DEAL 2021

CURRENT EXPLORATIONS IN ENGLISH APPLIED LINGUISTICS

EDITED BY GYULA TANKÓ & KATA CSIZÉR

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

DEAL 2021:

Current Explorations in 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 & Kata Csizér

DEAL 2021:
Current Explorations in English Applied Linguistics

Edited by
Gyula Tankó & Kata Csizér

Budapest, 2021

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

ISBN 978-963-489-423-0
ISBN 978-963-489-424-7 (pdf)
ISBN 978-963-489-427-8 Ö

KUCSÁK
KÖNYVKÖTÉSZET ÉS NYOMDA
www.kucsaknyomda.hu

Executive Publisher: the Dean of Faculty of Humanities,
Eötvös Loránd University
Layout: Gyula Tankó
Printed by: Kucsák Könyvkötészet és Nyomda

Table of Contents

Contributors

Ágnes Albert (PhD, ORCID ID https://orcid.org/0000-0001-8339-7119) is an assistant professor at the Department of English Applied Linguistics at Eötvös University, Budapest and holds a PhD in Language Pedagogy. Her research interests include task-based language learning and individual differences with a focus on learner creativity and positive emotions related to foreign language learning. Currently, she is a member of a research team investigating the motivation, autonomy, and positive and negative emotions of secondary school learners about their English classes.

Zsuzsanna Andréka (https://doktori.hu/index.php?menuid=192&lang=HU&sz_ID=37585) is a PhD student in the Language Pedagogy and English Applied Linguistics PhD programme at Eötvös Loránd University, Hungary. Her main research area is language testing, but she is also interested in digital literacy. She is currently teaching language development courses at the university, focusing on improving students' language and communication skills through the use of different media of popular culture.

Kata Csizér (PhD, ORCID ID https://orcid.org/0000-0003-1755-8142) is an associate professor at Eötvös Loránd University, Budapest. Her main research interests are the social psychological aspects of L2 learning and teaching, as well as foreign language motivation. She has published over 100 academic papers and has co-authored several books, including the recent Palgrave Macmillan handbook of motivation for language learning with Martin Lamb, Alastair Henry, and Stephen Ryan. She is one of the two associate editors of the Journal of Second Language Writing (Elsevier).

Brigitta Dóczi (PhD, ORCID ID 0000-0002-1581-0079) teaches at the Department of English Applied Linguistics of Eötvös Loránd University, where she offers courses in vocabulary research, business English, pedagogical grammar, and pedagogical phonology. Her research interests are second language vocabulary development as well as task-based language teaching. Recently, she has been doing research on the role of verbal and visual modalities in fine arts education, resulting in the publication of a specialized bilingual art dictionary, as well as participating in a large-scale study investigating English language teaching in Hungarian secondary schools.

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. Her recently published book, *Understanding Context in Language Use and Teaching: An ELF Perspective* (Routledge, 2020) addresses issues relating to her main areas of interest.

Csaba Kálmán (PhD, ORCID ID https://orcid.org/0000-0003-4798-8637) 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. Currently, he is researching L2 teachers' extraordinary motivating agency. In addition to his academic activities, he has been running his own business offering a variety of corporate language training and skills courses for three decades.

Katalin Piniel (PhD, ORCID ID https://orcid.org/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.

Dávid Smid (ORCID ID https://orcid.org/0000-0003-4484-8591) is a PhD candidate in the Doctoral School of Education at Eötvös Loránd University (Budapest, Hungary). His research interests include the cognitive, affective, motivational, and social dimensions of L2 learner and L2 teacher psychology. His doctoral dissertation is about Hungarian pre-service English teachers' motivation. He has presented his research on three continents (Europe, Asia, and Australia) at various international conferences and co-organized the *1st International Conference on Sociolinguistics* (Budapest, Hungary). He has taught at tertiary level in Hungary and Japan.

Gyula Tankó (PhD, ORCID ID https://orcid.org/0000-0002-0721-8171) is an assistant professor at the Department of English Applied Linguistics at Eötvös Loránd University, Budapest, and has a PhD in Language Pedagogy. He teaches and researches English for Academic Purposes, Language Assessment (General English, English for Academic Purposes, and English for Diplomatic Purposes; Assessment Justification; Item Writing), and Discourse Studies.

Jamil Toptsi (https://doktori.hu/index.php?menuid=192&lang=EN&sz_ID=37641) is a PhD student at Eötvös Loránd University at the Language Pedagogy and English Applied Linguistics program. His main research interests include language learning motivation and autonomous language learning with digital technology. He also works as a lector at the department of English Applied Linguistics at ELTE, where he has developed a recent interest in the field of English for academic purposes.

Attila M. Wind (PhD, ORCID ID https://orcid.org/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 acquisition, more broadly, with a special focus on second language writing development and self-assessment. He is also interested in the application of Complex Dynamic Systems Theory to the study of second language acquisition.

Anna Zólyomi (ORCID ID https://orcid.org/0000-0002-9280-5775) is a PhD candidate in the applied linguistics and language pedagogy program at the Department of English Applied Linguistics at Eötvös Loránd University, Hungary. She is teaching at the university voluntarily as an external lecturer and is taking part in the New National Excellence Program (ÚNKP) in which she is conducting research on language aptitude and explicit/implicit learning. Besides these, she is mainly interested in foreign language learning-related emotions.

Editorial

We are proud to present the second reviewed volume in the applied linguistics series of the Department of English Applied Linguistics, Eötvös Loránd Univerity, an initiative proposed by our Head of Department, Enikő Öveges, in 2019. When she put forward the idea for the series, multiple aims were presented, among others, to further our understanding of the interdisciplinary research work done in our department, to share our research results with our students and fellow researchers, as well as to develop our personal communication and review skills. We think that we have managed to meet these aims as we worked through another year fraught with difficulties. We are profoundly grateful for her guidance.

This year's volume contains three different strands in applied linguistics through which we provide comprehensive insights into our field. Éva Illés examines the way the changing circumstances of English language use and teaching over the past decades have resulted in the form and meaning changes of five notions related to the study of English language use and teaching. In terms of theoretical approaches to the understanding of how foreign and second languages are learnt, Ágnes Albert's investigation of flow takes individual difference research in a largely uncharted territory, and through a review of the methods that have been used to measure flow in psychology and education as well of related empirical studies conducted in our field encourages applied linguists to investigate flow experiences in language learning. Gyula Tankó and Zsuzsanna Andréka present the results of an empirical investigation conducted in order to reveal the shortcomings of the Use of English paper of the B2 level advanced EFL school-leaving examination administered in Hungary, partly due to which the examination fails to fulfil its intended gatekeeping function, and to provide a framework of investigation for future studies.

The second chapter explores the most important context of second language acquisition research: the classroom. It is an unintended but welcome development that out of the four papers, three concentrate on the work of teachers, who are still a very much neglected population in our field. Brigitta Dóczi and Kata Csizér present the results of a classroom observation study, and in addition to providing a rare and informative insight into English teachers' daily work, they also contribute an observation protocol for the investigation of classroom processes in second language teaching. The study conducted by Dávid Smid and Anna Zólyomi intends to improve our understanding of English language teachers' beliefs about language learning and to inform pre-service L2 teacher education through the examination of the diverse effects of the social context on English language learning. Csaba Kálmán's results explain the teacher's roles in motivating adult learners of English in Hungarian corporate contexts, showing the importance of the personality and preparedness of the teacher in addition to such other key factors as appearance, personal branding, and syllabus design. Gyula Tankó's study motivated by concerns about potential threats to test validity contributes to the understanding of gender effect on fairness in the high-stakes yet sporadically researched, small-scale tertiary-level assessment of academic writing performance as well as informs our knowledge of gender-differences in foreign/second language learning. The chapter closes with an article in which through the exploration of the co-development of self-assessment accuracy and linguistic complexity over one academic semester Attila Wind provides insights into the dynamic development of self-assessment abilities and formulates their implications for L2 writing courses.

The final chapter contains two studies related to digital learning during the pandemic. Ágnes Albert and Katalin Piniel explore how students' emotions, motivation, autonomy and self-efficacy changed over the schoolyear marred by the difficulties induced by the COVID-19 crisis. Continuing the discussion of language learning under lockdown conditions, in the last paper Jamil Toptsi focuses on the autonomous learning behavior of students and how this learner characteristic together with mental states,

learning environment, and language learning experiences shaped their language learning during the lockdown.

We would like to thank the contributors for their high-quality work. As it is really rewarding to read and learn from colleagues' research, it was a joy to cooperate with all of you in the course of the preparation of this multiauthor work. Our reviewers shall remain anonymous, but their input greatly improved the quality of this volume. Monika Ford and James Griffin provided proof-reading of the manuscripts, for which we are thankful. Finally, we thank our university for the Textbook Grant (K15201/16) and the continuous support of our work.

We are ready to pass the baton to the next editorial team.

Happy reading!

Gyula Tankó & Kata Csizér

1.

Old Notions in New Contexts

Éva Illés
Illes.eva@btk.elte.hu

Abstract

The paper examines how the changed circumstances in the use of English over the past decades have resulted in changes in the terminology of the study of English language use and teaching. With English having become a global language used both intra- and internationally by speakers from diverse linguacultural backgrounds, the conceptualisation of established notions had to follow suit. As a result, there have been changes affecting both form and meaning. By drawing on the literature, the present study comprises the analysis of five notions which have been affected by the World Englishes context and the prevalent functioning of English as a global lingua franca. The five notions have undergone different changes, including the pluralisation of English, complementing definition by adding a new dimension, suggesting alternative terms, reviving side-lined conceptualisations, and using a novel term–all in order to be able to capture the current conditions of use more accurately. The analysis highlights the connection between the circumstances that have given rise to the modifications as well as the changes themselves.

Keywords: spread of English, changed circumstances of use, adjusting terminology

Old Notions in New Contexts

The global spread of English has resulted in fundamental changes in the use of English. The changed circumstances have affected not only the composition of users and the number and hegemony of native speakers, but also where and how English is spoken. In ex-colonial countries, such as India and Pakistan, English has been used as a second and/or official language fulfilling mainly *intranational* purposes, reflecting multilingual social realities which are very different from those countries where English is used mainly by monolingual native speakers (Kirkpatrick, 2007; Schneider, 2007). In many contexts of use, English also serves as a means of *international* communication where it functions as a lingua franca among speakers who come from a wide variety of linguacultural backgrounds (Seidlhofer, 2011).

Changes in use were followed by research and a reconceptualization of English in both intra- and international contexts and contexts where English functions as a lingua franca (ELF). The results of empirical and theoretical research have found their way into English language teaching (ELT), too, albeit somewhat slowly (Seidlhofer, 2012). Despite extensive research, which has managed to close the conceptual gap Seidlhofer (2001) highlighted when referring to the prevalent non-native-speaker use of English and native speaker norms as the target, traditional views of English still seem to prevail in some quarters of ELT (see, for example, Medgyes, 2014).

The efforts to reconceptualise English have inevitably resulted in the adjustment of terminology. Changes in form, frequency of use and meaning triggered by novel conditions of use is a process which can be observed in all language use. A representative example of this can be the word *pandemic* which was the Dictionary.com Word of the Year for 2020 (https://www.dictionary.com/e/word-of-the-year/) but was rarely used in everyday discourse before Covid-19. On 11 March, 2019, the Word Health Organization declared that Covid-19 was a pandemic:

> That day, when COVID-19 had then only taken 4,291 lives around the world, searches for *pandemic* skyrocketed 13,575% on Dictionary.com compared to 2019. *Pandemic* joined a cluster of other terms that users searched in massive numbers, whether to learn an unfamiliar word used during a government briefing or to process the swirl of media headlines: *asymptomatic*, *CDC*, *coronavirus*, *furlough*, *nonessential*, *quarantine*, and *sanitizer*, to spotlight a few. (https://www.dictionary.com/e/word-of-the-year/)

It was not only the frequency of use of the word *pandemic* which has increased dramatically but also the foregrounding of words which have been around for centuries and were used in restricted contexts before the Covid pandemic (e.g., furlough in military use, https://www.etymonline.com/word/furlough). Similarly to everyday discourse, in the study of English, the changes in use have resulted in formal, meaning-related and other modifications of terminology. A case in point may be the term World Englishes which has a plural form for a conventionally uncountable noun, and indicates a particular type of variety of English. In this paper, I will be investigating this and another four notions which are illustrative of the changes that have taken place as a result of the global spread of English. In doing so, I will discuss relevant linguistic and semantic issues as well as the circumstances which have given rise to the novel features of some selected terminology.

Throughout the paper, language change will be seen as a natural phenomenon. The approach adopted here corresponds to David Crystal's reply in an interview to the question of whether language change is a sign of progress or decay:

> The notion of decay makes no sense to me. For me, that's like going down to the seashore and looking at the waves coming in and going out and saying, 'That's a good wave, that's a bad wave.' The notion of change does not have any kind of evaluative significance for me, at all. (Mayevski, 2009, p. 60)

In effect, this study is also concerned with language change which is seen here as a natural and ongoing process which leads to novel features that are different but neither better or worse than the ones before them. The first term to be investigated is the already mentioned World Englishes.

World Englishes

It is probably an experience shared by many Word users that whenever the word Englishes appears, the plural form is underlined, meaning that Englishes is a grammatically incorrect form. This happens despite the fact that the journal *World Englishes* has been around since 1981 (https://onlinelibrary.wiley.com/journal/1467971x), so the form is not really that new. The question then arises as to what the plural form is meant to indicate and why does not reference to varieties of English suffice?

The World Englishes (WE) are used mainly in the Outer Circle of Kachru's model of the spread of English (Kachru, 1991). They are indigenised/nativised varieties that have been developed as a result of the wide range of functions English, often as an official and/or institutionalised language fulfils in countries such as India, Nigeria, Singapore or the Philippines (Pakir, 2009; Rubdy & Saraceni, 2006; Seidlhofer, 2011). In the literature, Englishes in the Outer Circle were often seen as less developed, faulty varieties of English. This long-held deficit view has been challenged by what Quirk called liberation linguistics (Kachru, 1991), which has been promoting linguistic emancipation and, with it, an endonormative perspective whereby WE speakers develop and apply their own norms, no longer conforming to the rules prescribed by privileged Inner Circle speakers. The WE perspective therefore views English as "pluricentric, with many new Englishes showing hybrid forms" (Pakir, 2009, p. 229). These Englishes are "defined not as flawed imitations of Standard English, but as legitimate forms of language that serve the needs of second-language speech communities" (Modiano, 2009, p. 217).

The Englishes spoken in many Outer Circle countries are thus seen as varieties of English in their own right, with their own particular lexical and grammatical features as well as distinct pragmatic use and discourse structure (Mahboob, 2009). They serve as rich resources for their users who exploit the meaning potential of the language to express their own values and identities. In the Outer Circle, English has become an "important tool to impart local traditions and cultural values" (Kachru, 1991, p. 219). They can function as an Islamic language in Pakistan, reflecting "a Pakistani Muslim identity" (Mahboob, 2009, p. 187), but can also carry the weight of the African experience of writers such as Achebe (as cited in Widdowson, 2003).

The plural form "Englishes" aims to highlight the diversity of these Outer Circle varieties of English and the fact that "English no longer has one single base of authority, prestige and normativity" (Mesthrie & Bhatt, 2008, p. 3). In addition, the plural form as well as the other term, New Englishes, used in reference to legitimate and often codified Outer Circle varieties reflect the fact that these Englishes are more recent than the varieties in the Inner Circle. Whereas English spoken in North America, Australia and New Zealand emerged as a result of the first dispersal of English when the language was "transported to the 'New World'" (Jenkins, 2009b, p. 5), it was the second diaspora, colonialisation and English being transported to Asia and Africa, which led to World Englishes.

As has been pointed out above, World Englishes are defined as nativised varieties as opposed to the native varieties of English in the Inner Circle. According to Kirkpatrick (2007), nativised varieties are "newer varieties that have developed in places where English was not originally spoken and which has been influenced by local languages and cultures" (p. 5). Interestingly, it is Kirkpatrick himself who challenges the native v. nativised variety distinction, claiming that the definition of nativised varieties, in fact, applies to native varieties, too, as there were other languages before English in England or English in North America. What differs, however, is the length of time since the different varieties of English

have been around. In this respect, the less frequently used New Englishes appears to be the term which more aptly describes the varieties used in Asia and Africa than the much more widespread notion of World Englishes. The subtitle of Mesthrie and Bhatt's book (2008), "World Englishes: *The study of new linguistic varieties"* (my emphasis), lends support to this argument.

