ágnes Albert
albert.agnes@btk.elte.hu

Abstract

The qualitative research reported in this article investigated whether English major students at a Hungarian university experience flow in any kind of writing that they engage in and what dimensions of flow they encounter when they feel highly absorbed in their writing tasks. Moreover, it highlighted what might prevent participants from experiencing flow in writing and what factors affect their engagement in writing. The participants in this study (N = 7) were interviewed using a semi-structured interview protocol, then their interview transcripts were analyzed thematically. The collected data revealed that English majors participating in this study experienced flow in writing. Based on their own accounts, participants' flow experiences were mostly characterized by having an altered sense of time, concentrating on the task at hand, having positive feelings that made this experience intrinsically motivating for them, feeling in control of the task, losing their self-consciousness, and having clear goals. Furthermore, it has been found that factors like certain task features, external distractions, and time limits might hinder participants from experiencing flow in writing. The findings of the current study have clear implications for language teachers and instructors. They should consider a range of factors when designing and administering language writing tasks such as their level of difficulty, the time assigned to complete these tasks, their degree of relevance and familiarity to the

DOI: 10.21862/ELTE.DEAL.2022.9.227

language learners, and their clarity. Teachers should also pay attention to their learners in terms of what task types and genres they prefer to write and how motivated they are to perform a particular writing task.

Keywords: flow, writing, qualitative study, flow dimensions, flow preventions

English Majors' Flow Experiences in L2 Writing: An Interview Study

Education, especially its institutionalized form, faces great challenges these days. Due to a wide range of technological developments, the world we live in has changed in numerous significant ways, yet educational institutions do not always keep up with the changes. The educational system, and the way knowledge is transmitted within it, has remained unchanged in many ways since the industrial revolution when public schooling was introduced on a large scale (Carl, 2009). No wonder that the need for change in education has been on the agenda for some time, which led to the birth of numerous conceptualizations as to what education in the 21st century should be like (Geisinger, 2016; Trilling & Fadel, 2009).

One of the many problems schools seem to struggle with these days is that they fail to engage students (Fredricks, et al., 2005). In our digital age, there are so many sources of information competing for the limited attentional capacity of individuals that schools face a fierce competition. Nevertheless, as is clear from numerous studies, attention is crucial for any kind of learning, including language learning (R. Schmidt, 2010). Being provided the best possible language input will be completely useless unless it becomes intake as a result of being attended to (R. Schmidt, 2010). Hence, it is no wonder that engagement is becoming a catchphrase these days.

Being in flow is undoubtedly an engaging experience; people feel that they become one with the action, doing it for its own sake and not for some external reward (Csíkszentmihályi, 1975). For this reason, a number of researchers consider flow one form or even the prime example of engagement (Seligman, 2011). Thus, in line with earlier attempts at researching flow in educational contexts (Shernoff & Csíkszentmihályi, 2009), it would be important to establish whether language learners are prone to flow experiences when performing certain activities and tasks and to shed light on the factors that are likely to induce flow in them. Since research in flow in language learning mainly targeted speaking and reading activities, we set out to investigate the potential flow experiences

of language learners during writing tasks. In this paper, we first provide a brief overview of the flow construct, followed by a summary of the empirical research carried out in connection with flow in language learning so far. Then we report on an exploratory interview study in which we attempted to investigate the flow experiences of second language writers.

Literature Review

Flow Theory

Flow, referred to as the optimal experience by Csíkszentmihályi (1975), can be defined as "the holistic sensation that people feel when they act with total involvement" (Csíkszentmihályi, 1975, p. 36). According to J. A. Schmidt (2010), a person in flow will be totally immersed in the task at hand, merged with the action, unconscious of both time and self. As inferred from the previous definition, flow experience has certain characteristics or dimensions, which have often been found to be reported by people who experience it. Based on interview data, Csíkszentmihályi (1975) claimed that flow experience is characterized thus: (1) a skills-challenge balance, (2) having clear goals, (3) having control over one's performance, (4) loss of self-consciousness, (5) unambiguous feedback, (6) focused attention, and (7) action-awareness merging while someone performs an activity. However, the list of flow dimensions was later extended partly due to the increased interest in researching the concept of flow in different activities and areas of life and partly due to the growing popularity of and use of questionnaires, which necessitated a more comprehensive coverage of the construct. For example, Jackson and Csíkszentmihályi (1999, p. 16) expanded the previous seven into nine different dimensions, which are explained in detail below:

1. Challenge-skills balance: When performing an activity, there should be a balance between the person's skills or abilities and the

challenge posed by the activity in order for the person to experience flow. If the challenge level is too high, the individual might experience anxiety, whereas if the challenge level is too low, the person might experience boredom or apathy (Egbert, 2003).

2. Action-awareness merging: According to Jackson and Csíkszentmihályi (1999), when someone is in flow, or in the zone, "there is no awareness of self as separate from the actions one is performing" (p. 18); thus, their actions are performed in an automatic or spontaneous way.
3. Clear goals: Having clear goals in mind about what to achieve and what to do in a task or activity increases people's attention; thus, it enhances their flow experience.
4. Unambiguous feedback: Informing people that they are succeeding in what they are doing facilitates their flow experience.
5. Concentration on the task at hand: People who experience flow often report that they experience a state of intense focus while performing the activity or task.
6. Sense of control: When experiencing flow, a person is usually in control of their performance, and they are able to make decisions regarding their progress in the activity or task they are doing.
7. Loss of self-consciousness: According to Csíkszentmihályi (1975) and Jackson and Csíkszentmihályi (1999), people in flow become freed from their self-concerns and lose their ego.
8. Transformation of time: When in flow, sometimes people feel that time alters by passing either more quickly or slowly.
9. Autotelic experience: An autotelic experience is an experience which the individual considers to be beneficial or intrinsically rewarding and would not mind redoing.

Although the two new items on the list, which are the transformation of time and autotelic experience, were already present in earlier conceptualizations of flow (Csíkszentmihályi, 1975), they were not included among the seven original flow dimensions.

The hypothesized relationships between these nine components or dimensions of flow were clarified by Moneta's (2012) Hybrid Componential Model of Flow State. In this model, Moneta hypothesized that these nine components of flow can be divided into two categories: flow antecedents and flow facets. According to him, flow experience is moderated by flow antecedents like goals, feedback, challenge-skill balance, and the person's concentration, leading to experiencing flow state. Once in flow, the various facets of flow unfold and vary in degree from one person to another; these facets include control, merging of awareness and action, autotelic experience, loss of self-consciousness, and distortion of the notion of time.

Flow in Language Learning

Besides studies investigating the presence of flow in educational contexts (Shernoff et al., 2003), the possibility of flow existing in the language learning process and in foreign language (FL) classrooms was also explored. Egbert (2003) conducted one of the first studies of flow in foreign language learning, and her study is considered fundamental regarding task-specific flow in language learning, which inspired further research into this topic (see, e.g., Aubrey, 2017a, 2017b; Czimmermann & Piniel, 2016; Hong et al., 2017; Liu & Song, 2021). Egbert's study targeted 13 Spanish language learners' flow experiences when performing seven different language learning tasks. Egbert aimed to find out whether flow exists in foreign language classrooms or not, and if it does, what type of foreign language learning tasks induce it. The results from this study showed that flow exists in foreign language classrooms. Moreover, participants reported higher levels of flow when they performed a computer-based speaking task in Spanish and when they wrote about topics that interested them. She also concluded that the flow experience differs from one person to another based on their characteristics and the classroom environment.

Although on the whole there appears to be less research on tasks in the written modality involving literacy skills, there are a number of studies investigating flow in reading. For example, McQuillan and Conde (1996) explored the possibility of experiencing flow while reading in English in the case of both English native and non-native speakers. Their results indicated that most of the readers who participated in their study reported experiencing flow when they read for pleasure, when they were interested in the reading topic, and when they felt that the reading text was rewarding and useful for them. Interestingly, McQuillan and Conde claimed that a number of their non-native English user learners reported experiencing flow only when reading texts written in their mother tongue; however, there were some non-native English language users who indicated experiencing flow when reading texts written in both English and their first language. In a later study, Azizi and Ghonsooly (2015) argued that their participants' level of flow experienced while reading several texts was affected by the genre of the text. In particular, the participants of this study reported experiencing higher flow levels when reading texts from the expository genre rather than the argumentative genre.

In an attempt to measure flow in writing, Abbott (2000) investigated the flow experience of two fifth-grade students when writing non-academic texts in their L1 over a period of four months. Results showed that students reported experiencing flow when they felt autonomous and when they were able to make decisions regarding several aspects of their writing such as genre, style, and the length of the text to be produced. Similarly, Larson et al. (1985) reported that high school students engaged in writing research projects experience engagement and total absorption while writing their papers. Investigating flow in writing on another population, Perry (1999) studied whether professional creative writers experienced flow while writing. The researcher interviewed 62 publishing writers: 33 poets and 29 fiction writers. The results showed that writers indeed experienced flow most of the time while producing their literary works, and they also indicated that they learnt how to enhance their

optimal experience over time with more practice and increased knowledge.

In summary, we can conclude from the aforementioned studies that language learners might experience flow in the process of language learning and that flow exists in foreign language learning and FL classrooms. Moreover, language learners' flow experiences depend on various factors such as individuals' characteristics and language learning task features. However, what can also be inferred from the literature is that there is a lack of studies on the flow experiences of EFL learners in writing, who are neither native speakers of English nor professional writers. Thus, as there is a need to investigate this, the aim of the current paper is to investigate the flow experiences of English language major students at a Hungarian university when they perform writing tasks. The research questions to be answered in the proposed study are the following:

1. How likely are English major students to report experiencing flow in writing?

2. What characterizes English major students' self-reported flow experiences when writing?

3. What factors are perceived by English major students as hindering them from experiencing flow in writing?

Methods

To answer these questions and gain a deeper insight into the phenomenon of flow in writing, an exploratory qualitative study was designed. The qualitative approach was chosen as it provides an opportunity for the in-depth exploration of understudied phenomena, such as flow in writing. We interviewed advanced learners of English about their flow experiences in connection with writing in English with the help of a semi-structured interview. Flow experiences of the PhD students who participated in this

study were thought to be representative of flow in the writing processes of English major language learners in general and were investigated as such. Besides encouraging them to share their flow experiences in their own words, we also used the theoretical conceptualizations of flow discussed above to shed light on different aspects of this multifaceted phenomenon. Thus, the conceptual framework (Jackson & Csíkszentmihályi, 1999) introduced in the literature review was also used when analyzing our data. The following sections are going to provide more details about the methods applied in this study.

Participants

Seven participants took part in this study. Based on the fact that all of them were pursuing their PhD studies in English at the time of the study at a Hungarian university and on the information they provided, their English language proficiency level was advanced. In order to ensure anonymity, the participants were given pseudonyms. There were six female participants and one male participant, and they were selected based on convenience sampling.

The first participant, Worood was a 24-year-old female who had been learning English for up to 13 years. According to Worood, at the time she was using English for educational purposes, socializing, work, and for communication in everyday life. She believed that she was talented in writing, especially in writing in English rather than in her mother tongue, Arabic. She also enjoyed writing research papers and articles.

The second participant, Melis, was a 33-year-old female who had been learning English for 22 years. Melis used English for everything in her life: with friends, with her foreign partner, at work, and for her studies. She reported enjoying writing essays and short stories in English rather than in her mother tongue, Turkish. Although she thought that her writing skills were okay, she saw writing as a burden, especially the process of starting to write.