English as a lingua franca

Interestingly, English as a lingua franca became an established term in applied linguistics with some delay after English had already been widely used as a lingua franca, with research following suit. Even though Jenkins, an ELF researcher who started her empirical research into pronunciation in the late 1980s (Jenkins, 2015), published the findings of her investigation with the title "The phonology of *English as an international language*" (2000, my emphasis). The reason for choosing English as an international language rather than lingua franca was the fact that the notion of ELF was not really known to applied linguists at the time (Jenkins, 2015). As Jenkins (2015) put it in a footnote:

> Although I attempted to introduce the term 'English as a Lingua Franca' and its acronym 'ELF' in Jenkins 1996, this proved difficult (see Jenkins 2000: 11), and for the few years that followed, ELF researchers, myself included, tended to use the more transparent 'English as an International Language' and its acronym 'EIL'. (p. 53)

Even though the article which initiated the corpus linguistic research of the lingua franca use of English made reference to ELF (Seidlhofer, 2001), the report of the initial findings of the then already materialised Vienna-Oxford International Corpus of English (VOICE) hypothesised the existence of Euro-English, a possible variety in its own right (Jenkins, Modiano & Seidlhofer, 2001). Researchers of the corpus comprising 1 million words of spoken data arrived at this early conclusion because the corpus data

suggested that there may be features which are specific to the kind of English used by the participants, a thousand, mainly European speakers of English with about 50 different first languages. At this stage, which Jenkins (2015) calls "ELF phase one ('ELF 1')" (p. 52), ELF was seen as a new variety with the potential of becoming the new model that would be able to provide an alternative to English as a native language with the possibility of constituting a "new resource for the design of English instruction" (Seidlhofer 2001, p.151).

However, further research into the nature of ELF communication revealed that features assumed to be typical of ELF forms can, in fact, be found in both native and post-colonial varieties of English (Jenkins, 2012; Sewell, 2013). The universal question-tag *innit* is used in Inner Circle Estuary English (Eitler, 2006) and when it comes to the omission of the third person singular *-s*, similar tendencies for regularisation can be observed postcolonial varieties of English (Schneider, 2007).

These findings as well as observations such as the high "online variability" of ELF (Jenkins, 2012, p. 490) communication have led to the conclusion that "ELF cannot be conceptualised as a language variety" (Jenkins, 2012, p. 490). Therefore, ELF researchers nowadays refer to ELF as a "legitimate branch of modern English" (Jenkins, 2007, p. 17) with many emerging and changing varieties, and as "continuously negotiated, hybrid ways of speaking" (Seidlhofer, 2011, p. 4).

The shift of focus from form to function and the definition of ELF in pragmatic terms (for example, Illés, 2020; Jenkins, 2009a; Park & Wee, 2011) has resulted in the consolidation of the term English as a lingua franca, which stresses the function English fulfils in the communication of speakers from diverse linguacultural backgrounds. In this sense, the word 'as' in ELF indicates *how* English is used rather than the spread of use as in the case of English as an international language.

Applied to the new context, the notion of lingua franca has also undergone changes in meaning. A lingua franca is traditionally defined as a language used to facilitate communication between native speakers of different mother tongues (Meyerhoff, 2006; Wardhaugh, 2006). As has been argued above, ELF is conceptualised in pragmatic rather than linguistic terms. In addition, the use of English as a lingua franca is not restricted to a particular geographical location like Swahili in West Africa. ELF is, rather, a global, "non-local lingua franca" (Mauranen, 2018, p. 7) which is used in all kinds of contexts all over the world. Another novel feature of ELF is that it includes native speakers of the language who make up the minority of all users of English.

Native and non-native speakers of English

The native–non-native English speaker dichotomy has been in the centre of heated debates for several decades (Medgyes, 2017). Issues surrounding the distinction were raised already in the 1990s (Medgyes, 1994). Medgyes (1994) highlighted the difficulties the definition of the native speaker entails. With the help of a selection of criteria found in the literature (e.g., a native speaker is someone who speaks English as their first language or was born in an English-speaking country, etc.), Medgyes, in fact, proved that all of investigated categories were very loose and fuzzy.

World Englishes (Mesthrie & Bhatt, 2008) and ELF research, with their focus on non-native speaker use and ownership of English, has brought the native and non-native speaker distinction to the fore again. When discussing the viability of the dichotomy, Jenkins (2009b) argued that since speakers "'own' their lingua franca English" (p. 87), it makes no sense to refer to them as non-native speakers. However, the distinction can be applied to English as a foreign language where the aim is to learn the language to communicate with L1 speakers.

In order to replace the irrelevant native and non-native speaker distinction, Jenkins (2009b) proposed the following tripartite alternative: monolingual English speaker (MES), bilingual English speaker (BES) who is a proficient speaker of English and at least another language, and the non-bilingual English speaker (NBES) who can speak English "at a level of reasonable competence" (p. 90). The suggested reconceptualisation aimed to remove the negative connotation the notion of the non-native speaker carries and present BES as more favourable than MES, thus reflecting the view that bi- and multilingualism is the prevalent and preferred condition when it comes to the use of English worldwide. However, as Jenkins (2014) herself admitted, the distinction between BES and NBES is fuzzy because, due to the lack of stringent criteria, it is difficult to determine who counts as a proficient speaker and what the level of reasonable competence entails. The suggested solution to the problem in the future was the development of devices which can measure "intercultural ELF communication skills and the use of accommodation strategies" (Jenkins, 2014, p. 39).

When reflecting on the various possible outcomes regarding the use of the native and non-native dichotomy, Jenkins (2014) offers a hypothesis which centres on the changing connotations of the native and non-native terms. During the time of the British Empire, the natives were indigenous people living in the colonies who were considered primitive and uneducated, thus attaching negative connotations to the notion. However, with the global spread of English and the empowerment of speakers in the Outer and Expending Circles, the term native has gained positive connotations. Native speakers were and are still often seen as the ultimate authority with superior language skills. In contrast, in ELF and World Englishes contexts, non-native speakers have the advantage in that they are more competent and can communicate effectively in international contexts of use. Jenkins (2014) thus envisages a reversal in connotation where the notion of native acquires pejorative meaning again. Interesting thoughts from a native speaker.

Seidlhofer (2011) assumes that the native versus non-native distinction will not only become obsolete, but will disappear altogether:

> My feeling is that, after so much agonizing about the terms 'native' and 'non-native', the problem will actually resolve itself in that new and appropriate words will emerge. It may well be that in the not too distant future we will be wondering why we ever thought that we needed these terms at all. (p. 6)

Seidlhofer's prediction may prove reliable in that what will disappear with the native versus non-native dichotomy is, in fact, an idealisation. Widdowson (2012) argued that the often evoked notion of the educated native speaker is "an idealized construct, a convenient abstraction which is, paradoxically enough, on a par with Chomsky's ideal speaker-hearer" (p. 13). In real-life, native speakers, like other speakers of English or any language, come in all shapes and sizes, and who is considered educated may also vary. As Widdowson (2012) pointed out, the standard is sanctioned by authority and comprises the language that has been codified by linguists rather than the language used by its native speakers. It seems that, as in the case of everyday language use, time will tell and the communicative needs of language users will determine what happens to the much debated dichotomy of native versus non-native speaker.

Authenticity

In the case of authenticity, the changed contexts of use of English have resulted in the use of the same term but with a meaning that has been around for a long time but has been side-lined. World Englishes and ELF in particular created the conditions in which this neglected conception of authenticity appeared to be more suitable than the currently prevalent understanding of the notion. Although much has been written about authenticity, especially in the ELT literature (e.g., Buendgens-Kosten, 2014;

Clavel-Arroitia & Fuster-Márquez, 2014; Guariento & Morley, 2001; Siegel, 2014), the different meanings the notion carries can be best captured from a pragmatic perspective.

The prevailing conception of authenticity is "actually attested language produced by native speakers for a normal communicative purpose" (Widdowson, 1983, p. 30). Authentic texts, therefore, are those which are not pedagogically modified and which students come across when they communicate with speakers of L2 (Harmer, 2007, p. 273). Authenticity is thus necessarily linked up with the target language community (Guariento & Morley, 2001), its speakers and the speakers' knowledge of the world, including knowledge about their particular culture and subculture(s). In ELT, where appropriateness entails complying with native speaker norms (Seidlhofer, 2011), it is assumed that through exposure to authentic texts and tasks, students will learn to respond appropriately, that is, in a native-like manner. It is claimed that authentic texts are also motivating because they present real language (Guariento & Morley, 2001) and meet students' long-term communicative needs, which is being able to communicate successfully with the speakers of the target language (Yuk-chun Lee, 1995).

According to Widdowson (1990), authenticity whose ultimate aim is native speaker correctness of language and appropriateness of language use is "bound to be, to some extent, an illusion" (p. 44). The reason for this lies in the nature of communication. Texts are produced for a particular audience and assume some shared knowledge which provides the common ground, and usually remains unsaid. With the help of cues provided by the language, the reader/listener activates the knowledge that is necessary to fill the gaps the language has left, and interprets the text. For example, the following text was written for real-life purposes and for readers with whom knowledge about the current political situation in Britain is taken for granted, but can the reader of this paper make sense of it? The text is the headline of an article published on a news-site: 'No wonder dumb-loyal Bojo babe Therese Coffey got Universal Credit sums wrong' (https://newsotime.com/uk-news/no-wonder-

dumb-loyal-bojo-babe-therese-coffey-got-universal-credit-sums-wrong/). The headline is authentic for a particular readership who can relate it to their reality, but probably remains, by and large, incomprehensible for those not belonging to this group.

The alternative conception of authenticity, learner authenticity (Yul-chun Lee, 1995) or process authenticity (Murray, 1996), focuses on the learner, their reality and knowledge of the world. Authenticity here "is related to the learner as language user's context and is a matter of presenting language that learners can authenticate for themselves" (Widdowson, 2003, p. 115). As a consequence, an authentic text in this view is not necessarily what has been written by and for native speakers but one which engages the learner's own reality and relates to the learner's experience of the world. Without this connection, no understanding or learning takes place: "You cannot make sense of anything without bringing it within the confines of what is preconceived as familiar. Everything new has to be related to what is given" (Widdowson, 2007, p. 28). Therefore, for materials to be authentic, they need to be made real and local for the particular group of learners in the classroom. This may necessitate pedagogical tampering with the materials that have been produced for a different or wider audience.

It is learner authenticity which seems better suited if the aim is to prepare learners for communication in ELF contexts of use. The main reason for this is that the moment learners step outside of the classroom, they function as users of the language (Seidlhofer, 2011), which implies that they have to engage on their own terms both linguistically and schematically–the way they do it in their first language. In such interactions correctness of language and appropriateness of language use is not measured against the norms of native speaker use but in relation to the success of the communication as judged by the interlocutors themselves. In the classroom, therefore, teachers need to create conditions which enable learners to function as language users who engage in the process of authentication, rather than being given the authentic response of native speakers.

Learner autonomy

ELF communication has given rise to an amended definition of learner autonomy which entails complementing the notion with an additional dimension. As in the case of authenticity, an early conception of autonomy originating in adult education (Benson, 2007; Little, 2007) has dominated ELT. It has its roots in the most prevailing definition of autonomy by Holec (1981). According to Holec, autonomy is "the ability to take charge of one's own learning" (p. 3). In more practical terms, this means that the responsibility for various aspects and stages of language learning is transferred from the teacher to the learner. This includes setting goals, selecting resources and strategies, and assessing progress (Cotteral, 2000). Since being able to control the learning process requires pedagogic decision making, in order to be autonomous, learners need explicit knowledge of language learning which, in turn, necessitates training–as is the case with teachers. This training then has to be added to the language courses and should comprise the acquisition of metaknowledge and metalanguage (Illés, 2012). Critiques of this perception of autonomy often question whether the added-on training can indeed result in more effective learning and better proficiency. The alternative definition of autonomy has grown out of the dissatisfaction with the preponderance of pedagogic decisions which should be within the remit of teachers.

Whereas the prevalent definition focuses on pedagogical issues, the main concern of the amended definition is language use. The reason for this is the fact that in ELF contexts, learners function as language users in their own right who need to exploit the linguistic and other resources at their disposal in order to be able to successfully cope with the diversity and increased demand for negotiation ELF communication presents. As a consequence, the aim of ELT should be to facilitate learners' becoming competent language users who can successfully grapple with the requirements of ELF interactions under their own steam. The amended definition of autonomy, therefore, should expand the scope of autonomy and

include the ability to manage and control language use, which should in return, affect the learning process (Illés, 2012).

The conception of autonomy with a focus on the learner as a language user also resolves the conflict that exists between the traditional perception of authenticity whose concern is native-speaker language use and the concept of learner autonomy. Widdowson (1996) argued that the two prioritise different realities, and that "the language which is real for native speakers is not likely to be real for the learners" (p. 68). When, however, authenticity entails the engagement of the learners' reality, the two notions stop being incompatible.

Conclusion

Drawing on the literature, this paper aimed to investigate how the changed circumstances in the use of English have affected some of the terminology in applied linguistics. The analysis of five notions has revealed that there have been alterations in both form and meaning. The plural form in World Englishes indicates a change in the usage of an otherwise uncountable noun in order to highlight the diversity of the varieties involved and the distinction from the kind of English spoken in the Inner Circle. On the other hand, English as a lingua franca is a novel term which emphasises the function English fulfils in international communication. As a result, 'as' in the expression refers to how English is used rather than where it is spoken as is the case with the notion of English as an international language. It seems that while there appear novel terms, there are notions which are becoming obsolete and which will either peter out or will be replaced by new concepts. An example of this in this paper is the native and non-native speaker dichotomy which does not pertain to ELF contexts of use where all participants are native speakers of their ELF.

Changes in the use of English are beginning to filter through to the teaching of English. The sidelined conception of authenticity, which relates understanding and producing the target language to the learners' reality, has proven to be more suitable for preparing learners for ELF interaction where learners as language users engage in and on their own terms. The last type of change in the terminology discussed in this paper is the addition of a new dimension to the definition of learner autonomy. In this case, the prevailing classroom learning-focussed perception of the notion has been complemented by a language-use oriented dimension which aims to enable learners to become autonomous language users.

When reading the literature, it is obvious that the changes in the use of English have been more profound than the investigation in this paper might suggest. World Englishes and ELF researchers call for the rethinking of many well-established notions, such as community, competence, and legitimate speakers of a language (https://www.slideshare.net/engnet/prof-barbara-seidlhofer). Widdowson (2012) goes further and advocates an overhaul of the categories and methods used for the description of English. Lots of food for thought and plenty of challenges for applied linguistics these days.

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

Benson, P. (2007). Autonomy in language teaching and learning. *Language Teaching*, *40*(1), 21–40.

Buendgens-Kosten, J. (2014). Authenticity. *ELT Journal*, *68*(4), 457–459.

Clavel-Arroitia, B., & Fuster-Márquez, M. (2014). The authenticity of real texts in advanced English language textbooks. *ELT Journal*, *68*(2), 124–134.

Cotteral, S. (2000). Promoting learner autonomy through the curriculum: Principles for designing language courses. *ELT Journal*, *54*(2), 109–117.

Eitler, T. (2006). Identity construction, speaker identity and Estuary English. In L. Varga, (Ed.). *The even yearbook 7* (pp. 1–14). ELTE SEAS.

Guariento, W., & Morley, J. (2001). Text and task authenticity in the ELF classroom. *ELT Journal*, *55*(4), 347–353.

Harmer, J. (2007). *The practice of English language teaching* (4th ed.). Pearson Education Limited.

Holec, H. (1981). *Autonomy in foreign language learning*. Pergamon.

Illés, É. (2012). Learner autonomy revisited. *ELT Journal*, *66*(4), 505–513.

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

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford University Press.

Jenkins, J. (2007). *English as a lingua franca: Attitude and identity.* Oxford University Press.

Jenkins, J. (2009a). English as a lingua franca: Interpretations and attitudes. *World Englishes*, *28*(2), 200–207.

Jenkins, J. (2009b). *World Englishes: A resource book for students* (2nd ed.). Routledge.

Jenkins, J. (2012). English as a lingua franca from classroom to classroom. *ELT Journal*, *66*(4), 486–494.

Jenkins, J. (2014). *English as a lingua franca in the international university*. Routledge.

Jenkins, J. (2015). Repositioning English and multilingualism in English as a lingua franca. *Englishes in Practice*, *2*(3), 49–85.

Jenkins, J., Modiano, M., & Seidlhofer, B. (2001). Euro-English. *English Today*, *17*(4), 13–19.

Kachru, B. B. (1991). Liberation linguistics and the Quirk concern. In M. L. Tickoo (Ed.), *Languages and standards: Issues, attitudes, case studies* (pp. 206–226). SEAMEO Regional Language Centre.

Kirkpatrick, A. (2007). *World Englishes: Implications for international communication and English language teaching*. Cambridge University Press.

Little, D. (2007). Language learner autonomy: Some fundamental considerations revisited. *Innovation in Language Learning and Teaching*, *1*(1), 14–29.

Mahboob, A. (2009). English as an Islamic language: A case study of Pakistani English. *World Englishes*, *28*(2), 175–189.