The third participant, Dana, was a 31-year-old female who had been learning English for around 25 years by the time of the study. According to Dana, writing was her least favorite skill among the four language skills, and she found it complicated. However, she believed that her writing was good, and she found it easier to write in English rather than in her mother tongue, i.e., Arabic, especially when she needed to write academic papers. Writing in English made Dana feel more liberated—as if she was a different person.

Next is Khawla, who was 37 years old and had been learning English for around 30 years. Khawla used English for academic purposes as well as for interacting with her international friends. At the time when the study was conducted, she used to write academic papers and assignments and wrote Facebook posts and Twitter tweets in English for fun. Khawla liked writing in English more than writing in her mother tongue, Arabic, and she thought that it was easier for her to write in English than in Arabic. She also felt a sense of freedom in writing in English because she did not feel that she might be judged if she made a mistake in her writing.

The fifth participant, Lawrence, was a 31-year-old male student who used English for teaching, studying, reading, and writing. He stated that he had been learning English for around 21 years at the time of the study. He enjoyed writing in English and, as he stated, he felt that he was a different person when he wrote in English. Lawrence enjoyed writing and believed that he had good writing skills.

The sixth participant, Huyen, was a 34-year-old female student who had been learning English for around 21 years at the time of the study. She stated that she used English for work and educational purposes, as well as for communication. Huyen believed that she was a good writer, and she liked the feeling she experienced whenever she wrote in English. Moreover, she thought that writing in English was more straightforward than writing in Vietnamese, her mother tongue.

The seventh and last participant, Sofia, was a 29-year-old female student who had been learning English for around 10 years at the time of the study. Similarly to her peers, she used English for work, socializing,

and for educational purposes. Sofia did not seem to be very confident about her writing skills. She also thought that writing in English was simpler than in her mother tongue, Spanish.

Instrument

The instrument designed for data collection was a semi-structured interview. This type of interview offers flexibility for the researcher as the order of the questions on the interview schedule can be changed, and further questions can be added to it during the interview if needed. In order to cater for the multidimensional nature of flow, the nine dimensions of flow described by Jackson and Csíkszentmihályi (1999) were used as a conceptual framework when developing the interview protocol. The main topics covered by the interview schedule included (1) the characteristics, or dimensions, of the flow experience in English language writing; (2) factors that might prevent writers from experiencing flow; and (3) other factors affecting the flow experience in writing. In order to trigger learners' thoughts about the flow experience, the interview started with a quotation from Csíkszentmihályi's (1996) study describing the state of being in flow. Participants were then asked whether they had similar experiences while writing, and they were encouraged to provide a detailed account of what it felt like to be in flow while writing. Follow-up questions about the various aspects of the flow experience were asked only after learners provided their own accounts of flow.

The interview protocol comprised of 22 questions altogether; the majority of the questions referred to the flow experience in addition to a few general questions about the participants and their writing activities (see the Appendix). The interview protocol was piloted by first seeking expert judgement about the questions compiled and then by conducting a pilot interview with an English major student in order to find out whether the questions were clear and understandable. Minor adjustments were made to the interview protocol as a result of piloting.

Data Collection

After finalizing the interview protocol, the seven interviews were scheduled by contacting the participants and gaining their approval to participate in the study. It was made clear to all the participants that their participation was going to be anonymous and voluntary. Interviews were conducted online via either MS Teams or Skype, and the first author of the article acted as the interviewer. The interviews were conducted in English as the interviewer did not share the first language of the interviewees in all of the cases, and the participants' level of English was deemed adequate to enable them to express their thoughts accurately in English. Each interview lasted approximately for 45 minutes, and the entire process of interviewing lasted for three weeks. For ethical reasons, participants were asked to confirm their voluntary participation in the study, and their oral consent that they do not mind being recorded during the interview was recorded. As the final step of data collection, all the interviews were transcribed, and the resulting transcripts were used for the subsequent phases of the data analysis.

Data Analysis

Thematic analysis (Howitt, 2016) was used to analyze the interview data and identify the main themes that emerged from the interviews in connection with writing flow. Since identifying themes related to writing flow was our main aim, we strived to use an inductive approach and attempted to rely on the interview texts as our primary source of data. However, our background knowledge about the different flow frameworks described in the literature review section clearly influenced both the follow-up questions included in the interview guide and our interpretation of the participants' answers. Consequently, a deductive approach was also used when coding the interview segments. The initial coding for establishing units of meaning was done after the second reading of the interview transcripts, which was followed by several

rounds of rereading and recoding. Then, the final codes were categorized into four separate overarching themes that could be used to provide a comprehensive account of these participants' views about writing flow. The four themes included the (1) participants' own accounts of flow, (2) participants' views concerning different flow dimensions, (3) factors preventing flow, and (4) factors affecting flow in writing.

Results and Discussion

This section contains the results of our study as they emerged from the interviews. The discussion of the findings is presented together with the results, and they are arranged according to the emerging themes that the thematic analysis yielded.

Theme 1: Participants' Own Accounts of Flow

When asked about their own flow experiences in writing, all seven respondents indicated that they had had experiences like the one described in the quote (see the Appendix), which means that all of the advanced L2 writers taking part in the study had experienced flow while writing in English at some point in their lives. Melis stated:

> I experienced this state when I had to write my thesis… The part where I had to report the results of my analysis, I wrote this part in three days… I realized that I was in the zone… as it was about my previous job, I knew what I was doing, and I liked what I was doing. I didn't feel anything. I lost all my sensations.

Although drawing general conclusions about the frequency or widespread occurrence of writing flow based on the experiences of only seven participants taking part in a qualitative study is clearly unwarranted, this finding still lends support to the claim that it is possible

for L2 writers to experience flow while writing in their L2 since all of the respondents were familiar with this experience.

When recounting their own experiences, there were a number of features that participants often mentioned in connection with flow in writing. One of the most common flow characteristics reported was an altered sense of time, which sometimes meant that the participants were unaware of the passing of time, or they sensed that time flew by quickly. They made the following remarks in connection with time: "I forgot about time for a couple of minutes" (Sofia); "I lost track of time" (Lawrence); "the time went really fast" (Huyen); or "I remember that I didn't feel time, I spent many hours many hours working on it... but time didn't feel like it was very long... time just flew by" (Dana). The altered sense of time is a characteristic feature of flow also based on the literature as it appears for example among Jackson and Csíkszentmihályi's (1999) flow components.

Another frequently occurring feature in the descriptions was total concentration; flow was described by the participants as a state where nothing can distract them from the task. Worood described this state in the following way:

> So I have to be completely focused, otherwise I know I won't be able to write something good. I definitely get involved especially when I am working on my own, in my room… and not really having a sense of worry, completely disconnected from Facebook and social media, no music… no one around me.

Dana stated:

> while doing the task, I was pretty much occupied… I remember I didn't daydream and my ideas didn't drift away from the topic... I was so concentrated, I didn't think about other things at that time, so focused, I felt passionate about doing it.

Similarly to the altered sense of time, concentration on the task at hand can also be found in Jackson and Csíkszentmihályi's (1999) framework.

Some participants also mentioned positive feelings in connection with flow, which might be surprising since emotions are not usually included in flow frameworks. Nevertheless, Dana for example stated that "… I remember that I really enjoyed it... before starting this writing task, I didn't feel worried or anxious, I felt okay before", while Khawla reported that "I started organizing my ideas feeling no pressure at all. I felt comfortable about it, I was at ease". The positive feeling mentioned by the participants can probably be linked to the autotelic nature of flow. This is likely to be the reason why individuals find flow experiences intrinsically rewarding (Jackson & Csíkszentmihályi, 1999).

Besides the above more frequently mentioned ones, there were a number of flow characteristics which were articulated by one participant only. Lawrence, for example, talked about being in control of the writing activity; he noted that "I felt like I was in control of my performance". He also experienced a loss of self-consciousness, which he expressed in the following way: "I wasn't worried about my performance… I felt that I performed well. I was not looking for validation or evaluation, it's what I think". Having a definite idea about the task ahead was also mentioned in the flow experience description provided by Melis. She stated "I realized that I was in the zone… as it was about my previous job, I knew what I was doing, and I liked what I was doing." This statement can probably be linked to the notion of having a clear goal, which along with the previously mentioned control and loss of self-consciousness also appears among flow characteristics in various flow frameworks (e.g., Jackson & Csíkszentmihályi, 1999).

When comparing our participants' flow descriptions to the flow characteristics found in Jackson and Csíkszentmihályi's (1999) framework, it can be seen that six out of the nine flow components were present in our participants' descriptions. When giving an account of their flow experiences during writing in English, our participants tended to talk about having an altered sense of time, concentrating on the task at hand,

having positive feelings that made this experience intrinsically motivating for them, feeling in control of the task, losing their self-consciousness, and having clear goals.

Theme 2: Participants' Views Concerning Different Flow Dimensions

Participants were asked to provide further information on their flow experiences, especially on the flow dimensions that were not mentioned by them when they were asked to describe their own flow experiences in writing. When asked about the role of challenge-skill balance in writing tasks, participants stated that their abilities and the challenge of the writing task should match in order to experience flow. They made the following remarks in connection with challenge-skills balance: "If the writing task was up to my level, I will feel more absorbed and focused" (Dana) and "the challenge and my skills should match if I need to be engaged" (Melis). The challenge-skills balance is a basic feature of flow as it already appears in early conceptualizations of flow (Csíkszentmihályi, 1975) and among Jackson and Csíkszentmihályi's (1999) flow components. However, Worood had a different view regarding the challenge-skills balance. She stated the following, "Even if the writing task was difficult, and the challenge was high, I might still be focused and get absorbed if there is/was? a sort of benefit out of performing the task". According to Worood, even if the level of task challenge was high, but the task was rewarding, she might experience flow, thereby shifting the focus from her skills to the perceived importance of the task. Similar findings appeared in Engeser and Rheinberg's (2008) study, in which they only found partial support for the substantial effect of the challenge-skills balance in determining flow, so they concluded that the effect of the interaction between skills and challenge also depends on other factors such as "the (perceived) importance of the activity and the individual achievement motive" (p. 168).

Another flow dimension which was not mentioned in the participants' own account of flow in writing was action-awareness

merging. When asked about this dimension, only one participant, Huyen, commented that having clear task goals helped her perform the writing task automatically without exerting much effort. Huyen said: "If the task aim is clear, its framework appears automatically in my mind." This probably indicates automatic or spontaneous task execution, which is in line with Jackson and Csíkszentmihályi's (1999) conceptualization of action-awareness merging.

The only flow dimension in reference to Jackson and Csíkszentmihályi's (1999) flow dimensions which was not mentioned either in the participants' own accounts or in response to the follow-up questions regarding their flow experiences in writing was the unambiguous feedback dimension. This could be attributed to the lack of immediate external feedback during the process of writing in general. However, the issue of task grading was mentioned when participants were asked about what might hinder or enhance their flow experience in writing, so this will be discussed in the following subsection of the results discussion section.

Based on the abovementioned results, it can be concluded that participants in this study reported experiencing flow in writing. Moreover, in line with Jackson and Csíkszentmihályi's (1999) flow framework, their flow experiences were characterized by (1) having an altered sense of time, (2) staying focused, (3) having positive feelings that make this experience autotelic, (4) feeling in control of the task, (5) losing their self-consciousness, (6) the merging of action and awareness, (7) having clear goals, and (8) the feeling of balance between their skills and the challenge posed by the writing task.

Theme 3: Factors Preventing Flow

Participants were also asked about what hindered them from experiencing flow in writing. One of the most often mentioned factors that hindered their flow experience was time limits and strict deadlines. If the writing task was due by a tight deadline or there was a time limitation for

performing it, they claimed that this negatively impacted their performance as well as their flow experience. Dana said:

> If I have a limited time to complete the writing task, I will not be focused on my ideas and how well the task is performed… I will be occupied with time. I will not be absorbed because I will be checking the time to see how much do I have left, and then my ideas will be cut and I will be distracted… Time limitation is stressful.

According to Dana and the other participants, time limit and tight deadlines caused stress and distraction and affected the overall completion of the writing task. This finding was also evidenced by one of the participants in Larson et al.'s (1985) study on flow in writing research projects, where one of the students expressed that as the deadline approached, he began losing focus and started feeling anxious.

Another factor that has been mentioned by the participants was related to the degree of their familiarity and interest in the topic of the writing task. Both Sofia and Khawla mentioned that their unfamiliarity with the writing task topic would not help them in getting engaged in the writing process. Lawrence, for example, thought that being either overly familiar or entirely unfamiliar with the writing topic might hinder the writer from experiencing flow in writing. He made the following remark: "If the topic is very familiar and you are too knowledgeable about it, you will not know from where to start, and if you are absolutely unfamiliar with it, it is really difficult to get engaged." Worood also thought that being too familiar with the topic of the writing task might prevent her from experiencing flow; however, she claimed that she might still experience flow if the topic was unfamiliar to her. According to her:

> Many things could prevent me from experiencing flow in writing, for example, … if the topic is too exposed or over-used… If I am not familiar with a topic, but I do find it interesting, I will try and find

> my way around it... I will try to collect as much information as I can about it.

Moreover, participants stated that if the topic of the writing task was not interesting to them, they might not be engaged in the writing process. This finding aligns with Egbert's (2003) claim that foreign language students might experience flow if they could deal with interesting topics while performing language tasks.

There were some other factors mentioned by some of the participants that were claimed to hinder them from experiencing flow in writing. One such example is external distractions while performing the writing task, such as noise coming from the surrounding environment, social media notifications, or lack of motivation to perform the task. In Worood's words: "… when there are people around me, and if I lack motivation." Knowing that the task will be graded also seems to affect learners' optimal experience negatively. "If the task will be graded, I will be worried, and this will impact my engagement negatively" (Khawla). In connection with task grading, Sofia thought that knowing that the task would be graded might improve her performance, but it might not help her in getting engaged as she would be worried about grading. She said:

> If I know that the task will not be graded, I will write whatever comes to my mind. If it is graded, I will invest more in it… The fact that the task will not be graded decreases my performance; at the same time, it increases my engagement.

In contrast, Lawrence claimed that knowing that a writing task would be graded would not affect his performance and level of engagement in the writing task. Nevertheless, both Khawla and Lawrence believed that unclear task goals would make them anxious and prevent them from experiencing flow in writing. Lawrence said:

> If you do not know what the aim of the task is, you are aimless and the activity falls apart. If I was not sure what I am expected to do, I would be all over the place, and this will negatively affect my optimal learning experience.

Furthermore, according to Khawla and Huyen, if the task was hard or highly demanding, it would decrease their level of engagement. Huyen said: "If the requirements of the task were beyond my knowledge, I would feel anxious." This comment is in line with flow theory (Csíkszentmihályi, 1975), which claims that task demands exceeding the learner's skills lead to anxiety and also lends support to what Zollars (2018) observed about the effect that tasks lacking a challenge-skills balance have on learners' flow experience. Finally, Sofia and Huyen pointed out that a writing task with many instructions or requirements would make it hard for them to get involved in the process of writing; thus, they may not experience flow. This finding further supports the idea of avoiding restrictive guidelines that has been suggested by Lo and Hyland (2007) for increasing learner engagement.

To sum up, according to the participants in this study, having time limits or strict deadlines for performing a writing task, being unfamiliar with the writing task topic or not finding it interesting, and having distractions would hinder their flow experience in writing. In addition, lacking motivation to complete a task and knowing that the task is going to be graded might prevent flow from happening. Finally, tasks with unclear goals, tasks with a lot of instructions and requirements, and tasks which are considered highly demanding decrease the level of flow the participants experienced in writing. It should be noted that some of these hindrances, like the lack of interest and motivation, excessive challenge, and lack of clear goals, are in direct contrast with the previously identified flow dimensions. Thus, learners' claims that these have a role in preventing flow only accentuates the importance of these factors when present in creating flow.

Theme 4: Factors Affecting Flow in Writing

As this study investigated the flow experiences of English majors in writing, it was important to explore in more detail the factors that affected our sample's flow experiences when performing a writing task. As expected, participants indicated a number of factors which they thought might have affected their flow experience in writing. The first and most important feature of the writing task that would affect language learners' flow was the task topic. All of the participants in this study stated that their familiarity with the writing task topic increased their engagement. Dana said: "My familiarity with the topic affects my writing experience and makes me more absorbed in the task at hand." Secondly, the genre of the writing task seemed to play an important role in writers' optimal learning experience. For example, all of the participants, except for Melis, preferred to write argumentative writing tasks. Alongside with the argumentative genre, both Lawrence and Sofia reported feeling more engaged in creative, academic, and reflective (personal) writing, whereas both Dana and Huyen felt more engaged in academic writing, reflective writing, and narrative writing tasks. Melis and Worood preferred narrative tasks and technical writing. Finally, Khawla felt more engaged in creative writing as she often writes poems. Khawla emphasized the importance of genre in writing by saying: "Genre is important in writing. If I do not like the genre, I will feel like I am compelled to do the task and I would not enjoy it." Lawrence also added: "Task genre plays an important role in my level of engagement". However, according to Huyen, if she did not like a particular genre but had to perform a high-stake task in that particular genre, she might still focus on it and get immersed. This finding is in line with what Azizi and Ghonsooly (2015) concluded regarding the effect of genre on flow in reading.

Interestingly, all the participants in this study stressed that they do not experience flow in writing in their English language classrooms, even if the task was achievable and interesting. This finding contradicts Egbert's (2003) claim that flow occurs in foreign language classrooms

when performing clear and interesting writing tasks. Moreover, Worood, Dana, and Melis stressed that they experienced flow in writing individually but not in collective work, that is, in pairs or groups.

Time limitation seemed to play a major role in language learners' writing and flow experiences. As mentioned previously in the section discussing causes that prevent flow, limited time and tight deadlines may result in anxiety and less engagement. However, it might also help students in experiencing flow in writing. Lawrence pointed out that time limitation can have both a good and bad impact on his writing experience. He said: "Having no time limits gave me a sense of freedom… It could also add a little challenge and enjoyment, but if the time limit was short, it could be a source of frustration." On the topic of task requirements and features, Huyen stated: "… If I had to write something very long, my engagement would be reduced." Thus, the required length of the writing task can facilitate or impede flow in language learners' writing experience. In connection to this finding, remarks about the length of writing have been made by Abbot (2006), who concluded that students experienced flow in writing when they had greater control of the important aspects of their writing, including the length of the paragraph they had to write.

Finally, writing tasks with clear instructions seemed to play a major role in language learners' flow experience. According to the participants in this study, the clearer the task instructions were, the better their performance was, and the more they were engaged in the writing process. In addition, both Khawla and Lawrence expressed their enjoyment of free writing tasks as this technique helped them to brainstorm their ideas.

In summary, several features of writing tasks seemed to affect our participants' flow experience. For example, based on participants' self-reports, the topic and genre of the writing task, its length, its instructions, and time limitation might affect flow experiences. According to the participants of this study, the classroom environment does not enhance getting in the zone while writing; instead, they tend to experience flow in writing individually when they are not in their English language classrooms.

Conclusion

The present study aimed to investigate the flow experiences of English language major students at a Hungarian university when they perform writing tasks. In order to achieve this aim, a qualitative study was conducted by interviewing seven students who were pursuing their PhD studies in English at a Hungarian university. The first two research questions investigated whether English major students experience flow in writing or not, and what characteristics of flow shape their flow experience. According to the findings, our participants were found to experience flow often in writing. Furthermore, participants commented on the following dimensions of flow in writing: challenge-skills balance, clear goals, concentration on the task at hand, sense of control, loss of self-consciousness, transformation of time, action-awareness merging, and interest in the task performed as it becomes autotelic. Although these do not represent all possible flow dimensions, they are generally in line with the findings of earlier research on flow (Csíkszentmihályi, 1975; Jackson & Csíkszentmihályi, 1999).

Regarding the third research question, which investigated what hindered English language majors from experiencing flow in writing, it was found that external distractions, such as noises from the surroundings and mobile phone notifications, prevented them from being focused and in the zone while writing. In addition, the learners' lack of motivation, time constraints of the writing tasks, and tight deadlines affected their optimal learning experience as well as their performance. Special attention was paid to the topic of the writing task. For most of the participants, if the topic was unfamiliar to them or not interesting, they would feel anxious, and this might hinder them from experiencing flow in writing. Moreover, graded writing tasks and writing tasks which had unclear goals and many restrictive guidelines were likely to prevent English majors from experiencing flow. Writing tasks which were considered to be difficult or highly demanding were also thought to hinder the flow experience of our participants. Some of these hindrances are in direct contrast with the flow dimensions proposed by Jackson and Csíkszentmihályi (1999).

There were also other factors mentioned by our participants in connection with what might have affected their flow in writing. For example, along with time limitation and the topic of the writing task, our participants reported that the clarity of task instructions, the required length of the writing task, and the genre of the writing task might affect both their performance and their flow experiences. Furthermore, according to the participants of this study, they were more likely to experience flow when they were alone and not in the EFL classrooms, as according to their self-reports, being surrounded by students and teachers hindered their flow experience.

In conclusion, this study lent support to the possibility of experiencing flow in writing, providing evidence that under the right conditions, writing tasks have the potential to engage learners, creating optimal experiences for language learning. We also pointed out what challenges English majors might encounter while trying to get engaged in the process of writing and what might influence their flow experiences, which in return might affect their performance in writing tasks; this possibility should be the topic of further investigations. Based on the results of the current research, we suggest that in order for teachers to induce flow and enhance English major language learners' writing experiences, they should try to avoid giving language learners in-class writing tasks involving tight deadlines. Moreover, language teachers are advised to pay attention to the writing tasks themselves: their level of difficulty, their degree of relevance and familiarity to the language learners, and their clarity. Teachers should also pay attention to their learners: what task types and genres they prefer and how motivated they are to perform a particular writing task. Finally, we need to draw attention to the limitations of the present study. As this is a small-scale qualitative study, the findings are illustrative of this particular sample only, so further research, possibly using a mixed methods design, would be necessary to investigate this phenomenon with a larger sample and from other possible angles.

Note: The authors are members of the MTA-ELTE Foreign Language Teaching Research Group. This study was funded by the Scientific Foundations of Education Research Program of the Hungarian Academy of Sciences.

References

Abbott, J. A. (2000). "Blinking out" and "Having the touch": Two fifth-grade boys talk about flow experiences in writing. *Written Communication*, *17*(1), 53–92. https://doi.org/10.1177/0741088300017001003

Aubrey, S. (2017a). Inter-cultural contact and flow in a task-based Japanese EFL classroom. *Language Teaching Research,* *21*(6), 717–734. https://doi.org/10.1177/1362168816683563

Aubrey, S. (2017b). Measuring flow in the EFL classroom: Learners' perceptions of inter-and intra-cultural task-based interactions. *TESOL Quarterly, 51*(3), 661–692. https://doi.org/10.1002/tesq.387

Azizi, Z., & Ghonsooly, B. (2015). Exploring flow theory in TOEFL texts: Expository and argumentative genre. *Journal of Language Teaching & Research*, *6*(1), 210–215. http://dx.doi.org/10.17507/jltr.0601.26

Carl, J. (2009). Industrialization and public education: Social cohesion and social stratification. In R. Cowen & A. M. Kazamias (Eds.), *International handbook of comparative education* (pp. 503–518). Springer.