Mauranen, A. (2018). Conceptualising ELF. In J. Jenkins, W. Baker, & M. Dewey (Eds.), *The Routledge handbook of English as a lingua franca* (pp. 7–24). Routledge.

Mayevsky, P. (2009). Interview with David Crystal on language change. *Language Issues*, *20*(2), 56–63.

Medgyes, P. (1994). *The non-native teacher*. Macmillan.

Medgyes, P. (2014). The native/non-native conundrum revisited. In J. Horváth & P. Medgyes (Eds.), *Studies in honour of Marianne Nikolov* (pp. 176–185). Lingua Franca Csoport.

Medgyes, P. (2017). *The non-native teacher. Updated edition with new material*. Swan Communication.

Mesthrie, R., & Bhatt, R. M. (2008). *World Englishes: The study of new linguistic varieties*. Cambrige University Press.

Meyerhoff, M. (2006). *Introducing sociolinguistics.* Routledge.

Modiano, M. (2009). Inclusive/exclusive? English as a lingua franca in the European Union. *World Englishes*, *28*(2), 208–223.

Murray, N. L. (1996). *Communicative language teaching and language teacher education*. [Unpublished doctoral dissertation]. Institute of Education University of London.

Pakir, A. (2009). English as a lingua franca: Analyzing research frameworks in international English, world Englishes, and ELF. *World Englishes*, *28*(2), 224–235.

Park, J. S-Y., & Wee, L. (2011). A practice-based critique of English as a lingua franca. *World Englishes*, *30*(3), 360–374.

Rubdy, R., & Saraceni, M. (2006). Introduction. In R. Rubdy & M. Saraceni (Eds.), *English in the world: Global rules, global roles* (pp. 5–16). Continuum.

Schneider, E. W. (2007). *Postcolonial English*. Cambridge University Press.

Seidlhofer, B. (2001). Closing a conceptual gap: The case for the description of English as a lingua franca. *International Journal of Applied Linguistics*, *11*(2), 133–158.

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

Seidlhofer, B. (2012). The challenge of English as a lingua franca. *Anglistik: International Journal of English Studies*, *23*(1), 73–86.

Sewell, A. (2013). English as a lingua franca: Ontology and ideology. *ELT Journal*, *67*(1), 3–10.

Siegel, A. (2014). What should we talk about? The authenticity of textbook topics. *ELT Journal*, *68*(4), 363–375.

Wardhaugh, R. (2006). *An introduction to sociolinguistics* (5th ed.). Blackwell.

Widdowson, H. G. (1983). *Learning purpose and language use*. Oxford.

Widdowson, H. G. (1990). *Aspects of language teaching*. Oxford.

Widdowson, H. G. (1996). Comment: Authenticity and autonomy. *ELT Journal*, *50*(1), 67–68.

Widdowson, H. G. (2003). *Defining issues in English language teaching*. Oxford University Press.

Widdowson, H. G. (2007). *Discourse analysis*. Oxford.

Widdowson, H. G. (2012). ELF and the inconvenience of established concepts. *Journal of English as a Lingua Franca*, *1*(1), 5–26.

Yuk-chun Lee, W. (1995). Authenticity revisited: Text authenticity and learner authenticity. *ELT Journal*, *49*(4), 323–328.

2.

Flow in Language Learning: Issues of Measurement

Ágnes Albert
albert.agnes@btk.elte.hu

Abstract

As the effect of contextual factors is increasingly acknowledged in language learning, a greater emphasis tends to be placed on the quality of the learning experience (Csizér & Kálmán, 2019). Educationalists' quest for the optimal learning experience led them to the concept of flow, which has been researched in different school contexts ever since (Csíkszentmihályi & Larson, 1984; Shernoff & Csíkszentmihályi, 2009). As flow can be defined as "the holistic sensation that people feel when they act with total involvement" (Csíkszentmihályi, 1975, p. 36), and it is an experience which typically accompanies an activity that people are willing to perform for its own sake, its educational benefits seem straightforward. Interestingly, flow in language learning has not yet been widely researched although interest in the topic seems to be growing. This review paper aims to assist researchers who see a potential in investigating flow experiences in language learning. Besides defining flow and describing the conditions that lead to it, this paper provides insights into the evolution of Csíkszentmihályi's flow theory. This theoretical overview is followed by the examination of different methods that have been used in psychology and education to measure flow, in an attempt to highlight potential new tools to be used in applied linguistics. Finally, existing empirical studies of flow in our field are discussed from a measurement perspective, and conclusions are drawn.

Keywords: flow theory, flow interview, flow questionnaire, experience sampling method (ESM), flow induction

Flow in Language Learning: Issues of Measurement

Language learning is a complex process whose success depends on many different factors. One of these is the learning context, which encompasses the daily language learning experiences of the students. Although the importance of the learning situation, which may include encounters with the language to be acquired inside or outside the classroom, has been acknowledged by a number of researchers (see e.g., Csizér & Kálmán, 2019; Dörnyei, 1994; 2005; Gardner, 1985; Noels et al., 2000; Ushioda, 2009; Williams & Burden, 1997), it seems to have attracted less empirical research than other determinants of language learning success. Since flow, which can be defined as "the holistic sensation that people feel when they act with total involvement" (Csíkszentmihályi, 1975, p. 36), describes a certain type of experience which may occur in a variety of contexts including learning contexts, it can provide a useful framework for understanding how situational influences affect the language learning process. This paper aims to offer insights into the theory of flow as a potential framework for use in our field to complement ongoing work on student engagement (e.g., Noels et al., 2019), and to provide an overview of the instruments that can be used to conduct empirical research on it to expand researchers' existing repertoire.

The Flow Experience

The concept of flow was born as a by-product of investigating creativity; a research endeavour that started in the United States in the 1950s (Guilford, 1950, 1959). Although creativity is a rather complex phenomenon and can be viewed from different angles, such as person, product, process and press (that is, the environment), for example (Rhodes, 1961), in this line of research the main focus was on the creative process, more specifically on the individual's subjective experience while being involved in it (Getzels & Csíkszentmihályi, 1976). Later, the scope of investigation was broadened in a way that it included various kinds of intrinsically motivated actions, which

came to be referred to as autotelic activities (Csíkszentmihályi, 1990). Consequently, the researcher's aim was to provide a faithful description of the subjective, phenomenological aspects of such activities and processes (Csíkszentmihályi, 1990, 1997). Eventually, the term *flow* was coined for describing "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).

Apparently, there are several key characteristics that can be used to describe the state of flow (Csíkszentmihályi, 1990); therefore, flow should be conceptualized as a multifaceted experience (Engeser & Schiepe-Tiska, 2012). The flow state can be characterized by intense focus and concentration on the activity that is being performed at the moment; people feel in total control of their actions and are confident about being able to cope with whatever comes next. A merging of action and awareness results in intense concentration on the activity at hand, leading to a great reduction or complete loss of individuals' reflective self-consciousness. Consequently, individuals engage in the actions without observing themselves. When in flow, people's subjective experience is that time passes faster than normal; thus, distortions in perceiving time also characterize this state. Finally, the objective outcomes of the actions may be irrelevant because these experiences are perceived as related to an activity that the person considers intrinsically rewarding; thus, the activity itself is the reward (Csíkszentmihályi, 1990).

In order to have the complex experience described above, the so-called flow, a number of conditions have to be met (Csíkszentmihályi et al., 2005). First, a *balance* between the perceived *skills* of the individuals and the *challenges* faced by them need to be ensured. Upsetting this balance can lead to several different consequences: if high-level skills are not met by adequate challenges, the person first relaxes, then becomes bored while if the challenges faced exceed current skill levels, the individual is likely to become anxious. Some authors (Keller & Landhäußer, 2012; Rheinberg, 2008) argue that instead of challenges the term *demands* should be used, as the term

challenges already implies the evaluation of task demands with regard to the abilities of the person, that is, the term itself already hints at the compatibility of skills and demands. Despite the plausibility of this reasoning, I will continue to use the term challenge, as it is the more common and widely used term in the literature.

A further condition that increases the likelihood of flow is the presence of a clear set of *goals* (Csíkszentmihályi et al., 2005). By providing direction and purpose to individuals' behaviour, the importance of goals lies in channelling attention and structuring the experience. Closely linked to the previous condition is the last one, which claims that individuals need to have clear and immediate *feedback* about their actions. The link between these two conditions is obvious, given that providing such feedback is much more straightforward if clear goals are set. Interestingly, the feedback does not have to be positive; the key factor here is that it serves its purpose and helps the individual in maintaining the balance between challenges and skills. Keller and Landhäußer (2012) argue that not only are clearly set goals and feedback interdependent, but since maintaining the balance between skills and challenges is impossible without clear goals and immediate feedback, these two antecedents are said to be superfluous.

A Historical Overview of Flow Theory

The very first flow model proposed by Csíkszentmihályi (1975) is based on maintaining the balance between challenges and skills. Action capabilities (*skills*) are represented on the x axis of the model while action opportunities (*challenges*) are represented on the y axis. The diagonal line between the two marks the *Flow Channel* while above the diagonal, where challenges exceed skills, worry and anxiety can be found. Below the diagonal, where skills exceed challenges, we find boredom and anxiety (p. 49). In later representations of the model, the area characterized by high skill and low challenge is indicated to lead to boredom only (Csíkszentmihályi, 1990).

However, empirical investigations (Csíkszentmihályi & LeFevre, 1989; Massimini et al., 1987), have not confirmed this initial model: it turns out a mere balance between challenges and skills, which is possible even at low levels of both, is not necessarily going to result in flow experience. The model was changed in a way that the perceived fit of challenges and skills had to be above the average level for flow to occur. Accordingly, the *quadrant model* of the flow state was born, where the diagram is divided into four squares, and the lower section of the previous diagonal line is now replaced by apathy. In the square above it, in the high challenge low skill quadrant, we can still find anxiety, while the low challenge high skill quadrant is occupied by boredom. Boredom, apathy and anxiety are sometimes referred to as anti-flow experiences (Allison & Duncan, 1987; Csíkszentmihályi, 1975).

The need to consider the findings from empirical data later led to the birth of a more complex, so-called experience fluctuation or *Octant Model* (Massimini & Carli, 1988; Csíkszentmihályi, 1997). In this model, there are eight channels in total, each with low, moderate, or high challenge and low, moderate or high level of skills. Because of the arrangement of the channels in the diagram, we find that when superimposed on the previous quadrant model, each quadrant now contains three channels with the outer ones overlapping slightly, resulting in the following eight channels: arousal, flow, control, relaxation, boredom, apathy, worry and anxiety. This more complex model takes more account of positive, pleasant experiences outside the flow channel, and it acknowledges that low challenges combined with high skills often lead to experiencing relaxation instead of boredom, for example.

As a relatively recent development, Keller and Landhäußer (2012) proposed several modifications to the flow model. In their revised flow model, instead of concentrating on challenges and skills separately, which they found unfeasible for the reasons described above, they placed the perceived fit of skills and task demands on the x axis and the subjective value of the activity on the y axis. They argued that, when collecting data, respondents are much more likely to judge the fit between the task demands

or challenges and their abilities to perform the task than to provide an independent, abstract evaluation of the two. Furthermore, introducing the subjective value of the activity into the model, not only avoids the rather ill-defined concept of *above the average flow and challenge balance*, but also allows the measurement of flow intensity.

Another interesting aspect of their theory concerns the way they determined the subjective value of the skill-related activity. Keller and Landhäußer (2012) referred to the notion of regulatory compatibility, which can be understood as the compatibility of the characteristics of the individuals and the environment, more precisely the task they are facing. Keller and Bless (2008) defined regulatory compatibility as the "phenomenological experience that arises when individuals experience a compatibility of (personal and situational) factors that are involved in performing a task or activity" (p. 197), and these personal characteristics can be related to, for example, a person's goals, needs or strategies. On the level of personality traits, Keller and Landhäußer (2012) argue that individuals with high levels of autonomy, an internal locus of control, and action orientation are more likely to experience flow, as these orientations tend to lead to situations which require the execution of skills and competences. Therefore, certain individual differences (ID) variables, some well-established within applied linguistics like autonomy (Benson, 1996), seem to be directly relevant to the capacity to experience flow.

Although it is not necessarily obvious from the static description given about the models above, the concept of flow is inherently dynamic in nature. Quite clearly the quality of the experience changes as the individual switches between different activities, but it does not stay constant even within the same activity. As people improve at a given task, a previously challenging activity will lose its appeal; thus, new challenges need to be sought to match the increased level of their abilities (Nakamura & Csíkszentmihályi, 2014). This conceptualization is very much in line with the functioning of any learning activity and recent interpretations of language

learning within the Complex Dynamic System Theory framework (De Bot, 2008).

Besides the theoretical models reviewed above, some recent models of flow are directly linked to measurement methods such as questionnaires. These are componential models, which faithfully reflect the scales of the flow measuring instruments. For example, the nine-factor model of Jackson and Marsh (1996) and Jackson and Eklund (2002) contains 1) concentration, 2) control, 3) merging of action and awareness, 4) autotelic experience, 5) lack of self-consciousness, 6) altered perception of time, 7) clear goals, 8) immediate feedback, and 9) challenge and skill balance as components of flow, while Rheinberg's (2008) flow model has only six components because clear goals and feedback have been merged, and the altered sense of time and the challenge and skill balance are missing from it. Moreover, in Bakker's (2008) interpretation of the flow concept, and thus in his questionnaire, the flow components are even fewer: absorption, work enjoyment, and intrinsic motivation.

Measuring Flow

Since the flow models reviewed above are either indirectly or directly linked to specific measurement methods, in the next section I will provide an overview of the methods used for measuring flow and give a brief description of the most well-known instruments of the field.

Interviews

The concept of flow emerged from interview studies; Csíkszentmihályi (1975) first identified the notion of optimal experience in semi-structured interviews with rock climbers, chess players, dancers and surgeons. With the help of his interviewees, he managed to acquire qualitative accounts of the phenomenological characteristics of flow, which he later used to develop other measurement instruments. Interestingly, for a

long time after his initial attempts, questionnaires rather than interviews were the dominant research methods in the study of flow (Delle Fave et al., 2011). However, there seems to be a renewed interest in the use of interviews in flow research, both because it allows for an in-depth investigation of flow experiences in certain domains (see for example Hefferon & Ollis, 2006; Pace, 2004; Swann et al., 2012), and because it can also be applied to special populations such as children (Inal & Cagiltay, 2007), or in special circumstances (Seifert & Hedderson, 2010), where the use of questionnaires is not feasible. As a new trend, interview data are sometimes supplemented by observations, as in the case of a skateboarding study, where researchers used both types of data sources in their ethnographic research (see Seifert & Hedderson, 2010).

Questionnaires

It seems that the most common method for studying flow is to use questionnaires. The questionnaire that resulted from the initial interview studies is the so-called Flow Questionnaire (FQ; Csíkszentmihályi, 1975). The FQ contains interview quotations describing different aspects of the flow experience, and respondents are asked to recognize these, thus identifying themselves as flow-ers or non-flow-ers. Flow-ers are then required to identify those activities during which they experience flow and select one that best represents the experience described in the quotations. The last section contains Likert-scale items in which respondents are asked to rate their experiences while engaged in different activities like flow, work, and leisure (Delle Fave & Massimini, 1988). This questionnaire served as a basis of the first model of flow described in the previous section, where a balance between skills and challenges was hypothesized to lead to flow experiences. A major strength of the instrument according to Moneta (2012) is that it provides a clear definition of flow and allows researchers to estimate the prevalence of flow in specific contexts. Nevertheless, the psychometric properties of the test are not in line with modern standards and expectations,

and its application is limited in that it cannot be used for measuring the intensity of flow (Moneta, 2012).

More recent questionnaires attempt to comply with the requirements of classical test theory and use scales with appropriate content and construct validity (Moneta, 2012). They tend to adopt a componential approach and attempt to capture flow as a multidimensional variable. This, however, presents certain dangers. While the FQ, where flow is conceptualized "as an all-or-nothing phenomenon–qualitatively distinct from other experiential profiles" (Delle Fave et al., 2011, p. 65), identifies individuals as flow-ers or non-flow-ers, this is clearly impossible with a questionnaire that adopts a componential approach, where the presence or absence of flow is determined by the total score. In such a questionnaire, people may be identified as being in flow state even if they score zero on one of the crucial components represented by a subscale, which contradicts flow theory. Therefore, this approach might lead to the overestimation of flow in the population.