Csíkszentmihalyi, M. (1996). *Creativity: Flow and the psychology of discovery and invention.* HarperCollins.

Csíkszentmihalyi, M. (1975). *Beyond boredom and anxiety*. Jossey-Bass.

Czimmermann, É., & Piniel, K. (2016). Advanced language learners' experiences of flow in the Hungarian EFL classroom. In P. D. MacIntyre, T. Gregersen, & S. Mercer (Eds.), *Positive psychology in SLA* (pp. 193–214). Multilingual Matters.

Egbert, J. (2003). A study of flow theory in the foreign language classroom. *The Modern Language Journal*, *87*(4), 499–518. https://doi.org/10.3138/cmlr.60.5.549

Engeser, S., & Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion*, *32*(3), 158–172. https://doi.org/10.1007/s11031-008-9102-4

Fredricks, J. A., Blumenfeld, P., Friedel, J., & Paris, A. (2005). School engagement. In K. A. Moore & L. H. Lippman (Eds.), *What do children need to flourish? Conceptualizing and measuring indicators of positive development* (pp. 305–321). Springer Science +Business Media. https://doi.org/10.1007/b100487

Geisinger, K. F. (2016). 21st century skills: What are they and how do we assess them? *Applied Measurement in Education,* *29*(4), 245–249. https://doi.org/10.1080/08957347.2016.1209207

Hong, J.-C., Hwang, M.-Y., Tai, K.-H., & Lin, P.-H. (2017). Intrinsic motivation of Chinese learning in predicting online learning self-efficacy and flow experience relevant to students' learning progress. *Computer Assisted Language Learning, 30*(6), 552–574. https://doi.org/10.1080/09588221.2017.1329215

Howitt, D. (2016). *Introduction to qualitative research methods in psychology* (3rd ed.). Pearson.

Jackson, S. A., & Csíkszentmihályi, M. (1999). *Flow in sports*. Human Kinetics.

Larson, R., Hecker, B., & Norem, J. (1985). Students' experience with research projects: Pains, enjoyment and success. *The High School Journal*, *69*(1), 61–69. https://www.jstor.org/stable/40365188

Liu, H., & Song, X. (2021). Exploring "flow" in young Chinese EFL learners' online English learning activities. *System, 96*, Article 102425. https://doi.org/10.1016/j.system.2020.102425

Lo, J., & Hyland, F. (2007). Enhancing students' engagement and motivation in writing: The case of primary students in Hong Kong. *Journal of Second Language Writing*, *16*(4), 219–237. https://doi.org/10.1016/j.jslw.2007.06.002

McQuillan, J., & Conde, G. (1996). The conditions of flow in reading: Two studies of optimal experience. *Reading Psychology*, *17*(2), 109–135. https://doi.org/10.1080/0270271960170201

Moneta, G. B. (2012). On the measurement and conceptualization of flow. In S. Engeser (Ed.), *Advances in flow research* (pp. 23–50). Springer. https://doi.org/hch5

Perry, S. K. (1999). *Writing in flow: Keys to enhanced creativity*. Writers Digest Books.

Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan, & I. Walker (Eds.), *Proceedings of CLaSIC 2010, Singapore, December 2-4* (pp. 721–737). Singapore: National University of Singapore, Centre for Language Studies.

Schmidt, J. A. (2010). Flow in education. In E. Baker, P. P. Peterson, & B. McGaw (Eds.), *International encyclopedia of education* (pp. 605–611). Elsevier.

Seligman, M. E. P. (2011). *Flourish: A visionary new understanding of happiness and well-being*. Free Press.

Shernoff, D., & Csíkszentmihályi, M. (2009). Flow in schools: Cultivating engaged learners and optimal learning environments. In R. Gilman, E. Huebner, & M. Furlong (Eds.), *Handbook of positive psychology in schools* (pp. 131–145). Taylor & Francis.

Shernoff, D. J., Csíkszentmihályi, M., Shneider, B., & Shernoff, E. S. (2003). Student engagement in high school classrooms from the perspective of flow theory. *School Psychology Quarterly*, *18*(2), 158–176. https://doi.org/10.1521/scpq.18.2.158.21860

Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times*. John Wiley & Sons.

Zollars, J. (2018). *Flow theory and engagement: Observing engagement through the lens of flow in a middle school integrated maker space* [Unpublished doctoral dissertation]. University of Pittsburgh.

Appendix
Flow in writing Interview Protocol

This interview is aimed at finding out more about learners' experiences in connection with writing in English. I am going to ask you some questions, and I want you to answer them the best you can. Please be honest and share all of your thoughts openly. These are open ended questions, and there are no correct or incorrect answers. Your responses will help me answer my research questions. You may stop doing the interview at any time, and you do not need to answer any questions you feel uncomfortable with.

I want to assure you that this interview is going to be confidential and your identity is going to be anonymous, so feel free in answering our questions. This interview will last approximately for 40 minutes. I really appreciate your voluntary participation in our research and would like to thank you for your time. For data analysis purposes and better data quality, the interview will be recorded. Can you please confirm that you don't mind that the interview is recorded?

Interview questions:

a. Introductory questions:

1. How old are you?
2. How long have you been learning English?
3. What is your level of proficiency in English?
4. How do you use English right now? For which purposes?
5. What sort of writing tasks you need to perform in English? What sorts of things do you write in English?
6. Do you like writing? What type of texts do you enjoy writing?
7. What are some things you enjoy about writing in English?
8. What are some things you dislike about writing in English?
9. How do you feel about your writing skill?
10. How is writing in English different from your mother tongue?

b. Flow experience questions:
Now I want you to recall your writing experiences in English language.

I am going to read this quote for a poet describing the state of total or intense focus which he called flow experience (Csíkszentmihályi, 1996, p. 121), and then I'll ask you some questions regarding your writing experiences.

> You're right in the work, you lose your sense of time, you're completely enraptured, you're completely caught up in what you are doing . . . when you are working on something and you are working well, you have the feeling that there's no other way of saying what you're saying.

Based on this statement:

11. Did you have a similar experience, to some extent, while you were performing any writing task before?

Probe. 1: Do you experience this state of total engagement at your language classroom as well?

12. Can you describe this experience in details, please?
    Probes: If the description is not detailed enough, ask:
    a. When did it happen?
    b. Did it happen in the classroom or outside it?
    c. What sort of writing were you involved in at that time?
    d. How does it feel to be in the zone while performing a writing task?
    e. Did you feel like you are in total control of your performance?
    f. Did you devote your entire attention to the writing process?
    g. Were you worried about your performance? Why not?
    h. How well do you think you performed the task?
13. When do you usually experience this state of intense focus? Under what conditions?

Probes:

a. What helps you in getting absorbed in the writing task at hand?
b. Do you feel more absorbed when you perform an easy writing task or a difficult one?
c. Do you think that the level of task challenge and your writing skills should match in order for you to be totally engaged in the writing task? Why?
d. In your opinion, what makes a writing task more interesting?
e. Was the task interesting? Did you enjoy performing it?
f. Where you familiar with the task topic? Did this affect your degree of engagement?
g. When you have spent a long time writing a report or an essay, have you ever felt that time passed in an unusual way (e.g., in slow motion, very fast, time stopped)?

14. Was the aim of the task clear enough for you? How did this affect your writing experience?
15. Do you think that the teacher plays a major role in provoking your immersion in the process of writing in the classroom? If yes, can you explain how?
16. Does the topic of the task affect your amount of engagement? In what way does it affect your writing experience?

    Probes:

    a. Which topics interest you?
    b. Does your familiarity with the topic affect your writing experience? How?
17. Do you think that clear task requirements and instructions would improve your writing performance? If yes, can you explain how?
18. Writing has many different genres, such as the narrative genre (e.g., writing stories), descriptive genre (writing an essay describing a place or a person), argumentative genre (e.g., persuasive essays), academic writing (e.g., writing research papers & articles), and creative writing (e.g., poems). Can you think of any particular genres that usually make you feel that you are in the zone while writing?

Probes:

a. Do you get absorbed while writing a research article or an essay for your university courses?
b. Do you get absorbed while writing a narrative of any event?
c. Do you get absorbed while writing a description of anything, e.g., a city, a person, your house, etc.?
d. Do you get absorbed while writing an argumentative essay in which you express your opinion about some controversial issue?
e. Do you write poems or fiction? I f yes, do you experience the state of intense focus while doing so?

19. Do you think that having no time limits for completing a writing task would affect your performance and engagement in the task?
20. If you knew that your performance in a particular writing task would be graded or judged by someone (e.g., your teacher), do you think your level of engagement on the task might still be the same? If no, would you explain why?
21. Can you think of things that might prevent you from getting totally involved in any writing task?

I have no further questions. Do you have anything more you want to bring up, or ask about before we finish the interview? Thank you very much for participating in the research. If you are interested in our research results, we can send them to your email address.

10

The Charismatic Teacher: An Interview Study on the Motivating Agency of Charismatic Language Teachers in Hungarian Higher Education

Szabolcs Peták
pesza96@gmail.com

Csaba Kálmán
kalman.csaba@btk.elte.hu

Abstract

Language teachers' explicit motivational strategies and their motivational power have been thoroughly studied by a number of researchers in L2 motivation research; however, the personal qualities of language teachers have yet to be scrutinized as important sources in the generation of L2 motivation. One such personality trait is charisma. This study aims to investigate how university foreign language (FL) teachers and English language teacher trainees regard charisma as an implicit motivational tool. The research focuses on the qualities that make a FL teacher charismatic and on the motivational power of charismatic behavior. Semi-structured interviews were conducted with eight teachers and eight teacher trainees from a Hungarian university in order to explore the motivational impact of teacher charisma. Results show that subject knowledge, methodological knowledge, and positive character traits were found to be the most distinctive features of charismatic L2 teachers, who were also generally seen as outstanding motivators.

Keywords: charisma, L2 motivation, teacher's role in motivation

DOI: 10.21862/ELTE.DEAL.2022.10.259

The Charismatic Teacher: An Interview Study on the Motivating Agency of Charismatic Language Teachers in Hungarian Higher Education

The generation and maintenance of motivation is one of the most important tasks of L2 teachers, as acquiring an L2 is a rather lengthy, laborious, and possibly frustrating process (Dörnyei et al., 2006). Many techniques aimed at kindling learners' motivation have been developed, as seen in Dörnyei's (2001) taxonomy of motivational strategies, which lists several of them; however, teachers' personal qualities have yet to be thoroughly examined as possible vital factors in motivating learners. One such personal quality which may have a significant effect on learners' motivation is teacher charisma.

Charisma can be defined in various ways. People seen as charismatic have "a personal charm and magic that in others can wake up unusual devotion, personal loyalty, and enthusiasm and promote stronger self-leadership among the followers" (Blašková et al., 2018, p. 3). This explains why charismatic leaders are seen as exceptionally influential people. Although charisma has not been extensively studied from the perspective of teaching, researchers claim that the concept can be applicable to the classroom (Archer, 1994).