Despite the difficulties described above, Jackson and Marsh (1996) used the componential approach to construct the Flow State Scale (FSS), which was designed to measure flow in sport and physical activity settings. Jackson and Marsh (1996) hypothesized that flow has nine distinct components, such as 1) challenge-skill balance, 2) action-awareness merging, 3) clear goal, 4) unambiguous feedback, 5) concentration on task at hand, 6) sense of control, 7) loss of self-consciousness, 8) transformation of time, and 9) autotelic experience. In their interpretation, these dimensions can be thought of as correlated facets of the flow construct, which together determine the intensity or level of flow. Two years later, Jackson and her colleagues (1998) developed a Dispositional Flow Scale (DFS) designed to determine individuals' dispositional tendency to experience flow. These questionnaires were later revised by Jackson and Ecklund (2002), resulting in the FSS-2 and the DFS-2, which are based on the same theoretical foundation but have better psychometric properties.

Although the DFS (Jackson et al., 1998) and the DFS-2 (Jackson & Ecklund, 2002) intend to assess a person's general disposition to experience flow, most recent flow questionnaires do not address the measurement of flow as a personality trait. Most of them are more like the FSS in that they are specific to the state or situation in which flow is likely to occur; therefore, they can be considered situation-specific scales. The Flow Short Scale, developed by Rheinberg, Vollmeyer and Engeser in Germany (see the English language questionnaire in Engeser, 2012), is another example for such a scale. It can be administered after any activity to measure the extent of flow experienced by participants. The Flow Short Scale measures different components of the flow experience with ten items, and it contains additional items to measure the perceived importance of the activity, the experienced difficulty of the task, respondents' perceived skill, and the perceived balance between skills.

Another situation-specific flow scale is the WOrk-reLated Flow Inventory (WOLF), developed by Bakker (2008) in the Netherlands. This questionnaire is specific to the context of work, and it attempts to measure work-related flow through 13 items. The components of flow it is supposed to cover are absorption (four items), work enjoyment (four items), and intrinsic motivation (five items). Respondents must judge the frequency of having such experiences on a 7-point Likert scale, ranging from "never" to "always". Empirical data has confirmed the three-factor structure of the inventory and, indeed, it appears to be internally consistent and valid (Delle Fave et al., 2011).

Measuring the social aspects of flow is a novel topic as far as flow questionnaires are concerned. The idea that flow can be "shared" or even enhanced by the presence of others first emerged in an interview study conducted by Allison and Duncan (1988) with blue and white-collar women at work and in leisure contexts. They found that women working in a university setting were more likely to experience flow when working with their students, and mothers in general also reported more flow when

interacting with their children. Later, in a study involving music teachers and their students, Bakker (2005) found support for his assumption that flow might be "contagious", as music teachers' and their students' scores on WOLF appeared to be related. In a study involving college students, Walker (2010) found that participants tended to enjoy flow more when the experience took place in a social context. This result was further supported by the findings of Magyaródi and Oláh's (2017) experimental study, where both task absorption and the total flow score measuring flow intensity increased in social situations.

In an attempt to relate collective efficacy beliefs to collective flow experiences, Salanova and her colleagues (2014) administered a Collective Flow Scale measuring two facets of flow: group task absorption, which is part of group engagement (6 items) (for details see Salanova et al., 2003) and a group task enjoyment scale (2 items). In addition, group challenge and group skills were also measured. They found that groups with high collective efficacy beliefs were likely to experience more flow, and also perceived more challenges and felt more competent/skilled. Thus, it seems that besides the challenge and skill balance, self-efficacy beliefs also play an important role in experiencing flow in social contexts.

Despite the apparent interest in the social aspects of flow, only one instrument specifically designed to measure it has been found. This is the Flow Synchronization Questionnaire (FSyQ), developed by Magyaródi and Oláh, (2015, 2016) in Hungary. They defined the construct of flow synchronization as the subjective experience which occurs when members of a dyad work together. During this process they cooperate and react to each other in order to attain their common goal in a situation where the challenges posed by the task and their skills are balanced, that is, the conditions of flow are fulfilled (Magyaródi & Oláh, 2015). The final version of the questionnaire contains 28 items and 5 factors that focus on engagement and concentration as well as on the motivational and cooperation aspects of the experience. The five factors of the questionnaire are 1) synchronization and effective

cooperation with the partner (12 items); 2) experience of engagement and concentration (5 items); 3) motivation and positive impact on the partner (3 items); 4) motivation and learning for the person (4 items); and 5) coordination with the partner during the activity (4 items). The internal consistencies of the scales are adequate (Magyaródi & Oláh, 2015), and the reliability and validity tests show that both convergent and discriminant validity of the scales are acceptable (Magyaródi & Oláh, 2016).

Experience Sampling Method

A famous questionnaire-based method, which was invented to capture flow in everyday life, is the Experience Sampling Method (ESM) developed by Csíkszentmihályi and his colleagues in the 1970s (Csíkszentmihályi et al., 1977). The ESM aims to capture a random sample of subjective experiences over an extended period of time. Participants of ESM studies usually receive pagers or other electronic devices that they normally need to carry for a whole week; in more recent studies, respondents may use their own cell phones. During this time, six to eight self-reports are gathered daily at random time intervals from the participants. The procedure requires participants to fill in an experience sampling form either on paper or online whenever they are signalled. The form consists of several categorical and scaled items for recording different aspects of the experience they were engaged in at the time of the signal. The activity can be reconstructed with the help of the categorical items, while the scaled items measure the intensity of a range of subjective feelings. Data gathered this way can be analysed in several different ways, using different levels of analyses (Delle Fave et al., 2011).

Over the years, the ESM has been applied to record different aspects of real-life experiences in a variety of contexts; data have been gathered about family life, work, education, and across different cultures (for an overview see Hektner et al., 2007). Results of ESM studies have led to the

formulation of the previously described quadrant and octant models of flow. The greatest advantage of ESM over other methods appears to be its ecological validity, as it provides real-time assessment of actual experiences over a series of measurements. The very same features are also a source of potential problems: for example, the attrition of participants during the long and labour-intensive process and the limitations of the contexts and situations in which this method can be applied (Delle Fave et al., 2011). As for the psychometric properties of the scales, problems with construct and content validity have been reported (Moneta, 2012) as ESM scales do not cover the whole range of components that make up flow (Rheinberg, 2008).

Flow Induction in Experimental Settings

As experimental designs usually allow for a more controlled investigation of the phenomenon in question, attempts have been made to induce flow in laboratory settings. In experimental designs, flow is usually induced through learning tasks or computer games, as these are relatively easy to manipulate in terms of difficulty levels and thus the balance between challenges and skills (see for example Chanel et al., 2008; Engeser & Rheinberg, 2008; Keller et al., 2011; Kivikangas, 2006; Nacke & Lindley, 2009; Pearce et al., 2005; Soltész et al., 2012; Ulrich et al., 2014; Walker, 2010). In studies using flow induction, researchers expect participants to experience flow at medium levels of difficulty, while low difficulty levels are supposed to induce boredom and high difficulty levels lead to anxiety. Since the balance between challenges and skills is dynamic because practice leads to enhanced skills, challenges need to be readjusted constantly and flow induction tasks should also be adjustable. Computers might provide the solution in this respect, as computer games can be adapted to increase or decrease the difficulty level of the game depending on the performance of the individual (Keller & Bless, 2008; Keller & Blomann, 2008).

Although flow induction has the great promise of providing the possibility to control and manipulate flow dimensions and to provide insight into the temporal aspect, that is, the dynamics of flow, it also has several drawbacks. The tasks given in the laboratory might not always be interesting or meaningful enough for the participants to induce flow in them, and the artificial situation limited in time certainly lacks the spontaneity and complexity of real-life situations. The benefits these solutions offer may still outweigh the problems, and to verify the validity of flow experiences, participants' subjective experiences can also be considered.

Psychophysiological Measures of Flow

Flow induction in laboratory settings enables researchers to study the psychophysiological correlates of flow. Studies have been conducted using electroencephalography (EEG), which measures brain activity (Chanel et al., 2008; Soltész et al., 2012), facial electromyography (EMG) indicating emotional valence (De Manzano et al., 2010; Kivikangas, 2006; Nacke & Lindley, 2009), measures of electrodermal activity (EDA) reflecting arousal (Kivikangas, 2006; Nacke & Lindley, 2009), and electrocardiography (ECG) displaying effects of the autonomous nervous system (De Manzano et al., 2010; Keller et al., 2011), for example. Results of these studies are somewhat contradictory, which is not surprising given that flow can be experienced in a wide range of activities, from solving math problems to playing computer games or the piano. Moreover, most of the above psychophysiological measures are general and are probably influenced by many other factors besides the flow state experienced by the participant. Nevertheless, in an overview of the psychophysiological correlates of flow, Peifer (2012) concluded that the definition of flow should include that it is a state "that is characterized by optimized physiological activation" (p. 160), referring to "a decreased activation in default networks" (p. 160) on the one hand, and "moderate peripheral arousal following a u-shaped function of activation" (p. 160) on the other. According to her, the former characteristic describes the

state when task-irrelevant cortical processes (like self-reflection) are disengaged, while the latter suggests that the flow state becomes possible when arousal is optimal, that is, neither too low not too high.

In addition to the general measures described above, neuroimaging techniques provide a more detailed and accurate picture of the activation of different brain areas. In a study conducted by Ulrich et al. (2014) Magnetic Resonance Imaging (MRI) was used to measure regional cerebral blood flow in the brain. The authors found that the flow condition was associated with increased neural activity in the putamen and in the left inferior frontal gyrus. The first area was hypothesized to reflect the increased likelihood of achievement in the context of a difficult task, thus possibly linked to intrinsic rewards, whereas the latter suggests a deeper sense of cognitive control. Besides increases, reductions in neural activity were also observed. One of the affected areas was the medial prefrontal cortex, which is believed to be involved in self-referential processing associated with negative affectivity. The other area showing decreased activation was the amygdala, which might contribute to or reflect positive emotional experiences during flow. These findings suggest that certain phenomenological attributes of flow such as goal-orientation, sense of control, loss of reflective self-consciousness, and positive valence might be directly linked to changes in neural activity in specific areas of the brain.

Measuring Flow in Applied Linguistics

Although the growing number of empirical studies investigating flow in language learning and the recent publication of a chapter on its relationship with language learning motivation (Piniel & Albert, 2019) seem to indicate a growing interest in flow in our field, it seems that there is still much room for improvement in comparison with other areas of education. The first studies to investigate the contribution of flow to language learning motivation in Asian and Middle East contexts were conducted by Schmidt

and Savage (1994) and Schmidt and his colleagues (1996). While the first study adopted the classic Experience Sampling Method (ESM) which has been used in many other studies across different contexts (Delle Fave et al., 2011; Hektner et al., 2007), the second utilized questionnaires, the measurement tool that later became the dominant data collection instrument in flow studies within our field.

Interestingly, it was Egbert's study (2003), published years later, that seems to have inspired subsequent studies on flow in applied linguistics. With the intention of prompting further studies on optimal learning experiences in language classrooms, Egbert set out to test the applicability of flow theory to language learning tasks. Her study focused on secondary school learners of Spanish completing altogether seven language learning related tasks in the classroom, two of which involved reading activities on a computer. Her findings seemed to indicate that students experienced flow in the language classroom, especially in tasks where the challenge of the task and the learners' abilities were in balance, the task was interesting, and learners had some control over the task itself. Egbert used both questionnaires and stimulated recall post-task interviews to investigate the flow experiences of her 13 participants. Her Perceptions Survey, which was a modified version of Webster and his colleagues' (1993) and Trevino and Webster's (1992) instruments, contained 14 items belonging to four flow dimensions, such as interest, control, focus, and challenge. The fact that this instrument was later adopted by several researchers makes her study truly ground-breaking.

Aubrey (2017a; 2017b) was one of the authors who drew on the questionnaire published by Egbert (2003) to investigate how inter-cultural contact influenced first-year Japanese EFL learners' experiences of flow during five different dialogue tasks. Besides the questionnaire, he also used content analysis to retrospectively investigate the flow-enhancing and flow-inhibiting experiences during task performance recorded in his learners' diaries (Aubrey, 2017b). In a similar vein, Cho (2018) also used a modified

version of Egbert's (2003) questionnaire in her larger-scale study involving 141 participants, where she attempted to demonstrate the effects of task complexity and task modality on the perception of different flow components.

Although in some studies researchers have used different questionnaires to measure flow, they have retained Egbert's (2003) approach in the sense that they have used situation-specific instruments designed to measure the specific flow experiences that participants may have had in a given context. In their study exploring the relationships between intrinsic motivation, online self-efficacy, and flow experiences, Hong et al. (2017) used a total of five items to shed light on their learners' flow experiences while playing a game involving different Chinese characters. In contrast, Li et al. (2019) used a multi-dimensional questionnaire that, in addition to measuring antecedents of flow such as challenge-skill balance, clear goal, feedback and playability, also included outcome variables such as perceived learning and satisfaction, alongside the flow components of concentration, intrinsic motivation, and enjoyment. Similarly, Liu and Song (2021) drawing on Shin's (2006) conceptual framework also differentiated flow antecedents (skill, challenge, and clear goal), flow experience components (enjoyment, telepresence, focused attention, peer interaction, and time distortion), and flow consequences (achievement, satisfaction, and change of attitude and behaviour) while analysing learners' experiences during an English dubbing task.

Czimmermann and Piniel's (2016) study is also notable for two reasons. The first novelty of their approach is that, although they also used Egbert's (2003) Perception Survey as a starting point for their study, they supplemented it with measures of anti-flow experiences such as boredom, apathy, and anxiety (Spielberger, 1983). Secondly, in addition to measuring situation-specific flow, they also set out to measure classroom flow based on Oláh (2005), which instead of concentrating on the effects of a particular task, attempted to establish the likelihood of students' experiencing flow in

language classes in general; therefore, it appears to represent a transition between situation-specific flow and a dispositional flow scale. The same instrument measuring classroom flow was also used in Piniel and Albert's (2017) study which investigated high school students' flow and anti-flow experiences in relation to their motivated language learning behaviour and self-efficacy beliefs about learning English as a foreign language, the results of which suggested a circular relationship between these constructs.

Although questionnaires seem to dominate the field of flow research in applied linguistics, there are a number of studies that have used interviews. Abbott (2000) conducted a longitudinal interview study on children's flow experiences of writing, while Tardy and Snyder (2004) investigated the role of teachers' flow experiences in their understanding of their teaching practices, beliefs, and values. Recently Ibrahim and Al-Hoorie (2019) coined the term shared sustained flow (SSF) which occurs "when groupwork is coupled with flow over a period of time, potentially making learning both effective and highly enjoyable" (p. 52), and their findings were based on interviews they conducted with both language teachers and learners. Finally, Liu and Song (2021) also used follow-up interviews to collect further information about their participants' perceptions of the three different flow stages.

As the brevity of this overview indicates, the study of flow in language learning is still a developing field within applied linguistics and offers many opportunities for aspiring researchers. In terms of possible topics to be explored, at a theoretical level, the relationship between the increasingly popular notion of engagement (Noels et al., 2019) and flow should be clarified. Moreover, empirical studies could be carried out to provide further evidence for the hypothesized motivating effects of flow (Piniel & Albert, 2019). As flow appears to be very much a state-like concept (Csizér & Albert, 2021), it is probably best suited to be explored in context, with the help of different language tasks. Moreover, the dynamic nature of the construct (Nakamura & Csíkszentmihályi, 2014) makes it ideal for

inclusion in various Complex Dynamic Systems Theory studies of language learning, especially as the experience sampling method is explicitly mentioned as a method suitable for conducting such studies (Hiver & Al-Hoorie, 2019). This final remark also draws attention to the need to expand the methodological repertoire used to study flow within our field, to include new questionnaire-based tools such as the experiential sampling method, in addition to interviews and psychophysiological measures.

Conclusion

By providing an overview of the evolution of flow theory, one of the aims of the present paper was to demonstrate its relevance to learning in general and language learning in particular. Today, when motivation theories place an increasing amount of emphasis on the learning experience (see for example Csizér & Kálmán, 2019; Dörnyei, 2005, 2019), flow experiences which ensure high levels of engagement cannot be disregarded. Another aim of the paper was to review the numerous instruments that can be used to capture flow. Since it is clear that researchers in our field have a rather limited range of instruments to measure this construct, it might be useful to open a broader horizon and to present a range of options used by psychologists and educationalists. Since I am convinced that flow theory is a useful framework to investigate language learning experiences, I hope that having presented a wider repertoire of measurement tools available, researchers might be more inclined to conduct empirical studies in this area.

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

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

Allison, M. T., & Duncan, M. C. (1987). Women, work, and leisure: The days of our lives. *Leisure Sciences*, *9*(3), 143–161. https://doi.org/10.1080/01490408709512156

Allison, M. T., & Duncan, M. C. (1988). Women, work, and flow. In M. Csikszentmihalyi & I. S. Csikszentmihalyi (Eds.), *Optimal experience: Psychological studies of flow in consciousness* (pp. 118–137). Cambridge University Press.