There are a limited number of studies on teaching charisma; however, most of them have one common feature: charismatic teaching is generally affiliated with effective teaching. Charismatic teachers are usually seen as knowledgeable and tend to create confiding relationships with their students while also possessing zealous and energetic demeanors (Lee et al., 2014; Suryani, 2016). Although Archer (1994) made it clear that teaching charisma on its own does not equate with effective teaching, the potential positive effects of charismatic behavior inside the classroom could be of immense benefit to teachers of all subjects.

In order to contribute to this research niche and gain valuable insight into charismatic teachers' motivational agency on L2 learners, we conducted a qualitative interview study with eight English language

teacher trainees and eight teachers who teach English as a foreign language at a Hungarian university in Budapest. The teachers in the study were nominated by the student participants, who identified them as their most charismatic language teachers. With the help of the interviews, we intended to explore what qualities make L2 teachers charismatic and how charisma contributes to the motivational power of L2 teachers in a tertiary environment. In line with the literature, our hypothesis was that charismatic teachers would be seen as outstanding professionals and extremely effective motivators.

Theoretical Background

The Teachers' Role in Motivation

Research has shown that there is a significant correlation between the perceived role of teachers in L2 motivation and learners' motivation. Noels et al. (1999) found strong connections between learners' motivational orientations and their perceptions of their teachers' communicative style, showing that teachers who were perceived to provide ample feedback on the learning progress and support learners' autonomy increased intrinsic motivation. In turn, higher levels of intrinsic motivation are related to a decrease in anxiety, as well as improved competence in self-evaluation and an increased level of motivational intensity. Mezei and Csizér (2005) reinforced the importance of the effect that teachers can have on students' motivated language learning behaviour. The critical motivational role of teachers was also confirmed by Chan (2014) and Mezei (2014). Furthermore, Dörnyei and Guilloteaux (2008) found that in South Korea, where opportunities to use in-class motivational strategies are rather restrained because of the country's strict classroom traditions, even limited motivational practice can increase students' motivation in a compelling manner.

Explicit and Implicit Motivation

Explicit motivational strategies (MotSs) are "techniques deployed by teachers to deliberately enhance learner motivation" (Lamb, 2017, p. 2). Inspired by *The Modern Language Journal* debate on motivation research in the 1990s and with the help of pioneering Hungarian EFL teachers, Dörnyei (2001) created a 35-item list of MotSs. Later, 102 micro-strategies were assigned to the 35 main strategies (or macro-strategies) with explanations as to when they should be utilized in the teaching process. Dörnyei's taxonomy has been extremely influential in the field of L2 motivation (Lamb, 2019).

The way teachers utilize MotSs depends on their own beliefs about motivation and how they perceive the context in which the strategies can be put into action (Glas, 2016). Teachers' own professional identities and level of motivation, as well as a rational assessment of the value that different MotSs have for different groups, can also influence teachers' openness to MotS and willingness to test them with their students (Kubanyiova, 2012). However, Thorner and Kikuchi (2019) pointed out that there are several aspects of teacher behavior that affect learners without them being conscious about it. Furthermore, Lamb (2019) cautioned that over-using explicit MotSs may be detrimental to enhancing L2 motivation.

In a survey exploring students' perceptions of their teachers' motivating activities, Lamb and Wedell (2015) found that approximately 50% of the comments written by the students were in connection with the teacher's personal qualities, such as language competence, kindness, and cultural knowledge, instead of any particular MotS. Csíkszentmihályi (1997) stated that teachers' enthusiasm can considerably motivate learners. More recently, the role of teachers' enthusiasm in generating learner motivation has also been highlighted by Kunter et al. (2011). Ghanizadeh and Moafian (2010) showed that interpersonal relationships, enthusiasm, the happiness of the teacher, empathy, and support have the highest correlations with successful learning. Furthermore, Kálmán (2021)

claimed that *personality* and *behavior* were the most important aspects which determined a teacher's influence on motivation in Hungarian corporate contexts, although other aspects that are not related to MotS (e.g., appearance) also proved to be influential.

Based on such studies, it appears that numerous aspects of effective and motivating teaching are not related directly to motivational techniques, or perhaps are not related to them at all. Behavior and personality appear to play a significant role in the generation of motivation (Noels et al., 1999). In addition, there is another quality of motivating teachers which may unite several of the aforementioned aspects and is connected with behavior and personality. This quality is called charisma.

Charisma

Charisma is not easy to define. However, despite the abstract nature of the concept, several diverse definitions of the concept are available. Weber (1968) claimed that charisma is a gift "of the body and spirit not accessible to everybody" (p. 19). The first scholar who proposed an integrated psychological theory for charisma was House (1977), who assigned charisma to "leaders who by force of their personal abilities are capable of having profound and extraordinary effects on followers" (p. 189). However, it can also be regarded as a hypothetical and imaginative character trait that exists only in the imagination and feelings of others (Blašková et al., 2018). Blašková also defined charisma as "a personal charm and magic that in others can wake up unusual devotion, personal loyalty, and enthusiasm and promote stronger self-leadership among the followers" (p. 3), and Bolkan and Goodbogy (2014) described it as a quality of leaders based on their interactions and behaviors with subordinates.

Charismatic Leadership

Charisma has been extensively studied in corporate organizational literature. House (1977) identified the following defining characteristics in

his theory of charismatic leadership: high self-assurance, a strong faith in one's own ideals and principles, and a strong desire to affect and lead people. As a result of these characteristics, followers place considerable trust in their leaders in addition to accepting the leader's value system and setting higher performance goals (House, 1977). In his exploratory study of charismatic leadership in business, Conger (1985) concluded that charismatic leadership was recognized as a distinguishable quality and that a distinction can be made between charismatic and non-charismatic leaders based on explicitly separable qualities such as vision, inspiration, and the ability to excite. Conger and Kanungo (1987) also claimed that—instead of a personal characteristic—charisma is rather the perception of followers based on a leader's qualities, needs, values, beliefs, and their own perception of their followers.

Modern scholars do not consider charisma an unusual and uncommon quality; for example, neocharismatic theorists state that it can be studied in a variety of organizational contexts (Antonakis et al., 2011). They have also suggested that charisma in general is closely related to leaders' delivery mode as opposed to the content itself. Thus, it is leaders' communication in a vivid and emotional manner that makes followers act collectively towards a common vision, not the goal or the vision itself (Antonakis et al., 2011). Therefore, it seems clear that specific behaviors that leaders exhibit when they communicate with their subordinates and followers are particularly important in the study of charisma. This is especially true in instructional contexts (Bolkan & Goodboy, 2014): The numerous activities in which a teacher is engaged can entail a leadership role (Archer, 1994).

Teacher Charisma

Ștefănescu et al. (2020) argued that teachers can be seen as leaders inside the classroom, as they coordinate groups and activities. Moreover, Archer (1993) claimed that charismatic leadership theory is applicable to the classroom. Effective teaching is becoming an increasingly important topic, since knowledge is expanding at an accelerating rate. This can be observed

especially in tertiary contexts, meaning that university students have to cope with increased academic pressure (Blašková et al., 2018). Moreover, teachers' behaviors and attitudes are considered to be the most important factors in students' perceptions of the quality of higher education (Huang & Lin, 2014). In order to develop a better understanding of students' needs, it is important for teachers to learn and understand their students' personalities and motivation alongside while also activating their own personalities and motivation. Charismatic behavior may play a vital part in this process (Blašková et al., 2018).

One may contemplate what charismatic behavior is, as well as how it is manifested by teachers and how it can be assessed. Some teachers enjoy more popularity than others. As students are attracted by qualities that are not possessed by all teachers (Huang & Lin, 2014). Popular teachers are perceived as being able to guide, inspire, and encourage their students by broadening their perspectives and providing them with an enjoyable learning environment, thus decreasing the distance between students and themselves (Oin, 2008). However, teaching charisma is difficult to quantify; therefore, a consensus has yet to be reached amongst scholars regarding its qualities (Lee et al., 2014).

A limited number of empirical studies can be found on the subject. Huang and Lin (2014) created an inventory for assessing teacher's charisma in Taiwanese tertiary contexts, and their quantitative study examined charisma based on four basic determinants: *knowledge, character traits, teaching techniques,* and *humor*. Students perceived charismatic teachers to be highly knowledgeable, proficient in regard to teaching methodologies, having positive personality traits (e.g., approachability, friendliness, and patience) and to be humorous. Their findings were significant, as the qualities of charismatic teachers seem to be related to the attributes of good teachers, who are considered to be knowledgeable in both pedagogy and subject matter (Koehler & Mishra, 2009). Furthermore, good teachers are expected to be friendly, approachable, patient, and enthusiastic, character traits that make them role models for students. Good teachers must also have a large repertoire of teaching skills, meaning

that teaching methodology is also relevant and important (Voss & Gruber, 2006). Furthermore, a sense of humor, which has been proven to have a positive effect on learning, is also an important feature of a good teacher (Huang & Lin, 2014). In addition to being regarded as successful teachers based on the previously mentioned four perspectives, charismatic teachers were also shown to affect students' satisfaction and engagement with learning in a positive way.

Bolkan and Goodboy (2014) also researched teachers' and instructors' charisma. In their quantitative study, they measured 237 students' perceptions of their instructors. It was concluded that teaching charisma is communicated through *nonverbal immediacy, humour, caring,* and *confirmation. Nonverbal immediacy* refers to behaviors which reduce distance between students and the teacher in a psychological and physical sense (e.g., by using one's voice vividly when talking to the class). *Humor* involves the use of personal stories, anecdotes and jokes in an authentic style. *Caring* was regarded as teachers' interest in their students' progress and well-being (e.g., collecting feedback on a regular basis). *Confirmation* refers to teachers' beliefs in students' abilities inside the classroom (e.g., encouraging students to strive for goals not easily reachable). It was also found that similarly to charismatic leaders, charismatic teachers' performances are also perceived to be highly effective and satisfactory for their followers (i.e., their students).

Lee et al. (2014) also investigated charisma based on the same four aspects (i.e., *knowledge, character traits, teaching techniques,* and *humor*) in another Taiwanese quantitative study. Based on the responses of more than 800 junior secondary school students, the researchers found that *knowledge charisma* was the most important quality of charismatic teachers. This was followed by *teaching method* and *character charisma*, respectively. The three aspects and their associated charismatic qualities demonstrated a positive effect on situational and personal motivation. However, despite the fact that *humor charisma* was found to be helpful for enhancing situational interest, its effect on learning interest was not found to be as significant compared to the other three aspects.

L2 Teachers' Charisma

Suryani (2016) argued that students' behaviors are significantly affected by the behavior of the language teachers, as teachers in the second language classroom function as role models for their students and have a high potential to raise their students' motivation through their own communication. Students can adopt their teachers' styles, beliefs, and values, which can help them become better speakers of the target language. Suryani also listed several characteristics of charismatic language teachers. As she claimed:

> Charismatic teachers tend to communicate warmly, love challenge, inspire vision, trust, put high energy, be enthusiastic, be emphatic, inspire self-confidence, caring, encouraging and exciting. They also develop caring, motivating and trusting relationship. By having those characters and relationships, charismatic teachers have a potential role and capacity to nurture students' inner motivation. They can inspire their students' motivation by performing charismatic behaviors and develop warm communication. (p. 16)

Consequently, charismatic language teachers seem to be able to considerably influence their students and increase their motivation through their own personal attractiveness and the way they create and maintain relationships between students and themselves. This is in line with Lee et al. (2014), who claimed that charisma enhances situational and personal motivation. Furthermore, Bolkan and Goodboy (2014) highlighted the salutary effect of charismatic teaching on intrinsic motivation. Since intrinsic motivation is consistently affiliated with an improvement in the quality of learning (e.g., improved goal setting for studying, increased attention, and more thorough revision), the importance of the effect that teachers can have on it is particularly important (Ryan & Deci, 2000). Moreover, trusting relationships between students and teachers enhance students' self-worth, which also affects students' attitudes, engagement, and motivation in a fundamentally

positive way (Ryan et al., 1994). Suryani (2016) claimed that confidential relationships in the L2 classroom engender a supportive atmosphere in which learners can proclaim their opinions and express their perspectives; thus, they become independent learners with a significant amount of self-regulation and autonomy.