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

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

Bakker, A. B. (2008). The work-related flow inventory: Construction and initial validation of the WOLF. *Journal of Vocational Behavior*, *72*(3), 400–414. https://doi.org/10.1016/j.jvb.2007.11.007

Benson, P. (1996). Concept of autonomy in language learning. In R. Pemberton, E. S. L. Li, W. W. F. Or, & H. D. Pierson (Eds.), *Taking control: Autonomy in language learning* (pp. 27–34). Hong Kong University Press.

Chanel, G., Rebetez, C., Bétrancourt, M., & Pun, T. (2008). Boredom, engagement and anxiety as indicators for adaptation to difficulty in games. In A. Lugmayr, F. Mäyrä, H. Franssila, & K. Lietsala (Eds.), *Proceedings of the 12th International MindTrek Conference: Entertainment and media in the ubiquitous era* (pp. 13–17). Association for Computing Machinery, Inc. https://doi.org/10.1145/1457199.1457203

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. (1975). *Beyond boredom and anxiety: Experiencing flow in work and play*. Jossey-Bass Publishers.

Csíkszentmihályi, M. (1990). *Flow: The psychology optimal experience*. Harper Perennial.

Csíkszentmihályi, M. (1997). *Finding flow: The psychology of engagement with everyday life*. Basic Books.

Csíkszentmihályi, M., & Larson, R. (1984). *Being adolescent: Conflict and growth in the teenage years.* Basic Books.

Csíkszentmihályi, M., & LeFevre, J. (1989). Optimal experience in work and leisure. *Journal of personality and social psychology*, *56*(5), 815–822. https://doi.org/10.1037/0022-3514.56.5.815

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.

Csíkszentmihályi, M., Larson, R., & Prescott, S. (1977). The ecology of adolescent activity and experience. *Journal of Youth and Adolescence*, *6*(3), 281–294. https://doi.org/10.1007/BF02138940

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

Csizér, K., & Kálmán, Cs. (2019). A study of retrospective and concurrent foreign language learning experiences: A comparative interview study in Hungary. *Studies in Second Language Learning and Teaching*, *9*(1), 225–246. https://doi.org/10.14746/ssllt.2019.9.1.10

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.

De Bot, K. (2008). Introduction: Second language development as a dynamic process. *The Modern Language Journal*, *92*(2), 166–178. https://doi.org/10.1111/j.1540-4781.2008.00712.x

De Manzano, Ö., Theorell, T., Harmat, L., & Ullén, F. (2010). The psychophysiology of flow during piano playing. *Emotion*, *10*(3), 301. https://doi.org/10.1037/a0018432

Delle Fave, A., & Massimini, F. (1988). Modernization and the changing contexts of flow in work and leisure. In M. Csíkszentmihályi & I. Csíkszentmihályi (Eds.), *Optimal experience: Psychological studies of flow in consciousness* (pp. 193–213). Cambridge University Press.

Delle Fave, A., Massimini, F., & Bassi, M. (2011). *Psychological selection and optimal experience across cultures: Social empowerment through personal growth* (Vol. 2). Springer Science & Business Media.

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

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

Dörnyei, Z. (2019). Towards a better understanding of the L2 Learning Experience, the Cinderella of the L2 Motivational Self System. *Studies in Second Language Learning and Teaching*, *9*(1), 19–30. https://doi.org/10.14746/ssllt.2019.9.1.2

Egbert, J. (2003). A study of flow theory in the foreign language classroom. *The Modern Language Journal*, *87*, 499–518. https://doi.org/10.1111/1540-4781.00204

Engeser, S. (2012). Theoretical integration and future lines of flow research. In S. Engeser (Ed.), *Advances in flow research* (pp. 187–199). Springer.

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

Engeser, S., & Schiepe-Tiska, A. (2012). Historical lines and an overview of current research on flow. In S. Engeser (Ed.), *Advances in flow research* (pp. 1–22). Springer.

Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation.* Edward Arnold.

Getzels, J. W., & Csikszentmihalyi, M. (1976). *The creative vision: A longitudinal study of problem finding in art*. Wiley Interscience.

Guilford, J. P. (1950). Creativity. *American Psychologist*, *5*, 444–454. https://doi.org/10.1037/h0063487

Guilford, J. P. (1959). Three faces of intellect. *American Psychologist*, *14*, 469–479. https://doi.org/10.1037/h0046827

Hefferon, K. M., & Ollis, S. (2006). 'Just clicks': An interpretive phenomenological analysis of professional dancers' experience of flow. *Research in Dance Education*, *7*(2), 141–159. https://doi.org/10.1080/14647890601029527

Hektner, J. M., Schmidt, J. A., & Csíkszentmihályi, M. (2007). *Experience sampling method: Measuring the quality of everyday life.* Sage.

Hiver, P., & Al-Hoorie, A. H. (2019). *Research methods for complexity theory in applied linguistics*. Multilingual Matters.

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

Inal, Y., & Cagiltay, K. (2007). Flow experiences of children in an interactive social game environment. *British Journal of Educational Technology*, *38*(3), 455–464. https://doi.org/10.1111/j.1467-8535.2007.00709.x

Jackson, S. A., & Eklund, R. C. (2002). Assessing flow in physical activity: The flow state scale–2 and dispositional flow scale–2. *Journal of Sport and Exercise Psychology*, *24*(2), 133–150. https://doi.org/10.1123/jsep.24.2.133

Jackson, S. A., & Marsh, H. W. (1996). Development and validation of a scale to measure optimal experience: The Flow State Scale. *Journal of Sport and Exercise Psychology*, *18*(1), 17–35. https://doi.org/10.1123/jsep.18.1.17

Jackson, S. A., Ford, S. K., Kimiecik, J. C., & Marsh, H. W. (1998). Psychological correlates of flow in sport. *Journal of Sport and Exercise Psychology*, *20*(4), 358–378. https://doi.org/10.1123/jsep.20.4.358

Keller, J., & Bless, H. (2008). Flow and regulatory compatibility: An experimental approach to the flow model of intrinsic motivation. *Personality and Social Psychology Bulletin*, *34*(2), 196–209. https://doi.org/10.1177/0146167207310026

Keller, J., & Blomann, F. (2008). Locus of control and the flow experience: An experimental analysis. *European Journal of Personality*, *22*(7), 589–607. https://doi.org/10.1002/per.692

Keller, J., & Landhäußer, A. (2012). The flow model revisited. In S. Engeser (Ed.), *Advances in flow research* (pp. 51–64). Springer.

Keller, J., Bless, H., Blomann, F., & Kleinbohl, D. (2011). Physiological aspects of flow experiences: Skills-demand-compatibility effects on heart rate variability and salivary cortisol. *Journal of Experimental Social Psychology*, *47*, 849–852. https://doi.org/10.1016/j.jesp.2011.02.004

Kivikangas, J. M. (2006). *Psychophysiology of flow experience: An explorative study*. [Master's thesis, University of Helsinki.] https://helda.helsinki.fi/bitstream/handle/10138/19953/psychoph.pdf?sequence=2

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*, 1–22. https://doi.org/10.1080/09588221.2019.1619585

Liu, H., & Song, X. (2021). Exploring "Flow" in young Chinese EFL learners' online English learning activities. *System*, *96*, 102425. https://doi.org/10.1016/j.system.2020.102425

Magyaródi, T., & Oláh, A. (2015). Flow Szinkronizáció Kérdőív: az optimális élmény mechanizmusának mérése társas interakciós helyzetekben [Flow Synchronization Questionnaire: Measuring optimal experiences in social interactive situations]. *Mentálhigiéné és Pszichoszomatika*, *16*(3), 271–296. https://doi.org/10.1556/0406.16.2015.3.4

Magyaródi, T., & Oláh, A. (2016). A Flow Szinkronizáció Kérdőív pszichometriai jellemzői: reliabilitás-és validitásvizsgálat [Psychometric properties of the Flow Synchronization Questionnaire: Studying its validity and reliability]. *Mentálhigiéné és Pszichoszomatika*, *17*(2), 145–166. https://doi.org/10.1556/0406.17.2016.2.4

Magyaródi, T., & Oláh, A. (2017). The effect of social interaction on flow experience. *International Journal of Psychology and Behavior Analysis*, *3*, 126. https://doi.org/10.15344/2455-3867/2017/126

Massimini, F., & Carli, M. (1988). The systematic assessment of flow in daily experience. In M. Csíkszentmihályi & I. Csíkszentmihályi (Eds.), *Optimal experience: Psychological studies of flow in consciousness* (pp. 266–287). Cambridge University Press.

Massimini, F., Csíkszentmihályi, M., & Carli, M. (1987). The monitoring of optimal experience: A tool for psychiatric rehabilitation. *The Journal of Nervous and Mental Disease*, *175*(9), 545–549. https://doi.org/10.1097/00005053-198709000-00006

Moneta, G. B. (2012). On the measurement and conceptualization of flow. In S. Engeser (Ed.), *Advances in flow research* (pp. 23–50). Springer.

Nacke, L., & Lindley, C. (2009). Affective ludology, flow and immersion in a first-person shooter: Measurement of player experience. *Loading...: The Journal of the Canadian Game Studies Association*, *3*(5). Online resource available under http://journals.sfu.ca/loading/index.php/loading/article/view/72

Nakamura, J., & Csíkszentmihályi, M. (2014). The concept of flow. In M. Csíkszentmihályi (Ed.), *Flow and the foundations of positive psychology* (pp. 239–264). Springer. https://doi.org/10.1007/978-94-017-9088-8_16

Noels, K. A., Lou, N. M., Vargas Lascano, D. I., Chaffe, K. E., Dincer, A., Zhang, Y. S. D., & Zhang, X. (2019). Self-determination and motivated engagement in language learning. In M. Lamb, K. Csizér, A. Henry, & S. Ryan (Eds.), *The Palgrave handbook of motivation for language learning* (pp. 95–116). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-28380-3_5

Noels, K. A., Clément, R., & Pelletier, L. G. (1999). Perceptions of teacher communicative style and students' intrinsic and extrinsic motivation. *Modern Language Journal*, *83*, 23–34.

Noels, K. A., Pelletier, L. G., Clément, R., & Vallerand, R. J. (2000). Why are you learning a second language? Motivation orientations and self-determination theory. *Language Learning*, *50*, 57–85. https://doi.org/10.1111/0023-8333.00111

Oláh, A. (2005). Érzelmek, megküzdés, tökéletes élmény [Emotions, coping, perfect experience]. Trefort.

Pace, S. (2004). A grounded theory of the flow experiences of Web users. *International Journal of Human-Computer Studies*, *60*(3), 327–363. https://doi.org/10.1016/j.ijhcs.2003.08.005

Pearce, J. M., Ainley, M., & Howard, S. (2005). The ebb and flow of online learning. *Computers in Human Behavior*, *21*(5), 745–771. https://doi.org/10.1016/j.chb.2004.02.019

Peifer, C. (2012). Psychophysiological correlates of flow-experience. In S. Engeser (Ed.), *Advances in flow research* (pp. 139–164). Springer.

Piniel, K., & Albert, Á. (2017). L2 motivation and self-efficacy's link to language learners' flow and anti-flow experiences in the classroom. In S. Letica Krevelj & R. Geld (Eds.), *UZRT 2016 Empirical studies in applied linguistics* (pp. 90–103). FF 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.

Rheinberg, F. (2008). Intrinsic motivation and flow-experience. In H. Heckhausen & J. Heckhausen (Eds.), *Motivation and action* (pp. 323–348). Cambridge University Press.

Rhodes, M. (1961). An analysis of creativity. *The Phi Delta Kappan*, *42*(7), 305–310. https://www.jstor.org/stable/20342603

Salanova, M., Llorens, S., Cifre, E., Martínez, I. M., & Schaufeli, W. B. (2003). Perceived collective efficacy, subjective well-being and task performance among electronic work groups: An experimental study. *Small Groups Research*, *34*, 43–73. https://doi.org/10.1177/1046496402239577

Salanova, M., Rodríguez-Sánchez, A. M., Schaufeli, W. B., & Cifre, E. (2014). Flowing together: A longitudinal study of collective efficacy and collective flow among workgroups. *The Journal of Psychology*, *148*(4), 435–455. https://doi.org/10.1080/00223980.2013.806290

Schmidt, R., & Savage, W. (1994). Challenge, skill, and motivation. *University of Hawai'i Working Papers in English as a Second Language*, *12*(2), 1–25.

Schmidt, R., Boraie, D., & Kassabgy, O. (1996). Foreign language motivation: Internal structure and external connections. *University of Hawai'i Working Papers in English as a Second Language*, *14*(2) 1–70.

Seifert, T., & Hedderson, C. (2010). Intrinsic motivation and flow in skateboarding: An ethnographic study. *Journal of Happiness Studies*, *11*(3), 277–292. https://doi.org/10.1007/s10902-009-9140-y

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.

Shin, N. (2006). Online learner's 'flow' experience: An empirical study. *British Journal of Educational Technology*, *37*(5), 705–720. https://doi.org/10.1111/j.1467-8535.2006.00641.x

Soltész, P., Magyaródi, T., Mózes, T., Nagy, H., & Oláh, A. (2012). A flow-élmény elekrtofiziológiája [The electrophysiology of flow]. *Magyar Pszichológiai Szemle*, *67*(1), 77–103.

Spielberger, C. D. (1983). *State-Trait Anxiety Inventory for Adults (STAI-AD)* [Database record]. APA PsycTests. https://doi.org/10.1037/t06496-000

Swann, C., Keegan, R., Piggott, D., Crust, L., & Smith, M. F. (2012). Exploring flow occurrence in elite golf. *Athletic Insight*, *4*(2), 171.

Tardy, C. M., & Snyder, B. (2004). 'That's why I do it': Flow and EFL teachers' practices. *ELT Journal*, *58*(2), 118–128. https://doi.org/10.1093/elt/58.2.118

Trevino, L. K., & Webster, J. (1992). Flow in computer-mediated communication: Electronic mail and voice mail evaluation and impacts. *Communication Research*, *19*(5), 539–573. https://doi.org/10.1177/009365092019005001

Ulrich, M., Keller, J., Hoenig, K., Waller, C., & Grön, G. (2014). Neural correlates of experimentally induced flow experiences. *Neuroimage*, *86*, 194–202. https://doi.org/10.1016/j.neuroimage.2013.08.019

Ushioda, E. (2009). A person-in-context relational view of emergent motivation, self and identity. In Z. Dörnyei, & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 215–228). Multilingual Matters.

Walker, C. J. (2010). Experiencing flow: Is doing it together better than doing it alone? *The Journal of Positive Psychology*, *5*, 3–11. https://doi.org/10.1080/17439760903271116

Webster, J., Trevino, L. K., & Ryan, L. (1993). The dimensionality and correlates of flow in human-computer interactions. *Computers in Human Behavior*, *9*(4), 411–426. https://doi.org/10.1016/0747-5632(93)90032-N

Williams, M., & Burden, R. L. (1997). *Psychology for language teachers*. Cambridge University Press.

3.

Probing the Advanced Level EFL School-leaving Examination: The Use of English Paper

Gyula Tankó
tanko.gyula@btk.elte.hu

Zsuzsanna Andréka
andreka.zsuzsanna@btk.elte.hu

Abstract

The two-level school-leaving examination system was introduced in Hungary in 2005. Since then 16 years have passed, and the time is ripe to review the foreign language school-leaving examination, which was modified in 2017 and which is the most important examination in terms of the English major tertiary education programmes in Hungary as a gatekeeping test. The reason why it is necessary to probe the examination is that experience shows that the language proficiency of the students admitted into the English major education programmes, in which English is the medium of education, is low and insufficient for coping with the challenges posed by the courses. As a consequence, this exploratory investigation set out to analyse the Use of English (UoE) paper of the B2 level advanced EFL school-leaving examination, namely the paper which assesses the constructs constituting the core of communicative language ability without which there is no reading, listening, speaking, or writing ability to discuss, that is, lexis, grammar, and lexico-grammar. It aimed to answer the following research question: What are the potential shortcomings of the Use of English (UoE) paper that may partly explain why the B2 level advanced EFL school-leaving examination does not fulfil its intended gatekeeping function?

Altogether 32 tasks from eight UoE papers from the 2017-2020 period were analysed. Through the use of the official nationwide descriptive statistical data available on UoE tasks, the solution and close critical analysis of the UoE items, and double-coding based on a codebook developed for this investigation, it was revealed that the UoE paper suffers from serious flaws which threaten its construct validity, impact the reliability, fairness, and predictive validity of the test, and which may partly explain the low language proficiency level of the students admitted into English major programmes.

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

Probing the Advanced Level EFL School-leaving Examination: The Use of English Paper

The two-level school-leaving examination system was introduced in Hungary in 2005. Originally, the foreign language school-leaving examination consisted of an intermediate A2/B1 level examination, as defined by to the Common European Framework of Reference for Languages (Council of Europe, 2020), and an advanced B2 level examination. The system was last modified in 2017, when, amongst other modifications, the level of the intermediate foreign language school-leaving examination was changed from A2/B1, which was a Hungaricum in the field of European language assessment, to B1, while the level of the B2 exam was kept the same.

The foreign language school-leaving examination is compulsory for all secondary school leavers in Hungary (Hungarian Government, 2021), and English is one of the foreign languages that students can and do select for the examination. In 2020, altogether 33,427 secondary school leavers took the B1 level and 14,264 the B2 level examination in English compared to 9,996 B1 level and 2,449 B2 level examinations in German (Oktatási Hivatal, 2020).

An EFL secondary school leaving certificate is necessary for admission to English major tertiary education programmes in Hungary, namely the English and American Studies (BA) or the Teacher of English as a Foreign Language (MA) programmes. According to the Felsőoktatási felvételi tájékoztató (Higher Education Admission Bulletin, 2021), students applying for an English and American Studies (BA) programme must have a B2 level EFL secondary school leaving certificate, while those applying for a Teacher of English as a Foreign Language programme must have a B2 level certificate only if they select a non-foreign language second major or a foreign language major for teachers of minor languages (e.g., Japanese, Latin, or Slovenian). If they wish to major as teachers of two major languages (i.e., English combined with French, Dutch, German, Italian, or Spanish), they are eligible with one B2 level school leaving certificate in either of the two languages.

The majority of the students have been admitted lately to the Teacher of English as a Foreign Language (MA) programmes. For example, in the autumn semesters of 2017, 2018, and 2019 altogether 288, 318, and 301 ($\sum = 907$) students were admitted to the 10- and 11-semester MA in English teacher training programme in contrast to the 152, 170, and 250 ($\sum = 572$) students admitted to the BA in English programme of Eötvös Loránd University (Oktatási Hivatal, 2021b). This means that some of the students who were admitted did not have a B2 level certificate, which is a problem in itself, given that the minimum level recommended, for example, for being able to summarise texts in English, a key study skill necessary for learning content at tertiary level, is B2 (Kirkland & Saunders, 1991). This in itself could partly explain the low language ability of first year English major students that has been causing problems, namely that for B1 certificate holders proficiency in English is not available as a tool for learning but is a learning goal in programmes where English is the medium of instruction. However, the more likely and serious source of the problem may be that the B2 level EFL school-leaving examination itself does not assess language ability as intended.

Given the implications of the decisions made on the basis of these high-stakes examination results, primarily for the students and teachers in the English-medium tertiary-level instruction programmes, this investigation was carried out in order to analyse the UoE paper of the B2 level EFL school-leaving examination, that is, the paper which assesses the constructs that constitute the core of communicative language ability: lexis, grammar, and lexico-grammar.

Literature Review

Use of English in Language Assessment

A Use of English (UoE) paper assesses the knowledge and ability to use language structures (grammar) and words (lexis) in context (Alderson & Csereszny és, 2003a; Docherty, 2015). Early in language assessment, both grammar and lexis used to be assessed separately (Weir, 2013). However, Halliday (2004) argued that they should be considered the "two poles of a single cline, or continuum" (p. 43). Language assessors also found that the two constructs were inseparable (cf. Read, 2000 for a review of relevant studies) and should be assessed embedded in context in actual language use tasks that engage the test taker's abilities to combine form and meaning in pragmatically appropriate ways (Purpura, 2004; Read, 2000). Furthermore, the separate assessment of structures and words came to be augmented due to corpus linguistics findings (Pearce, 2007; Sinclair, 1991) with an integrated lexico-grammatical approach, which recognises the interdependence of grammar and lexis.

Halliday (2004) also claimed that lexico-grammar lies in the middle of the grammar-lexis cline. Together with other linguists, he preferred the term lexical grammar (Halliday, 2004; Sinclair, 2004), considering lexis a "delicate grammar" (Halliday, 2004, p. 44). Their rationale is formulated by Sinclair (2004, p. 164) as follows:

> Lexico-grammar is now very fashionable, but it does not integrate the two types of pattern as its name might suggest–it is fundamentally grammar with a certain amount of attention to lexical patterns within the grammatical frameworks; it is not in any sense an attempt to build together a grammar and lexis on an equal basis.

The same rationale was articulated by Purpura (2004, p. 17) from a language assessment perspective in the following way:

> In my opinion, the literal meaning of an utterance expressed by lexico-grammatical forms is inextricably associated with the intention that a speaker has in generating a proposition in context, and this association of form and meaning is the essence of grammatical ability.

Language assessors added lexico-grammar to grammar and lexis, which, as Gledhill (2012) noted, was especially required for the realisation of the expressive and communicative functions of language. Language assessors started to assess grammar, lexis, and lexico-grammar in a rich context where not the test takers' knowledge about form, meaning, and pragmatic appropriacy were tested but their ability to use language meaningfully in context (Alderson & Cseresznyés, 2003a). Consequently, the direct assessment of language ability through performance elicitation replaced indirect testing.

Context in Use of English Tasks

Read (2000) argued that since the same lexical item can have a different meaning in different contexts, in language tests lexis is to be presented in a context where test takers can rely on contextual clues. As already indicated in the above quotation on lexico-grammar, Purpura (2004) also stressed the language assessment implications of the relationship between form, meaning, and pragmatic use, thereby reinforcing the importance of context-appropriate language use.

The context-dependent assessment of UoE constructs resulted in tasks with items embedded in reading passages. This entailed the engagement of not only grammar, vocabulary, and lexico-grammar but also of reading comprehension, a development that was not unexpected. Hale et al. (1988) reported in a study on the rational cloze task type that vocabulary and grammar knowledge and reading comprehension were highly interrelated. Alderson (2000) also presented empirical evidence demonstrating that tests of reading were closely related to tests of vocabulary and grammar.

Furthermore, Spolsky (1995, p. 164) discussed an unpublished paper by J. B. Carroll and summarised his conclusion formulated already in 1954 concerning the relationship between vocabulary and reading measurement in foreign language: "If the stimulus were one word long, one was dealing with a vocabulary test; if longer, it was a reading comprehension test".

As a consequence, for example in the Cambridge Assessment English tests which assess UoE, since 2013, UoE has been measured in combination with Reading ability—with separate UoE and Reading tasks but within one paper (Zeronis & Elliot, 2013). One of the reasons given for this revision was that test takers are expected to and were actually found to process (Docherty, 2015) UoE and Reading tasks the same way. With the exception of the key word transformation items, the items in both components are presented in context, embedded in reading passages. Consequently, both rely on reading comprehension. Test takers are advised to start by reading the texts through in order to understand them. In both cases they engage in local (i.e., within sentence) and global (i.e., across sentences) reading behaviours (Khalifa & Weir, 2009). This is rather different from the 1970s, when the items in decontextualized, discrete-point test tasks like the cloze test targeted only local grammar and lexis (Read, 2000).

In order to reduce the effect of reading comprehension, the reading passages in the UoE tasks are to be less complex than in the Reading paper (Zeronis & Elliot, 2013). Alderson and Cseresznyés (2003a) also elaborated on this, stating that a UoE

> item or task will not focus on reading skills or strategies, and will therefore normally put the task or item in a context that is relatively easy to understand, that 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. A Use of English test, in other words, will require students to read and understand the context, but that context should be *considerably easier* [emphasis added] than the texts in a reading test, because the text is not the

> focus–the item (structure, word formation, vocabulary item) is the focus. (p. 26)

As a consequence, the reduction of the contamination of a UoE paper with reading comprehension can be achieved partly if it is ensured that the level of the lexis, grammar, and lexico-grammar targeted by the items matches the proficiency level the test is aimed to assess, whereas the level of the context necessary to answer an item is at or preferably lower than the measured proficiency level. There is an increasing amount of empirical evidence available on the actual language ability of learners at the various CEFR levels (e.g., English Vocabulary and Grammar Profile, 2015) and language test task developers should capitalise on these.

Additional Factors Affecting the Assessment

In addition to reading, there are other factors that can affect the performance of test takers attempting UoE tasks. For example, the illustrations accompanying a test task must be clear, age- and content-appropriate, and functional both in terms of thematic relevance and reproducibility through test booklet printing (Alderson & Csereszny és, 2003b; Tankó, 2005). Empirical evidence on their impact is available, for example, from Lindner et al. (2017), who investigated the effect in educational assessment of adding such representational pictures (RPs) to text-based items that effectively visualise the text without providing solution-relevant information. They found that RPs positively affected performance, test score reliability, and the validity of score interpretation and reduced rapid-guessing behaviour. The authors also argued that RPs help reading comprehension and the avoidance of misinterpretations because they foster the construction of coherent mental models of text representation. Similarly to illustrations, the titles of input texts have been researched, and there is evidence (cf. Lemarié et al., 2012 for a review) that they affect text content interpretation through prior knowledge activation.

The level of the instructions provided in the target language must not be higher than the assessed level, and the response format must not favour or disfavour test takers (Bachman & Palmer, 2010). These are to be ensured through task review. Furthermore, operational test tasks must also be reviewed (Fulcher, 2010) to make sure that their content matches the specification; the key is correct; the task is not biased in favour or against a subgroup; and there are no spelling errors, incorrect grammar, or formatting problems that affect processing. Additionally, the input text must be well-formed for efficient processing, and punctuation must also be correct because, as Tankó (2012) argued, it helps text processing and comprehension through its disambiguating function.

Construct Validity

The factors discussed above are likely to affect the validity of the measured construct, definable as "an integration of any evidence that bears on the interpretation or meaning of test scores" by Messick (1989, p. 17). Messick (1995) differentiated between two major types of threats to construct validity: construct underrepresentation, meaning that "the assessment is too narrow and fails to include important dimensions or facets of the construct" (p. 742); and construct-irrelevant variance, meaning that "the assessment is too broad, containing excess reliable variance associated with other distinct constructs as well as method variance, such as response sets or guessing propensities, that affects responses in a manner irrelevant to the interpreted construct" (p. 742). The two threats can occur simultaneously.

Construct-irrelevant variance always affects the consequential aspects of construct validity adversely and has two subtypes: construct-irrelevant difficulty, meaning that the "aspects of the task that are extraneous to the focal construct make the task irrelevantly difficult for some individuals or groups" (p. 742), which Messick exemplified with the contamination of a not reading-focused test with "undue reading comprehension" requirements that

negatively affect the scores of weaker readers; and construct-irrelevant easiness, meaning that "extraneous clues in item or task formats permit some individuals to respond correctly or appropriately in ways irrelevant to the construct being assessed" (p. 743), which Messick exemplified with familiarity with the contents of a reading passage that results in high but invalid scores for the advantaged test takers. Messick also stressed that construct-irrelevant variance is especially pertinent 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). In keeping with the above-discussed characteristics of UoE tasks and the associated assessment considerations, the current exploratory investigation aimed to answer the following research question: How suitable is the Use of English paper in the advanced level EFL school-leaving examination for the level-appropriate assessment of the aspects of language ability that it intends to assess?

Methods

In order to answer the research question, UoE papers were selected and analysed from past EFL school-leaving examinations. The first subsection gives a general introduction to the EFL school-leaving examination investigated. The second subsection describes the UoE papers analysed. The last subsection 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. Language learners can choose between two levels: The intermediate level is intended to be at Level B1 and the advanced level at Level B2 of the Common European Framework of Reference for Languages (Council of Europe, 2020). The examination consists of a written part, which features four papers (Reading,

Use of English, Listening, and Writing) administered in one sitting with a break in the middle; and of an oral part, which is administered on a separate day after the written part.

A B2 level state-accredited language examination certificate is issued for those test takers who achieve a minimum of 60% on both the written and oral parts of the advanced level EFL school-leaving examination. Those who achieve 40-59% get a B1 level state-accredited language examination certificate (Hungarian Government, 1997). However, the reverse does not apply: A language certificate issued by a certified language centre cannot substitute the EFL school-leaving examination.

The UoE Papers Analysed

The advanced level Use of English (UoE) paper in the focus of the current investigation is administered together with the Reading paper in the first half of the written examination. Test takers have 50 minutes to complete the UoE paper and can score 35 raw points, which are converted to 30 points (Hungarian Government, 2021). The UoE 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 EFL school-leaving examination, test takers must achieve a minimum of 12% on each part of the examination, which is rather low compared to the 40% minimum performance expected from all the other state-accredited language examinations by the Department for Language Examination Accreditation of the Educational Authority (Oktatási Hivatal, 2021a). Therefore, the UoE cut-score set at 12% means that test takers must answer correctly merely five UoE items, the equivalent of four weighted points, in order to pass this part of the examination, but they must achieve 25% overall in the complete examination (oral and written part) for the successful completion of the EFL school-leaving examination.

The advanced level UoE papers were selected for this study from the eight examinations administered in the 2017-2020 period. This period was

chosen because the last modified version of the EFL school-leaving examination was introduced on 1 January, 2017. Although according to the EFL school-leaving examination documentation (Hungarian Government, 2002 modified in 2017) the modifications affected only the level of the intermediate examination, and the oral and the writing parts of both levels, it did result in the modification of the examination system and therefore serves as a reference point in the history of the examination.

Each UoE paper analysed consisted of four tasks, and each paper contained altogether 35 items. The task types included were Error identification (n = 8), Open gap-filling (n = 8), Word formation (n = 8), Multiple-choice cloze (n = 7), and Banked gap-filling (n = 1). The order in which the tasks were presented varied across papers except for the Error identification task that was the last task in all the papers. In terms of response format, the items were selected response items except for those in the Open gap-filling and Word formation tasks, which were limited production items (cf. the Results and Discussion Section for a comprehensive description of the UoE papers and tasks analysed).

Data Analysis

The eight UoE papers containing four tasks each were first solved independently by the two authors. Following this, each author coded the items following a codebook compiled based on the reviewed literature containing information on the constructs the UoE paper intends to measure.

According to the EFL school-leaving examination specification (Hungarian Government, 2002), the UoE paper aims to measure the lexical, grammatical, semantic, and pragmatic abilities necessary for independent communication; more specifically, it tests whether the test takers are able to recognise, complete, or produce lexical items and grammatical structures in-text. The closely related lexical and semantic concepts were merged into the Vocabulary construct in this study. The specification does not define or

elaborate the concept of either lexico-grammar or pragmatic ability, but on the basis of the description of the tested UoE abilities for the purposes of this study pragmatic ability was interpreted as the context-specific, structured, and functional language use measureable through the contextually determined use of vocabulary, grammar, and lexico-grammar—a lexical extension of grammar. As all the UoE items were presented in context, it was assumed that all of them intended to measure pragmatic ability, and therefore they were not coded separately for this feature. Thus, in this study, the UoE paper was considered to operationalise three constructs. An item was coded as vocabulary if in order to answer it correctly the test taker had to make only a semantic decision (i.e., retrieve from memory or select from a set of options a word with a specific meaning); as grammar if the correct answer required a syntactic decision (i.e., a form-related decision); and as lexico-grammar if an item required the combined use of semantic and syntactic decisions.

The codebook was created through a number of development-piloting-revision cycles during independent and joint analyses of advanced level UoE papers administered before 2017. The final pilot was conducted with eight advanced level UoE tasks taken from two examinations administered in 2009. For each coding unit within each construct (eight main- and 16 subunits/Vocabulary, 20 main units/Grammar, four main- and 11 subunits/Lexico-grammar) a descriptive label, an abbreviation, a definition, and examples were provided. Six decision rules were also formulated. The repeated discussions and trials of the codebook versions resulted in a comprehensive yet straightforward and practical tool that both authors could use with confidence and precision when they coded the UoE tasks selected for this study.

Cohen's κ was run to determine if there was agreement between the two authors' independent codings of the items in the 2017-2020 UoE papers. There was very good agreement (Landis & Koch, 1977) between the two coders, $\kappa = .982$ (95% CI, .965 to .999), $p < .0005$. The items that the two

authors coded differently were discussed, and a final coding decision was made and recorded for each differently coded item before the analysis of the data.

The title of the reading passages in the UoE tasks and the images that accompanied each task were also analysed. The authors examined the functional and propositional connection between the titles and the reading passages as well as described the relationship between the content of the images, the title, and the content of the reading passage.

Each UoE item, including the example items, was also coded according to the CEFR level of the targeted lexical items and grammatical structures. The authors systematically looked up the CEFR level assigned to the specific meaning targeted by the lexical items in the English Vocabulary Profile (EVP; 2015) and the CEFR level of the grammatical structures in the English Grammar Profile (EGP; 2015). The EVP and EGP are based on the study of learner language, that is, on what the learners actually know at a given CEFR level, as shown by the Cambridge Learner Corpus and the Cambridge English Profile Corpus. They specify the CEFR levels only for those lexical items and grammatical structures that occurred in the corpora with high enough frequency to be reliably calibrated. As a result, not all the items from the UoE papers could be assigned a CEFR level.