As highlighted in the studies above, if language teachers and teachers of all other subjects are knowledgeable, proficient in teaching methodologies, and able to create trusting relationships between learners and themselves, they can have a notable influence on learners' motivation. Since there is a need for motivating and motivated teachers, and charisma appears to be a vital factor for generating motivation, one might ponder whether it can be learnt and taught (Antonakis et al., 2011).

Research Method

Based on the theories drawn from the literature and in accordance with the purpose of our investigation, we intended to explore what makes a teacher charismatic and what role charisma plays in the generation and maintenance of students' L2 motivation in tertiary contexts. To address these aims we formulated the following two research questions:

1. What makes a language teacher charismatic in a tertiary context?

2. How does charisma contribute to the motivational power of L2 teachers in a tertiary environment?

Research Paradigm

The intricate nature of the topic of our investigation and its exploratory nature called for a qualitative research paradigm in order to obtain as rich, accurate, and detailed information as possible. An interview study appeared as a favorable option to answer our research questions, as

interviewing allows researchers to capture both verbal and non-verbal cues, emotions, and behaviors (Mackey & Gass, 2005). The open-ended nature of the questions in our interview guide further enhanced the possibility to explore the sub-topics in a more subtle and sophisticated way.

Participants

Altogether, we interviewed 16 participants who were either attending or teaching at a Hungarian university in Budapest. Eight of the interviewees were English teacher trainees who were selected through convenience sampling. Seven of them were graduate students from the Unified Teacher Trainee Program, and one of them was an MA student in the Instruction of English as a Foreign Language MA program (see Table 1).

Table 1

Student Participants of the Study

Pseudonym	Gender	Age	Other major	Years of studying English	Number of L2 classes/week	Number of English teachers	Still motivated to develop language skills
Bonnie	female	24	-	19	4	3	yes
Dorothy	female	23	Spanish	15	6	2	yes
Haley	female	25	German	16	6	4	yes
Marcus	male	24	IT	15	6	5	no
Mary	female	24	Hungarian	15	7	5	yes
Miles	male	26	history	20	6	3	yes
Reya	female	24	history	14	7	4	yes
Tessa	female	24	Italian	18	6	4	yes

The other eight teacher participants were identified by the students as their most charismatic EFL teachers at the same university. We contacted

them via email after they had received a nomination from one of the student participants. Table 2 shows their biographical details.

Table 2

Teacher Participants of the Study

Pseudonym	Gender	Age	Years of teaching English	Number of L2 classes/week at the university	Number of English classes elsewhere	Still motivated to develop language and teaching skills
Charlie	male	48	29	2	36	yes
Felix	male	56	24	4	0	yes
Julia	female	25	38	0	0	no
Lucas	male	70	40	4	6	yes
Magnolia	female	61	40	2	0	yes
Margaret	female	64	40	8	0	no
Ulbricht	male	60	36	12	0	yes
Zachary	male	69	41	8	0	yes

Research Instrument

We created two semi-structured interview guides based on the literature related to teaching charisma (one for the learners and one for the teachers) in order to find answers to the research questions. The two instruments were similar in the sense that we aimed to find answers to the same or rather similar questions, although the queries sometimes had to be altered considering the participants represented two sub-samples. We opted for the semi-structured format as it provides the researcher with freedom and control over the development of the interview (Wallace, 1998). Besides an introductory section, the interview guide contained two main parts. Each part was based on one research question and had several sub-questions in order to elicit as detailed data as possible.

Procedures

In order to make sure that the developed instrument would yield rich and valid information, both versions of the interview guide were piloted. After the pilot interviews, we decided to omit one question from both versions of the interview guide as it was potentially misleading. This question was the same in both the teachers' and students' version ("In what way do charismatic language teachers contribute to the charisma of teacher trainees?"). The average interview time was approximately 45 minutes. The language of the interviews was Hungarian except in the case of the two teachers whose mother tongue was not Hungarian, with whom the language of the interviews was English. The interviews were conducted online between March and April in 2021.

Data Analysis

Following the consent from the participants, each interview was recorded and transcribed by the authors. The transcripts of the students' interviews contained 40,110 words, whereas the teachers' interviews yielded 36,782 words. For data analysis, the Constant Comparative Method (CCM) was used (Boeije, 2002). The interviews were analyzed in succession, and emerging themes were marked as the transcripts were being analyzed. When particular themes reoccurred in subsequent interviews, the number of occurrences was noted down (e.g., knowledge charisma: 5 out of 8 learners). Eventually, conclusions were drawn by comparing and contrasting the data retrieved from the two groups.

Results and Discussion

Student Interviews

The Qualities of Teaching Charisma

In order to answer the first research question (i.e., What makes a language teacher charismatic in a tertiary context?), we asked the students to

describe teachers whom they saw as able or unable to seize students' attention at university L2 classes, as this ability is associated with charisma in the literature (Bolkan & Goodboy, 2014). Several characteristics were associated with such teachers. One of the characteristics mentioned was knowledge in both the target language and in teaching methodologies, which was highlighted by five students during the interviews. This reinforces the findings of Huang and Lin (2014), according to which both competent and charismatic teachers are knowledgeable in their subjects and in regard to teaching methodology. This is also in line with the research of Lee et al. (2014), who stated that the two most important aspects of charismatic teaching are *knowledge charisma* and *teaching method.*

In connection with teacher-learner relationships, three students out of the eight mentioned that they felt regarded as equals by teachers able to command attention, and two students highlighted that charismatic teachers showed particular interest in their students. Moreover, charismatic teachers were also perceived to be constructive, straightforward, and honest by two different students. The findings of Ștefănescu et al. (2020) and Bolkan and Goodboy (2014) are reinforced by the findings, as charisma is likely to develop caring teacher-learner relationships (Ștefănescu et al., 2020) and is communicated through *nonverbal immediacy, humor, caring,* and *confirmation* (Bolkan & Goodboy, 2014). Furthermore, Ștefănescu et al. (2020) claimed that the presence of humor can also indicate that teachers pay attention to and care about their students. The following statement of Marcus (a student) about Charlie (a teacher) demonstrated how *nonverbal immediacy* and *caring* (Bolkan & Goodboy, 2014) are manifested by charismatic teachers:

> I remember when Charlie told us that it took him more than ten years to become confident as a teacher in the classroom. I liked it that he didn't try to hide this and he didn't want to look like a phoney omniscient teacher, but instead, he regarded us (university students) as human beings. When somebody shows their vulnerable side, it gives evidence of their confidence, because it makes me think that

> there's something behind this bloke... he has achieved something and it makes him credible in my eyes.

The quote from Marcus also confirmed Archer's (1994) idea that paying attention and showing concern about students are common features of charismatic teachers in the investigated context. Other positive character traits listed by the participants were as follows: confidence, creativity, empathy, flexibility, kindness, honesty, patience, and the ability to generate interest. Lee at al.'s (2014) findings are further reinforced by these results, as *character charisma* seems to greatly influence students. Moreover, the intertwined nature of the characteristics of teachers who were able to generate attention and those of charismatic teachers might suggest a relationship between recognized and charismatic teaching, which was argued for by Archer (1994).

Charisma and Motivation

In the next section of the interview, we explored how charisma might contribute to the motivational power of L2 teachers in a tertiary environment, which was the second research question of our study.

Motivation Inside the Classroom. Each student reported that charismatic teachers notably enhanced their motivation inside the classroom. This is reflected in an excerpt from Bonnie's interview: "I felt as if I had been in elementary school again, because I studied for the teacher back then". Three other students also claimed that they felt more eager to participate in classroom activities as charismatic teachers had a greater influence on them and affected them more than other teachers. They considered constructive feedback from charismatic teachers to be more valuable and important, as it seemed more authentic from them. This is in agreement with Blašková et al. (2018), as they stated that charisma can bring about "unusual devotion" and "personal loyalty" (p. 3). Furthermore, five students reported increased levels of intrinsic motivation, which also had a beneficial effect on their in-class participation. Lee et al.'s (2014) and Bolkan and Goodboys's (2014)

findings are also firmly supported by these results: Lee et al. (2014) claimed that charisma has a positive effect on both situational and personal motivation, while Bolkan and Goodboy (2004) stated that charismatic teaching affects intrinsic motivation in a very positive way. Dorothy and Bonnie also highlighted the salutary effect of good group dynamics, which further enhanced the two students' motivation to attend and actively take part in lessons. Their statements also corroborated Suryani's (2016) findings regarding the beneficial impact that charismatic language teachers' have on the classroom environment, which may further increase students' eagerness to develop their skills.

Motivation Outside the Classroom. Although all eight students claimed that preparing for the classes of their charismatic teachers was taken seriously, three of them said their motivation did not grow significantly outside the classroom, as they did not prepare using extra materials at all. In contrast, five students claimed that their motivation substantially increased, leading to them preparing considerably more between classes due to their teachers' motivating agency. Moreover, each student said they prepared their homework more diligently in the charismatic teachers' classes compared to others. Interestingly, Mary claimed that she did not study more for these lessons, nor did she find the materials more intriguing; however, she was more eager to prepare and always took learning seriously because of the teacher. Furthermore, Reya stated that in spite of the failures and lack of self-confidence she had experienced in the past, the trusting relationship between her and her teacher helped her regain confidence and motivated her to perform on a much higher level than before. Mary's claim is in agreement with Holladay and Coombs (1994) and Antonakis et al. (2011), as they argued that the perception of charisma depends on how information is communicated, not on what the content of the information is. Findings from Ryan et al. (1994) and Ștefănescu et al. (2020), showing that students' self-worth is strengthened by trusting and caring student-teacher relationships are also reinforced by Reya's statement.

Teacher Interviews

The Qualities of Teaching Charisma

The interviewees described teachers who are able to seize students' attention in very diverse ways. Interestingly, only three teachers highlighted the importance of subject knowledge and methodological expertise, albeit it can be argued that being knowledgeable is rather essential at this stage of one's career (the average age of the eight teachers was 61.3 years). Being prepared, creating a well-thought-out structure for the class, being empathic, and creating rapport between the teacher and the students by fostering meaningful interactions were all listed as important attributes of teachers able to capture students' attention. Being credible, meticulous, motivated, understanding, and well-intentioned were further qualities listed as well as being able to create a friendly atmosphere, provide constructive feedback, allow students to choose topics they are interested in, and inject their personality into the lessons.

When the teachers were asked to define teaching charisma, their answers were highly diverse; however, similarities did emerge in their answers to a certain extent. Julia claimed that "charisma is a gift of the personality with which one can influence students", which is in line with Weber's (1968) definition of charisma, which was described as "a gift of the body and spirit not accessible to everybody" (p. 19). Margaret said that charismatic teachers were similar to charismatic leaders and could influence students through their personal qualities as well as their presence. She defined presence as a quality which causes listeners to literally stare at the person possessing it. Ulbricht also highlighted how charismatic teachers were able to attract the students' gaze. Furthermore, Lucas argued that charismatic teachers are able to seize the attention of their students (and can do so by simply loving their job and leaving their bad mood outside the classroom). The definitions provided by Lucas, Ulbricht, and Margaret were also considerably congruent in the sense that all three of them stated charismatic teachers could influence and attract students by their different personal qualities.

Felix reasoned that charisma was a feature which was appealing to others and made it easier for charismatic individuals to connect with others. This was also stated by Magnolia, who listed several similar qualities, specifically being funny, easy-going, charming, and kind. Their statements confirm Bolkan and Goodboy's (2014) claim, as they suggested that interactions between charismatic leaders and subordinates played a pivotal role in the perception of charisma. However, Magnolia also claimed there could be other kinds of charisma, such as knowledge, strictness, and proficiency charisma. Furthermore, Zachary stated that an effective L2 teacher should be expected to perform on at least an average level in every possible aspect of L2 teaching (e.g., classroom management skills, familiarity with the teaching material, etc.); however, he claimed that if one could rise above the average in one or more aspects, their excellence would make that person charismatic. This idea is in line with House's (1977) definition, according to which the personal abilities of charismatic leaders give them the ability to significantly affect their followers. Moreover, Huang and Lin (2014) also stated that charismatic language teachers are characterized by outstanding *knowledge, character traits, teaching techniques,* and *humor.*

The most unique definition of charisma was provided by Charlie: "it is a personality trait (not a professional quality), an undefinable power which makes the charismatic person unafraid of everything. She or he can handle any situation, is open, accepting, interested in others, and happy". His thoughts are partly related to Bolkan and Goodboy's (2014) research, as they found *nonverbal immediacy*, *humor*, *caring,* and *confirmation* to be the most pivotal aspects of charismatic teachers' communication.

Charisma and Motivation

The third part of the interview guide aimed to explore whether the motivational power of L2 teachers in a tertiary environment is enhanced by charismatic teaching. As seven out of the eight teachers did not find themselves entirely charismatic, they opted to explain what they found central in the creation of motivation. They described how they attempted

to maintain students' motivation inside and outside the classroom. In addition, they were also required to think of what made them charismatic in the students' eyes, as they had been perceived to be charismatic by at least one out of eight student participants during the previous round of interviews.

Motivation Inside the Classroom. The eight teachers listed several aspects in connection with arousing and maintaining motivation inside the classroom, highlighting that classes should be diverse, dynamic, and well-structured; several teachers emphasized the importance of preventing boredom by energizing classes. As stated by different teachers, this can be achieved by keeping students engaged (e.g., using different activities or games), by creating and nourishing good group dynamics (e.g., by promoting interactions between students), and by using different task types during a lesson. Dörnyei's (1994) inclusion of the *group* as one of the main motivational dimensions in SLA is reinforced by these statements.

Furthermore, three teachers stated that being approachable, listening to the students' needs, and providing them with constructive feedback were also crucial in keeping them motivated inside the classroom. These views are aligned with Bolkan and Goodboy's (2014) findings that *caring* is a substantial quality of charismatic teachers. Suryani's (2016) claim that charismatic teachers are likely to "develop caring, motivating and trusting relationships" (p. 16) is also reinforced by the results. Ștefănescu et al. (2020) also mentioned that charismatic teachers are more inclined to bond with their students.

Moreover, participants also claimed that students need to be provided with the opportunity to express themselves, and that teachers should present topics which students are generally interested in. This is in line with Suryani's (2016) claim that charismatic L2 teachers promote a supportive atmosphere in which students are encouraged to express their perspectives and voice their opinions. Julia revealed how she attempts to maximize motivation inside the classroom. She claimed that

teachers have a major leadership role and should accept it with all of its pros and cons. As she put it:

> One must create a system in which it is worth it for students to work hard. In my opinion, creating such a system is the responsibility of the teacher together with consistently monitoring and supervising students' work. One needs to dig a canal, so students can flow their energy into it, but the dam has to be built high in order to prevent them from running over on one side or the other. They must stay in the channel which will take them to reaching their goals.

Julia's thoughts reinforce those of House (1977), as he stated that charismatic leadership has the following features: strong self-assurance, resilient belief in one's own principles, and a forceful aspiration to affect and lead others. The results are also congruent with a finding from Antonakis et al. (2011) that followers of charismatic leaders are likely to be dedicated to the cause of the leader.

Motivation Outside the Classroom. Julia's claim above can also be connected to motivation outside the classroom, as the atmosphere she mentioned is not restricted exclusively to in-class work. Thus, the leadership role of teachers does not cease to exist outside the walls of institutions. Two teachers highlighted the importance of the quality of tasks that students must complete between lessons. It was also stated that the exercises should fit into a well-thought-out structure and should be interesting to students. As the aforementioned aspects can be part of the atmosphere described by Julia, it can be implied that teachers' leadership role overreaches the borders of the classroom and confirms House's (1977) claims that the qualities of charismatic leaders (e.g., a resolute desire to affect others) significantly affect students between lessons, as well. Therefore, motivation inside and outside the classroom may be tightly intertwined.

Comparative Analysis

As the two interview guides contained questions addressing similar concepts from the students' as well as the teachers' perspectives, in this section of our research the responses of the students and the teachers are juxtaposed, resulting in a comparative analysis comparing the results of both sets of interviews.

The Attributes of Teaching Charisma

The intertwined nature of the features of teachers who are able to capture students' attention and the features of charismatic teachers became more and more apparent while conducting the research. There were several qualities mentioned by the students which overlapped both categories. These features included knowledge, humor, and positive character traits (e.g., confidence, empathy, kindness, and patience). In view of the above, charismatic teachers were regarded as knowledgeable by five students and three teachers, which reinforces the findings of Huang and Lin (2014) and Lee et al. (2014), as they found knowledge to be a crucial determinant of charisma. However, whereas charismatic teachers were perceived to be rather humorous by the students, only one teacher believed that charismatic teachers could be easy-going and funny. Therefore, similarly to previous research, humor seems to be a controversial quality: Huang and Lin (2014) and Ștefănescu et al. (2020) pointed out the importance of humor in determining charisma and facilitating motivation, whereas Lee et al. (2014) did not regard it as a focal aspect. Certain positive character traits were mentioned by both groups of participants, with charismatic teachers being perceived as communicative, empathic, kind and well-intentioned. As most (if not all) of the studies on the topic claim that they are generally perceived to have good inner qualities, both groups' statements reinforce the results of the existing research.

Interestingly, the biggest contrast between the two groups was in regard to the judgment of the teachers' own charisma, as each teacher was considered to be charismatic by at least one student (the "most

charismatic" teacher received three nominations), whereas only one teacher saw himself as charismatic. In spite of their disagreement regarding the students' perceptions of their hypothetical charisma, three teachers acknowledged that there might be some elements of their teaching which made them seem charismatic. On the other hand, four teachers did not find themselves charismatic at all. The disparity between the two groups further bolsters the legitimacy of Blašková et al.'s (2018) definition of charisma, as they claim it is a hypothetical quality.

Furthermore, there was a notable difference between what the students and the teachers considered to be the most prominent attributes of charismatic teachers. Students found knowledge, passion, and self-knowledge to be the most important traits. On the other hand, the teachers believed that the way they managed student-teacher relationships (through empathy, energetic interactions, and informal teaching) as well as their devotion and preparation made their students see them as charismatic. The characteristics of devotion and passion were the single overlap between the two groups' responses, as both the student and teacher participants emphasized the significance of enthusiasm inside the classroom. As each student highlighted positive character traits as important indicators of teaching charisma, it can be claimed that relationships (as emphasized by the teachers) have as crucial of a role as knowledge in the nature of charisma; however, students expect teachers to be first and foremost professional at an academic level. Nevertheless, both groups' responses are harmonious with the findings of Archer (1994), as he reported that both intellectual expertise and personal virtue can attract students.

Motivation Inside and Outside the Classroom. Each student claimed that their eagerness to participate in in-class activities was significantly increased by charismatic teachers, whereas five students said their charismatic teachers affected their intrinsic motivation in a considerably salutary way; thus, their diligence also developed outside the classroom. The students claimed their increased enthusiasm was facilitated by positive group dynamics, constructive feedback from, and

devotion to their teachers. Several teachers also claimed that creating energetic group dynamics, providing the students with constructive feedback, and listening to their needs were critical in regard to motivating their learners. These statements are in agreement with Dörnyei's (1994) theory on motivational components, according to which the three dimensions of motivation are *subject* (in this case English, as university students are likely to be motivated to develop their L2 skills), *group* (both groups emphasized the salutary influence of energetic group dynamics), and *teacher*, as students stressed their affection and devotion to their instructors. It was also stated by four teachers that personal contact outside the classroom and personal advice could have a beneficial effect on students' motivation. Intriguingly, the importance of structuring the course syllabus was only highlighted by the teachers, which may be explained by their holistic perception of their courses. Nevertheless, both the students' and the teachers' responses are consistent with the findings of Ștefănescu et al. (2020) and Bolkan and Goodboy (2014) that charismatic teachers are more inclined to form closer ties with their students.

Implications

Attempting to develop one's charisma is likely to be an uncertain and laborious procedure, and the learnability of the trait might even be called into question. However, a charismatic person likely possesses the ability to influence and inspire others; hence, aspiring to become more charismatic in the eyes of students may have a positive impact on the motivational agency of teachers. Therefore, even if it is seemingly impractical to organize university classes aiming to develop the charisma of teacher trainees, it is likely to prove useful in raising awareness of the abilities of charismatic leaders and teachers, as it was agreed upon by most interviewees that charisma could be developed to some extent.

The students interviewed found knowledge to be the most important quality of charismatic teachers, while the teachers regarded knowledge, preparedness, and performance inside the classroom as

pivotal elements for motivating students. As argued by Archer (1994), being able to attract and influence students without having substantial subject knowledge and methodological expertise could be considered mere charlatanism, as teachers possessing *good charisma* are expected to be highly knowledgeable both in their fields and in regard to teaching methodology (Lee et al., 2014). With this in mind, teachers who would like to be influential and motivating inside the classroom should always be intellectually prepared in addition to having the ability to attract students' attention. Possessing positive character traits is also generally hailed as a prominent quality of charismatic (and outstanding) teachers. Archer (1994), Huang and Lin (2014), and Lee et al. (2014) all stressed the importance of being approachable, caring, empathic, kind, and patient, as they are essential features of charismatic teachers. Furthermore, Suryani (2016) also associated charismatic language teachers with similar inner qualities.

Although opinions on the learnability and trainability of charisma are rather divided, most participants in the interviews argued that charisma could be acquired to a certain degree. As shown by Antonakis et al. (2011), the perception of people's charisma can be notably enhanced even in a short period of time; thus, providing language teachers and teacher trainees with courses focusing on MotSs, charismatic leadership tactics, and leadership skills might have a positive effect on the way they motivate learners.

Finally, charisma is likely to develop with age and experience. Several times during the interviews, both the students and the teachers implied that language teachers need time to become confident inside the classroom. Becoming knowledgeable and developing thorough self-knowledge is a long and difficult journey. If charisma is simply a set of behaviors, it indeed can be learnt (Archer, 1994), but as most participants of the research claimed, the concept of charisma is probably more complicated than that. Charismatic behavior is seemingly learnable, but if a teacher aims to develop real charisma, they need to master the language, become familiar with teaching methodologies and understand how their

personality works in order to be able to improve their motivational teaching practice. Charismatic behavior might be built on the aforementioned foundations.