A number of coding decisions were made in connection with CEFR level assignment: No CEFR level was assigned to the fully correct line items in the Error identification tasks because there was no meaningful way to do this. When the vocabulary (V) and grammar (G) components of an item testing lexico-grammar differed in terms of CEFR level, the item was coded according to the highest CEFR level specified in the EVP or EGP for either component (e.g., a G=B1 and V=B2 item was coded as a B2 level item). When several solutions were listed for an item in the official key, the level of the item was determined according to the level of the lowest CEFR level solution

with the application of the minimal effort necessary to solve the item correctly principle.

Results and Discussion

The outcomes of the analysis are presented in terms of the UoE papers included in the analysis and the UoE results of test takers who took the examination, the characteristics of the UoE paper according to the rubric and task types, the characteristics of the input, the distribution of the UoE items across task types and CEFR levels, the constructs assessed, the shortcomings of the scoring key, and some noteworthy task-type-specific problems.

School Leaving Examination Results

Within the 2017-2020 period investigated in this study, altogether 60,971 secondary school students registered for the advanced level EFL school-leaving examination. Of them, 58,183 took the examination and 57,946 (99%) passed (Oktatási Hivatal, 2021c). Based on the analysis of the results of the test takers who had a UoE score recorded, the majority of the test takers managed to get fairly high scores on the UoE paper (n = 57,922; *M* = 20.78, *Mdn* = 21, *SD* = 5.514; Q1 = 17, Q2 = 21, Q3 = 25). Furthermore, altogether 81 students had a converted UoE score of 4 points, and of these 53 (65%) passed the EFL school-leaving examination and became eligible, some (n = 8) with extra EFL points which are awarded for examination records of minimum 45%, for the English major programmes offered in Hungary. This serves as evidence that a student with the minimum acceptable EFL UoE score can actually become eligible for a tertiary English major programme.

Characteristics of the Rubric

Task Instructions and Response Format

The commendable features of the target language instructions are the layout, namely the bulleted format that potentially increases comprehension accuracy with the segmentation of the propositional content to be processed; and the repeated information structure featuring the announcement of the genre of the input text, its topic, task features, and the specification of how test takers should respond to the input followed by a statement about example items. The explicit instructions are also overall comprehensible. As they are not meant to measure language ability, the target language instructions should not require higher than B2 language ability level to understand them. However, the standard reiterated parts of the instructions occasionally contained higher level words (e.g., *grammatically*, C2). The topic announcement parts of the instructions contained B2 (*discovery*, *peculiar*) but also C1 (*relating to*, *commonly*) level words. Whereas the potential problems that the level of the standard instruction parts could cause are counterbalanced by test wiseness, that is, prior experience with the form of the tasks, the comprehension problems caused by the topic announcement part are likely to result in construct-irrelevant variance.

The format of the expected response was straightforward in the case of each UoE task type. The instructions, jointly with the information about the recording method given at the beginning of the paper, and the answers provided for the example items clearly indicated what the test takers must do. Familiarity with the UoE paper most likely also benefited the test takers in terms of the response format.

Task Types

Altogether 32 tasks representing five task types were used in the UoE papers (see Table 1). The regularly included task types were Word formation, Multiple-choice cloze, Open gap-filling, and Error identification; Banked gap-filling only occurred once. Whereas there was some variation in item number

Table 1

Task Types Used in the Analysed UoE Papers

Task type	∑	%	Year	Take 1		Take 2	
				Task no.	Items	Task no.	Items
Word formation	8	25	2017	1	8	1	8
			2018		9	2	9
			2019		9	3	9
			2020		9	1	9
Multiple-choice cloze	7	22	2017	2	9	2	8
			2018		9	-	-
			2019		8	1	7
			2020		9	2	8
Open gap-filling	8	25	2017	3	7	3	8
			2018		8	3	8
			2019		9	2	8
			2020		8	3	8
Error identification	8	25	2017	4	11	4	11
			2018		9	4	9
			2019		9	4	11
			2020		9	4	10
Banked gap-filling	1	3	2018	-	-	1	9
Total	32	100			140		140

per task, in the spring examination periods (Take 1) the number and order of the tasks was constant. Contrarily, in the autumn examination periods (Take 2) the order of the tasks varied except for the Error identification task. In the 2018 Take 2 examination only, the Multiple-choice cloze task was replaced with a Banked gap-filling task.

There was fairly good consistency in terms of the task types included in the UoE paper. With the exception of one, all the task types listed in the specification as possible UoE tasks were used in the analysed papers. Arguably, the unused task type, matching sentence segments to gaps on the basis of the cohesive devices that occur in the text, was not used because the task only allows for the measurement of a very specific discourse feature, namely textual cohesion, with an entire task, a luxury which no high-stakes test aiming to be valid and reliable can afford. The item type should have been removed from the specification or replaced with another functional one.

The task types used recurred consistently across the takes and were consistent across the years except for Take 2 in 2018, when the Multiple-choice cloze task was replaced by a Banked gap-filling task. Whereas the inclusion of the Banked gap-filling task was not in conflict with the specification, the task type is quite different even if we consider only the feature that there are more options per gap than in the case of the Multiple-choice cloze task, and as a result the consistency of tasks across takes was compromised, which does not serve the interest of the test takers.

The order in which the task types were presented was the same in all the Take 1 papers but different, except for the position of the Error identification task, in the case of the Take 2 papers. Although the test takers can choose freely the order in which they attempt the tasks within a paper, the test papers could be parallel in form given that there is no discernible rationale for the alteration of their internal structure, and the current practice may compromise consistency across takes.

Characteristics of the Input

The Length of the Input Texts.

According to the specification, the overall amount of the input text for the UoE paper must be between 700-900 words per UoE paper. The average length of the texts used was 859 words per paper, with a range of 800 to 908 words, which is not fully consistent with the specification. As can be seen in Table 2, the length specification was not observed twice, albeit it could have been avoided with the adaptation of the source texts, which the task designers anyway do, as was revealed by the comparison of the source and input texts.

Table 2

Amount of Text per UoE Papers

Year	Take	Text 1	Text 2	Text 3	Text 4	Total
2017	1	201	227	214	158	800
	2	218	251	230	187	886
2018	1	246	209	231	163	849
	2	207	260	217	140	824
2019	1	217	232	223	143	815
	2	215	257	266	166	904
2020	1	246	256	242	164	908
	2	224	245	256	164	889

Input Text Titles.

Those titles (n = 19) that were found to be functional and therefore most likely helped the test takers (Lemarié et al., 2012) were the ones that were clear summary statements about content (e.g., *World's oldest undamaged shipwreck discovered*) or which were descriptive (e.g., *A famous flight*). On the

contrary, opaque (e.g., *"Archaeology on steroids"*), ambiguous (e.g., *The real junk food project*), ludic (e.g., *Selfie madness: Too many dying to get the picture*), or culturally loaded (e.g., *Speakers' corner*) titles also occurred. These may well be authentic for the original native English speaker target readership within the native context and in the original form of the text; however, if they do not help but rather hinder a B2 language user's comprehension of the input text, they should not be used in a high-stakes test. One of the titles became incomprehensible due to the fact that during its adaptation for the test the subtitle with a much-needed disambiguating function was deleted. This way, the original *Poetry Pharmacy: Meet the poet who prescribes verse instead of pills to 'patients'* was reduced to *Poetry pharmacy*, which is not only opaque but also fails to match the content of the text. Whereas the comprehensibility of some of the titles was enhanced with the image that accompanied the tasks, it was not the case with this title.

Images Accompanying the Input Texts.

Each UoE task featured an image intended to function as a representational picture. Most of these (n = 21) were acceptable as they helped set the context and potentially helped the activation of the schema necessary for text comprehension. Those images were especially helpful which enhanced the comprehension of the titles by visually depicting their meaning and thus foreshadowing the content of the text. This was the case, for example, with *Ancient roots of cream tea discovered*, where the image next to the title illustrated what "cream tea" looked like. The remaining images were dysfunctional because they were indiscernible, watermarked, mismatched the content, or did not relate clearly to the text and, in the best case, became functional only in retrospect after the passage was read and understood. It is important that, if changing an image is inevitable, an original image be replaced with another one with a matching function. The original image accompanying the *Poetry pharmacy* source text was a photo of a clearly discernible bookshop with *Poetry pharmacy* written across the shop window

in large bold letters. The intention was clearly to help passers-by arrive at the inference explicitly stated in the subtitle of the source text. However, the test taker was deprived of this rich context-setting input when the image was replaced in the test task with a parchment and a quill, which could not possibly help disambiguate the abridged obscure title.

Faulty Input Text.

The text in the second task in the 2020 Take 2 paper contained a mistake that was not only left uncorrected but was turned into an unsolvable item. The task developers selected an authentic source text that contained a discourse level problem and adapted it for the task. However, whereas they did notice that there was a problem with the source, they failed to mend it. A careful reader can easily notice that the last paragraph in Example i given below does not make sense from Gap 25 on in either the source or the test version. The two consecutive paragraphs from the source and the test are reproduced below:

i. (Source version) The team reportedly said they intended to leave the vessel where it was found, but added that a small piece had been carbon dated by the University of Southampton and claimed the results "confirmed [it] as the oldest intact shipwreck known to mankind". The team said the data would be published at the Black Sea MAP conference at the Wellcome Collection in London later this week.

(Test version) Researchers believe that the ship is similar to a type of trading ship that has previously (23) _____ [Key: *been*] seen only on the side of items of ancient Greek pottery such as the 'Siren Vase'. The vase dates from about the same period and shows a ship like that with Odysseus tied to the mast (24) _____ [Key: *to*] resist the sirens' songs.

> The ship (Test version)/It (Source version) was among more than 60 shipwrecks found by the international team of archaeologists and scientists, (25) _____ (*which* in the source; Key: *which*) has been on a three-year mission to explore the (*depths of the*, Source) Black Sea to gain a greater understanding of the impact of prehistoric sea-level changes. (2020 Oct., Task 3)

(Original text available at https://www.theguardian.com/science/2018/oct/23/oldest-intact-shipwreck-thought-to-be-ancient-greek-discovered-at-bottom-of-black-sea)

The item writers changed *it* to *ship*, but this did not solve the discourse level problem: the referent in the first part of the text is the shipwrecked vessel, whereas the second part refers to the research vessel. It is rather unfortunate that the item writers decided to turn the conspicuously uncohesive and incoherent paragraph into an unsolvable item. Such items result in construct-irrelevant difficulty.

Faulty Punctuation.

Altogether 26 punctuation errors were found in the UoE papers tasks (see Table 3). There were six types of errors: Oxford comma (n = 3, Example ii), unnecessary comma (n = 2, Example iii), comma omitted before a non-finite clause (n = 1, Example iv), overpunctuation (n = 1, Example v) due to overwriting the correct English punctuation in the source with the application of Hungarian punctuation rules, comma after an introductory phrase (n = 11, cf. Example ix in the Item distribution across CEFR levels Section below), and comma between coordinated clauses (n = 8, cf. Example xii in the Scoring key Section below).

ii. ... bread, clotted cream and little strawberry jam. (2019 Oct., Task 4)
iii. ... a pub in (0) _____ (centre), Wimbledon, ... (2020 May., Task 1)

iv. The knocker-upper went around the neighbourhood with a long stick was tapping on people's windows. (2019 May, Task 4)

v. ... looking for a loan–but did not qualify –, Baschiera would ... (2020 Oct., Task)

Table 3

Frequency of Punctuation Errors per Task Type

Task type	Σ	Location		
		In-text	Instruction	Example item
Error identification	13	12	1	-
Word formation	7	6	-	1
Multiple-choice cloze	3	3	-	-
Open gap-filling	3	3	-	-
Total	26	24	1	1

Given that punctuation has a disambiguating role, there cannot be any punctuation errors in the texts whose comprehension is necessary for the correct solution of items. This is especially true for the error identification task, where the test takers have to rely on the precise understanding of the sense of the text to the largest degree. It is rather disconcerting that most of the punctuation errors were identified in this task type, especially because two of the tasks contained three mistakes each. This is a source of construct-irrelevant difficulty.

Analysis of the UoE Items

Item Distribution across Task Types.

Altogether 320 UoE items were coded of which 280 were regular and 40 example items. In each UoE paper, there was one example item per task

except for the Error identification task type, where two example items were given, one for a correct line and one for a line containing an error. The descriptive statistical data in Table 4 provides further details about the distribution of the items.

Table 4

Item Distribution across Tasks

Items per task type	Multiple-choice cloze	Error identification	Word formation	Open gap-filling	Banked gap-filling
N	65	79	70	57	9
Median/Mode	8	9	9	9	8
Range	4	4	2	1	0
Minimum	7	7	8	8	8
Maximum	11	11	10	9	8

A Kruskal-Wallis test provided very strong evidence of a difference ($p < 0.001$) between the mean ranks of at least one pair of task types. The Dunn's pairwise tests carried out showed that there was very strong evidence ($p < 0.001$ and $p < 0.008$) of a difference between the Open gap-filling and Error identification as well as the Multiple-choice cloze and Error identification task types. The median item number for the Error identification task type was 9.5 compared to 8 in the Open gap-filling and 8 in the Multiple-choice cloze task types. There was no evidence of a difference between the other pairs.

The differences identified mean that the UoE papers contained significantly more Error identification items than Open gap-filling or Multiple-choice cloze items. The strong preference for Error identification items would be acceptable if the task type tested UoE directly rather than indirectly, in other words the test taker's ability to use language rather than

their knowledge about language. By extension, the fact that there was no statistically significant difference between the items in the Error identification and Word formation tasks suggests that there was a similar preference for these item types. However, as discussed below, the items in the Word formation tasks target a very limited component of the UoE construct, and as a result this is indicative of construct underrepresentation.

Item Distribution across CEFR Levels.

The 23 fully correct line items in the Error identification tasks were excluded from the analysis of CEFR levels because they could not be assigned a meaningful CEFR level. Furthermore, no information was found in the English Vocabulary Profile (EVP) and the English Grammar Profile (EGP) about the CEFR level of 14 items. Therefore, altogether 37 items were not coded in the analysis of the CEFR level of the UoE items (see Table 5).

Table 5

Distribution of the UoE Items across CEFR Levels

CEFR level*	Σ	%	Cumul. %
A1	46	16	16
A2	52	18	35
B1	89	31	66
B2	82	29	95
C1	10	4	99
C2	4	1	100
Total	283	100	

*37 items could not be coded

The analysis revealed that 66% of the coded items were below and 5% above B2 level. Only 29% of the items targeted the level intended to be measured by the UoE paper. It is rarely the case that all the items in a test aimed at a specific level actually do measure the one level only, but the majority of the items must do. The fact that 66% of the items could be solved by candidates of language ability below the level at which the test is targeted is disconcerting. The test specification claims that the texts in which the items are embedded are at B2 level (Hungarian Government, 2002), which may actually be the case. However, even assuming that this is the case, a large number of the items can be solved by just looking at the immediate context of the gap, which often does not require B2 level text comprehension. Consider the following examples:

vi. ... to (6) _____ (form) a line ... (2017 May, Task 1, Word transformation/conversion; LG; Key: *form*)

vii. ... George Washington (13) _____ always remain on the $1 ... (2017 Oct., Task 2, Multiple-choice cloze/modal verb; G; Key: *must* selected from among 4 given options)

viii. ... he arrived (5) _____ Paris ... (2019 Oct., Task 1, Multiple-choice cloze/prepositional verb; V; Key: *in*)

In Example vi, *form* is a B1 and *line*, based on which the gap can be solved without the need to consult more co-text than the infinitive form indicator *to*, is an A2 level word. In Example vii, the proper name is a given, *always* is an A1, and *remain* is a B1 level word for a gap that is to be filled with the A2 level *must*. In Example viii, *arrive in* is an A2 level prepositional verb taught as one lexical unit. These examples illustrate the items that can be solved with very restricted local comprehension with lower than B2 level language ability.

The overall text level is also not very likely to be an applicable task feature because most of the items in the UoE paper require local comprehension, that is, they are not text-based but sentence-based items.

Few are the items for which the text taker must combine meaning from more than one sentence (see Table 6). A rare excellent item illustrating the need for global processing of the text is the following:

ix. At a meeting in Sweden [sic, missing comma] some designers and engineers got together to talk about this concept. (12) _____ of the developers wanted a better name for the project than MC-Link or Biz-RF. So in his presentation [sic, missing comma] he gave the new model a codename ... (2019 May., Task 2, Multiple-choice cloze: A/Some B/One C/None D/Both)

The test taker can select the correct solution based on the information available in the third sentence as it is twice cued by the pronouns *his* and *he*.