Conclusion

This study aimed to explore the characteristics of charismatic language teachers and how charisma contributes to the L2 motivation of students in tertiary contexts. The qualitative research was based on 16 interviews with eight teacher trainees and eight teachers at a Hungarian university in Budapest. Concerning the first research question (What makes a language teacher charismatic in a tertiary context?), language teachers' subject and methodological knowledge as well as the way they manage their relationships with their students play a crucial role in the perception of their charisma.

As the interviewees of both groups highlighted knowledge, preparedness, and positive character traits (e.g., approachability, caring, empathy, kindness, patience, etc.) alongside caring relationships with students as pivotal characteristics of charismatic teachers, it can be assumed that charismatic teachers seem to have both outstanding amounts of knowledge and the willingness to put considerable effort into building their relationships with their students. Methodological knowledge is also likely to be of major importance in the assessment of L2 teachers as reported by several interviewees**.** However, it has to be emphasized that some student participants scrutinized their teachers' methodological expertise in a more thorough way, as they did not exclusively regard them as their L2 teachers, but also role models to learn TEFL from.

The second research question (How does charisma contribute to the motivational power of L2 teachers in a tertiary environment?) investigated the motivational power of charismatic language teachers. Teaching charisma appears to have a highly positive effect on learners' motivation. Charismatic L2 teachers are devoted and use their positive character traits

to establish caring relationships with their students. Moreover, the expertise of charismatic L2 teachers can also inspire students to put more effort into their own development. Thus, it seems likely that the motivational power of charismatic teachers lies in their most prominent qualities, such as care, empathy, kindness, patience, and approachability. Their relationships with their students generate devotion and loyalty, while their knowledge seems to have a beneficial effect on their students' situational and intrinsic motivation.

While similarities regarding how the interviewees viewed charisma were discovered during the analysis of the interview transcripts, marked differences were also uncovered. The teacher and student participants found different behavioral aspects pivotal in the presence of charisma. The scope of this paper did not make it possible to reveal the underlying reasons behind the dissimilarities that emerged during the interviews. Further qualitative and quantitative studies would be required to analyze this variance.

Although the interview guide for this study was thoroughly designed, it still has its limitations. Each student interviewee had to rely on their own memory when they were asked to assess their L2 teachers' motivational power and charisma, as all of them were graduate students at the time of the interviews and had taken the courses focusing on language development a few years before the research. Likewise, several teachers had already retired or stopped teaching English as an L2 by the time the interviews were conducted.

We would like to conclude this paper by proposing possible future research directions. It would be interesting to repeat the interviews with secondary school students and teachers, or with adult learners and their teachers. It would also be exciting to see if L2 teachers' charisma could be developed markedly in a short period of time, similarly to the experiment conducted by Antonakis et al. (2011). Finally, a quantitative study carried out in Hungary to shed light on the most important components of teaching charisma would also provide valuable data for future research on this topic.

Note: This study was funded by the Scientific Foundations of the Education Research Program of the Hungarian Academy of Sciences.

References

Antonakis, J., Fenley, M., & Liechti, S. (2011). Can charisma be taught? Tests of two interventions. *Academy of Management Learning & Education*, *10*(3), 374–396. https://doi.org/10.5465/amle.2010.0012

Archer, A. C. (1994). *The measurement of charismatic teaching in the college classroom*. Conference Paper, Walla Walla College.

Avolio, B. J., Reichard, R. J., Hannah, S. T., Walumbwa, F. O., & Chan, A. (2009). A meta-analytic review of leadership impact research: Experimental and quasi-experimental studies. *The Leadership Quarterly*, *20*(5), 764–784.

Blašková, M., Blaško, R., & Polačková, K. (2018). Motivation and charisma of university teachers. *Human Resources Management & Ergonomics*, *12*, 21–31.

Boeije, H. (2002). A purposeful approach to the constant comparative method in the analysis of qualitative interviews. *Quality & Quantity*, *36*, 391–409.

Bolkan, S., & Goodboy, A. K. (2014). Communicating charisma in instructional settings: Indicators and effects of charismatic teaching. *College Teaching*, *62*(4), 136–142.

Chan, L. (2014). Effects of an imagery training strategy on Chinese university students' possible second language selves and learning experiences. In K. Csizér & M. Magid (Eds.), *The impact of self-concept on language learning* (pp. 357–377). Multilingual Matters.

Conger, J. A. (1985). *Charismatic leadership in business: an exploratory study. Unpublished doctoral dissertation*. School of Business Administration, Harvard University.

Conger, J. A., & Kanungo, R. N. (1987). Toward a behavioural theory of charismatic leadership in organizational settings. *Academy of Management Review*, *12*, 637–647.

Csíkszentmihályi, M. (1997). Intrinsic motivation and effective teaching: A flow analysis. In J. L. Bess (Ed.), *Teaching well and liking it: Motivating faculty to teach effectively* (pp. 72–89). Johns Hopkins University Press.

Dörnyei, Z. (1994). Motivation and motivating in the foreign language classroom. *The Modern Language Journal*, *78*(3), 272–284. https://doi.org/10.1111/j.1540-4781.1994.tb02042.x

Dörnyei, Z. (2001). *Motivational strategies in the language classroom*. Cambridge University Press. https://doi.org/10.1017/CBO9780511667343

Dörnyei, Z., Csizér, K., & Németh, N. (2006). *Motivation, language attitudes and globalization*. Multilingual Matters. https://doi.org/10.21832/9781853598876

Dörnyei, Z., & Guilloteaux, M. J. (2008). Motivating language learners: A classroom-oriented investigation of the effects of motivational strategies on student motivation. *TESOL Quarterly*, *42*(1), 55–77. https://doi.org/10.1002/j.1545-7249.2008.tb00207.x

Ghanizadeh, A., & Moafian F. (2010). The role of EFL teachers' emotional intelligence in their success. *ELT Journal*, *64*(4), 424–435. https://doi.org/10.1093/elt/ccp084

Glas, K. (2016). Opening up 'spaces for manoeuvre': English teacher perspectives on learner motivation. *Research Papers in Education*, *31*(4), 442–461. https://doi.org/10.1080/02671522.2015.1049287

Guilloteaux, M. J., & Dörnyei, Z. (2008). Motivating language learners: A classroom-oriented investigation of the effects of motivational strategies on student motivation. *TESOL Quarterly*, *42*(1), 55–77. https://doi.org/10.1002/j.1545-7249.2008.tb00207.x

House, R. J. (1977). A 1976 Theory of charismatic leadership. In J. G. Hunt & L. L. Larson (Eds.), *The cutting edge* (pp. 189–207). Southern Illinois University Press.

Huang, Y. C., & Lin, S.H. (2014). Assessment of charisma as a factor in effective teaching. *Journal of Educational Technology & Society*, *17*(2), 284–295.

Kálmán, C. (2021). The teacher's role in motivating adult learners of English. In G. Tankó & K. Csizér (Eds.), *DEAL 2022: Current explorations in English Applied Linguistics* (pp. 163–196). Eötvös Loránd University.

Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)? *Contemporary Issues in Technology and Teacher Education*, *9*(1), 60–70.

Kubanyiova, M. (2012). *Teacher development in action*. Palgrave Macmillan.

Kunter, M., Frenzel, A., Nagy, G., Baumert, J., & Pekrun, R. (2011). Teacher enthusiasm: Dimensionality and context specificity. *Contemporary Educational Psychology*, *36*(4), 289–301. https://doi.org/10.1016/j.cedpsych.2011.07.001

Lamb, M. (2017). The motivational dimension of language teaching. *Language Teaching*, *50*(3), 301–346. https://doi.org/10.1017/S0261444817000088

Lamb, M. (2019). Motivational teaching strategies. In M. Lamb, K. Csizér, A. Henry, & S. Ryan (Eds.), *The Palgrave handbook of motivation for language learning* (pp. 287–306). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-28380-3_14

Lamb, M., & Wedell, M. (2015). Cultural contrasts and commonalities in inspiring language teaching. *Language Teaching Research*, *19*(2), 207–224. https://doi.org/10.1177/1362168814541716

Lee, D. C., Lu, J. J., Mao, K. M., Ling, S. H., Yeh, M. C., & Hsieh, C. (2014). Does teachers charisma can really induce students learning interest? *Procedia - Social and Behavioral Sciences*, *116*, 1143–1148. https://doi.org/10.1016/j.sbspro.2014.01.359

Mackey, A., & Gass, S. M. (2005). *Second Language Research: Methodology and Design*. Lawrence Erlbaum.

Mezei, G. (2014). The effect of motivational strategies on self-related aspects of student motivation and second language learning. In K. Csizér & M. Magid (Eds.), *The impact of self-concept on language learning* (pp. 289–310). Multilingual Matters. https://doi.org/10.21832/9781783092383-017

Mezei, G., & Csizér, K. (2005). Második nyelvi motivációs stratégiák használata az osztályteremben [Using L2 motivating strategies in the classroom]. *Iskolakultúra*, *12,* 30–42.

Noels, K. A., Clément, R., & Pelletier, L. G. (1999). Perceptions of teacher communicative style and students' intrinsic and extrinsic motivation. *The Modern Language Journal*, *83*(1), 23–34. https://doi.org/10.1111/0026-7902.00003

Oin, L. (2008). Quality of college English teachers: Problems and suggested countermeasures. *US-China Foreign Language*, *6*(1), 7–12.

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

Ryan, R. M., Stiller J. D., & Lynch, J. H. (1994). Representation of relationships to teachers, parents and friends as predictors of academic motivation and self-esteem. *Journal of Early Adolescence*, *14*(2), 226–249. https://doi.org/10.1177/027243169401400207

Ștefănescu, A., Cârciag, P. D., Boeriu, D. M., & Morar, S. C. (2020). The importance of humour and charisma to facilitate students' motivation for learning. *Educatia*, *21*(19), 162–168. https://doi.org/10.24193/ed21.2020.19.21

Suryani, A. (2016). Redefining the potential role of charismatic language teachers in creating supportive academic atmosphere through students' motivational arousal. *Jurnal Sosial Humaniora*, *9*(1), 59.

Thorner, N., & Kikuchi, K. (2019). The process of demotivation in language learning: An integrative account. In M. Lamb, K. Csizér, A. Henry, & S. Ryan (Eds.), *The Palgrave handbook of motivation for language learning* (pp. 367–388). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-28380-3_18

Voss, R., & Gruber, T. (2006). The desired teaching qualities of lecturers in higher education: A means end analysis. *Quality Assurance in Education*, *14*(3), 217– 242. https://doi.org/10.1108/09684880610678540

Weber, M. (1968). *Max Weber on charisma and institutional building*. University of Chicago Press.

With the third volume in our series, we would like to share with students and colleagues at Eötvös Loránd University and beyond some of the research work conducted at the Department of English Applied Linguistics (DEAL) in 2022. The volume features a collection of empirical and theoretical articles that address challenging issues and problematize existing orthodoxies in our field. While reflecting our range of research interests, the studies are also informative, thought-provoking, and innovative within as well as beyond our local context. The practical implications of the findings alongside the constructive criticism formulated present opportunities for critical reflection and change. We especially recommend the volume to applied linguistics tutors who already use research articles in their undergraduate and graduate courses or would like to do so in the future.

GYULA TANKÓ & **ATTILA M. WIND,** Editors

ISBN 978-963-489-550-3

ELTE | FACULTY OF HUMANITIES

EÖTVÖS UNIVERSITY PRESS