Table 6

Items Requiring Local or Global Comprehension

Comprehension type	∑	%
Global	12	4
Local	308	96
Total	320	100

The predominance of local items is strongly suggestive that in spite of the vocabulary, grammar, and lexico-grammatical items embedded in a reading passage, the UoE test is markedly a sentence- or phrase-based test (cf. Examples viii and vi) in which independent items focus on single language ability components. This does not fully match the definition of the UoE construct and is therefore a potential construct underrepresentation problem. What is reassuring to a certain degree is that a Kruskal-Wallis test provided no evidence of significant differences between the UoE papers across takes in terms of the CEFR level of the items: It seems that each cohort of test takers is measured with a test that is equally not on the intended level.

Constructs Assessed.

The UoE paper was meant to assess three constructs: vocabulary, grammar, and lexico-grammar. Since the UoE items are presented in a text and therefore require test takers to process text and rely on textual meaning, the items inevitably also test reading ability. However, in order to make sure that the ability assessed is not reading comprehension, the UoE items must be presented in a text where the co-text of the items is less complex than the tested level. This was not found to be the case.

The specification (Hungarian Government, 2002) states that the difficulty level of the UoE texts is lower than that of the texts in the Reading paper, but that the texts are in terms of language and content B2 level texts. This is in itself conflicting information. It is also stated that the input text may contain lexis or structures that are above the tested level, but they are not necessary for the successful completion of the task. These statements raise a number of issues: If the UoE paper measures B2 level language ability and the input is of lower difficulty level than in the Reading paper of the same examination, what level does the Reading paper measure?

However, what is of immediate concern for this study is that the UoE texts are intended to be B2 level texts in terms of language and content. This contradicts the expectation that primarily the targeted lexis and structures are to be at the tested level in a UoE test or else the measurement becomes excessively contaminated by reading ability. Since this condition was not met, the UoE paper is most likely fundamentally affected. Furthermore, the analysis of the tasks revealed that most of the items did not target B2 level lexis or structures (see Table 5 above). If a B2 level text is to be interpreted by the test takers in order to answer UoE items below B2 level, the paper most likely turns into a reading comprehension test only. The papers in the written part of the examination are equally weighted; consequently, if the UoE paper measures primarily reading, then 50% of the written paper measures reading.

Table 7 presents a summary of the UoE constructs targeted by the example and regular items in the papers. The majority of the items targeted structures, the second most frequent item type measured lexico-grammar, and vocabulary items were the least frequent. Overall, more grammatical decisions had to be made by the test takers than lexical ones, even if we consider lexico-grammar items as items that require partly lexical and partly grammatical decisions.

Table 7

UoE Constructs Targeted and Their Distribution of Constructs across Task Types

UoE construct*	Banked gap-filling	Multiple-choice cloze	Open gap-filling	Error identifi-cation	Word forma-tion	Total
Vocabulary	6	18	14	43	0	81
Grammar	3	43	50	28	0	124
Lexico-grammar	1	4	8	1	78	92
Total	10	65	72	72	78	297

*The 23 correct line items in the Error identification tasks excluded.

Table 7 also reveals that lexico-grammar was tested only minimally in most task types, except for the Word formation task, in spite of the fact that it could also be tested in other task types.

Based on the analysis of the specific language phenomena targeted, it can be noted that the items do not reflect a concerted effort made in order to systematically and extensively assess the test taker's ability. Most vocabulary items assessed denotation, which would be acceptable if the majority of the targeted lexical items were B2 level items. Additionally, meaning relations

(synonymy, hyponymy, antonymy, meronymy) should also be tested. Given that level-specific information on grammatical structures became available in 2001 with the publication of the Vantage language learning objectives (Van Ek & Trim, 2001) and that a range of free resources based on EFL learner corpora (e.g., English Grammar Profile Online) are available and provide empirical level-specific information about the grammatical ability of B2 level EFL learners, typical B2 level structures or B2 level aspects of structures should be targeted by the grammar items in the UoE paper. Finally, lexico-grammar is almost exclusively measured in the Word formation task and mostly through suffixation. It should be measured in a more diverse way than primarily word formation. These features of the UoE tasks are indicative of construct underrepresentation. The UoE aspects targeted by the items in the analysed UoE tasks are included in the Appendix. A more detailed discussion about the ratio and range of the targeted aspects is beyond the scope of this paper, but a cursory look is enough to reveal some of the potential problems.

Scoring Key.

The scoring key provided for the UoE papers is problematic in a number of ways. First, the way the alternative solution options are given is confusing. Whereas for certain items the correct solutions are listed (e.g., *anxiety/anxiousness*), for other items an embedded list is also given in the *exhausted* (also acceptable: *exhaustion*) form. This is potentially confusing for scorers as it seems to suggest that some solutions are more acceptable than others.

The scoring key was not complete for all the tasks. The following examples illustrate two instances of this problem:

x. ...on the 400th anniversary of Shakespeare's death, archaeologists believe they have found evidence (9) _____ support the claims. (2019 Oct., Task 2)

xi. Weeks of (5) _____ (investigate) of "what's stopping them from connecting" may be necessary to convince someone ... (2020 May, Task 1)

In Example x, only *to* is given as a solution whereas *that* is also correct. In Example xi, in addition to *investigation(s)* given in the key, *investigating* (B2) is also correct. Both of the missing correct solutions are straightforward and *that* is unquestionably also an easy solution that some test takers will have provided as an answer. If the scorers followed the scoring key, part of the variance in the scores of test takers who provided the unlisted solutions did not represent their language ability level as it was affected by the incomplete scoring key.

On the contrary, in some cases the solutions given were implausible for the level and context of the test. For the following item,

xii. My friends walked on [sic, missing comma] and I stood there (7) _____ (tremble) with fear. (2020 Oct., Task 1)

the solutions given in the key were *trembling*, *atremble* and *a-tremble*, the last two of which are possible but very low frequency items that most dictionaries do not even list. It is highly likely that they are the results of task moderation where the level of the test was disregarded.

Two examples for incorrect solutions provided in the key, which is yet another source of construct-irrelevant variation, are the following:

xiii. They also established that a skull in a church in Beoley, 20 miles away, (11) _____ according to local legend belonged to the playwright, was in fact ... (2019 Oct., Task 2)

xiv. And yet, knowing when to be tight-lipped can give you the upper hand in everything from sales deals and pay (2) _____ (negotiate) to presentations and staff development. (2018 May, Task 1)

Example xiii, is a straightforward case. This specific non-restrictive relative clause cannot start with the relative pronoun *that* listed as a solution in the key. For the item in Example xiv, the key lists the following solutions: *negotiations*, *negotiation*, and *negotiating*. The second solution given in the key is wrong because it does not fit syntactically the parallel construction *deals ... pay negotiations ... presentations* because it is a singular noun in a list of plural nouns. The original source text also uses the plural form.

For several items, a number of alternative solutions of different CEFR levels are accepted as equivalent for the B2 level test. For the following item,

xv. And they're (22) _____ impressed with the replacement. (2018 Oct., Task 3)

the solutions given in the key are *not* (A1), *less* (A2), *hardly* (B1), *barely* (B2), and *scarcely* (C2). Arguably, the most likely solution to have been chosen is *not*. This is another task design problem, and items that allow such a range of CEFR levels should be avoided similarly to items that can be filled in correctly with almost any modal verb (cf. 2017 Oct., Task 3, Item 24 where any one of *can*/*could*/*may*/*might*/*would* is correct.)

The distractors in several items were faulty. The following examples illustrate this problem:

xvi. The name Bluetooth was meant to be temporary (14) _____ something official came together. (2019 May, Task 2, Options: A. before, B. so, C. until, and D. when)

xvii. He took small amounts from wealthy clients and transferred (10) _____ those who couldn't qualify for credit. (2020 May, Task 2, Options: A. for, B. to, C. them for, and D. them to)

In Example xvi, both A/*less frequent* and C/*more frequent* are correct solutions. In Example xvii, only C is accepted as a correct solution whereas both C and D are correct. Similarly to the other problems discussed in this

section, such faulty items result in construct-irrelevant variance and are as such a major threat to construct validity.

Task Specific Problems

Two of the task types, namely Word formation and Error identification, raised a number of issues that need to be addressed.

Word Formation.

As Table 8 shows, the items in the Word formation task tested primarily derivational morphology only through suffixation. There were few cases of prefixation only as well as of combined suffixation and inflection. It was also found that in some takes almost all the items tested only suffixation (e.g., 2018 May, Task 1, 8 suffixation items/9 total items). In a few cases, inflectional morphology was also used, and each case was an instance of the inflectional plural suffix –s (e.g., *profession*, N → *professional*, N → *professionals*, Ns or *change*, V → *change*, N → *changes* Ns). Other than conversion, no other word formation processes (e.g., compounding, blending, clipping, or back formation) were used. This and the overrepresentation of items that test only suffixation resulted in word formation construct underrepresentation. The inclusion of inflection, and only in the form of the plural suffix, is also indicative of construct underrepresentation. In addition to the above, the analysis revealed that the majority of word formation items (49%) were not B2 level items and most of those (44%) were below B1 level. As a result, the task is most likely to have a negative washback effect on teaching and learning in the form it was included in the UoE papers.

Table 8

Word Formation Processes Targeted by the UoE Items

Process	Total		Item type	
	∑	%	Example	Regular
suffixation	57	73	5	52
prefixation	7	9	0	7
suffixation + inflection	7	9	2	5
conversion	3	4	0	3
prefixation + suffixation	2	3	0	2
conversion + inflection	2	3	1	1
Total	78	100	8	70

Error Identification.

The task design problem found in the Error identification tasks was that in 6 out of the 8 tasks extra words that either did not fit semantically or were syntactically incorrect occurred at the very beginning (n = 3) or at the end (n = 6) of the lines. Because this positioning of items results in construct-irrelevant variation induced by the format of the task, it should be avoided (cf. Alderson & Cseresznyés, 2003a).

Conclusion

The analysis of the UoE paper in the high-stakes advanced level EFL school-leaving examination revealed a number of problems that raise validity, reliability, fairness, and consequential concerns. Given the actual CEFR level tested, it is very likely that through false positive classification errors even students with exceptionally low UoE scores can potentially become eligible for an English major programme in addition to the two-language-major students admitted with B1, popularly known as “tourist” (cf. Council of Europe, 2020) level of language ability.

Furthermore, the UoE paper is very likely to be affected by construct validity shortcomings. The construct underrepresentation problems identified are that the predominance of local items is strongly suggestive that the paper is markedly a sentence- rather than the text-based paper it is intended to be. If the UoE paper indeed measures B2 level reading, then 50% of the written paper assesses reading comprehension, mostly locally in the UoE paper. Most vocabulary items assessed only denotation and even that not primarily at B2 level. Several structures were not typical B2 level structures, or the B2 level structural aspects were not targeted. Word formation items measured a very limited part of the UoE construct and were significantly more frequent than two other task type items.

The analysis revealed a number of construct-irrelevant variance issues: The consistency of tasks across takes was compromised in terms of task types, as was the consistency of the internal structure of the UoE paper in terms of task type order. What could have made the tasks irrelevantly difficult were the following: The titles and images were occasionally dysfunctional or otherwise problematic. Some of the input texts were faulty in terms of coherence and punctuation mistakes. The scoring key was occasionally incomplete, wrong, or featured unlikely solutions, and alternative solutions of different CEFR levels were accepted as equivalents. Some items contained faulty distractors. The error identification tasks suffered from a basic task design error.

Admittedly, this investigation focused on only one paper of the school-leaving examination, so the remaining papers and the oral part may counterbalance its weaknesses. Moreover, the definition of lexico-grammar may be narrow. However, there is little doubt that for validity, reliability, and fairness concerns as well as for washback and consequential reasons, the UoE paper is to be considerably improved.

References

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

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

Alderson, J. C., & Cseresznyés, M. (2003b). *Guidelines for item writers. Into Europe: Reading and use of English*. British Council Hungary. https://www.lancaster.ac.uk/fass/projects/examreform/Media/GL_Reading&Use.pdf

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

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

Docherty, C. (2015). *Revising the Use of English component in FCE and CAE*. *Research Notes*, *62*, 15–20.

English Grammar Profile Online (2015). *English Profile: The CEFR for English*. Cambridge University Press. https://www.englishprofile.org/english-grammar-profile/egp-online

English Vocabulary Profile Online (2015). *English Profile: The CEFR for English*. Cambridge University Press. https://www.englishprofile.org/wordlists/evp

Felsőoktatási felvételi tájékoztató - 2021. szeptemberben induló képzések [Higher education admission bulletin for September 2021] (2020.12.18). https://www.felvi.hu/felveteli/jelentkezes/felveteli_tajekoztato/FFT_2021A

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

Gledhill, C. (2012). A lexicogrammar approach to checking quality: Looking at one or two cases of comparative translation. In I. Departere (Ed.), *Perspectives on translation quality* (pp. 71–97). De Gruyter Mouton. https://doi.org/10.1515/9783110259889.71

Hale, G. A., Stansfield, C. W., Rock, D. A., Hicks, M. M., Butler, F. A., & Oiler, Jr., J. W. (1988). *Multiple-choice cloze items and the Test of English as a Foreign Language*. *TOEFL Research Reports*, *26*. Educational Testing Service.

Halliday, M. A. K. (2004). *An introduction to functional grammar* (3rd ed. Revised by Christian M. I. M. Matthiessen). Hodder Arnold.

Hungarian Government (1997). 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://net.jogtar.hu/jogszabaly?docid=99700100.kor

Hungarian Government (2002). OM rendelet az érettségi vizsga részletes követelményeiről [(Ministry of Education) on the Specific Requirements of Upper Secondary School Leaving Examinations] Ministerial Decree 40/2002. (V. 24.). http://njt.hu/cgi_bin/njt_doc.cgi?docid=65297.357363

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.

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

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

Lemarié, J., Lorch, R. F., & Péry-Woodley, M.-P. (2012). Understanding how headings influence text processing. *Discours*, *10*, 3–21. https://doi.org/10.4000/discours.8600

Lindner, M. A., Lüdtke, O., Grund, S., & Köller, O. (2017). The merits of representational pictures in educational assessment: Evidence for cognitive and motivational effects in a time-on-task analysis. *Contemporary Educational Psychology*, *51*, 482–492. https://doi.org/10.1016/j.cedpsych.2017.09.009

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–104). American Council on Education and Macmillan.

Messick, S. (1995). Validity of psychological assessment. *American Psychologist*, *50*(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Oktatási Hivatal (2020). Vizsgateljesítmények az egyes vizsgarészeken 10 vizsgatárgyból (a nem megjelenteken és a mentességgel rendelkezőkön kívül). [Exam performance in each of the ten examination subjects (excluding those who did not attend and those with an exemption)] https://www.ketszintu.hu/publicstat.php

Oktatási Hivatal (2021a). Akkreditációs kézikönyv [Accreditation handbook.]. Department for Language Examination Accreditation of the Educational Authority. https://nyak.oh.gov.hu/nyat/doc/ak2021/word/Akkreditacios_Kezikonyv_2021.pdf

Oktatási Hivatal (2021b). Elmúlt évek statisztikái. [Statistics of recent years.] https://www.felvi.hu

Oktatási Hivatal (2021c). Statisztikák, vizsgaeredmények. [Publicly available statistical data and school leaving examination results.] https://www.ketszintu.hu/publicstat.php

Pearce, M. (2007). *The Routledge dictionary of English language studies*. Routledge.

Purpura, J. E. (2004). *Assessing grammar*. Cambridge University Press.

Read, J. (2000). *Assessing vocabulary*. Cambridge University Press.

Sinclair, J. (1991). *Corpus, concordance and collocation*. Oxford University Press.

Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. Routledge.

Spolsky, B. (1995). *Measured words*. Oxford University Press.

Tankó, Gy. (2005). *Guidelines for item writers. Into Europe: The writing handbook*. British Council Hungary. https://www.lancaster.ac.uk/fass/projects/examreform/Media/GL_Writing.pdf

Tankó, Gy. (2012). *Professional writing: The academic context* (Rev. 2nd ed.). Eötvös University Press.

Van Ek, J. A., & Trim, J. L. M. (2001). *Vantage*. Cambridge University Press.

Weir, C. J. (2013). *Measured constructs: A history of Cambridge English language examinations 1913 – 2012*. *Research Notes*, *51*, 2–10.

Zeronis, R., & Elliot, M. (2013). *Development and construct of revised Cambridge English: Proficiency*. *Research Notes*, *51*, 22–30.

Appendix

UoE aspects targeted by the items in the analysed UoE tasks

UoE Aspect	Specific aspect	Subcategory	Σ
Grammar (n = 124)	tense use	past simple (n = 9) past perfect simple (n = 5) present perfect simple (n = 3) present perfect passive (n = 3) going to future (n = 1) going to future (past)(n = 1) past continuous tense (n = 1) past perfect simple passive (n = 1) present simple (n = 1)	25
	preposition**		21
	conjunction		15
	relative clause	defining relative clause (n = 6) non-defining relative clause (n = 6) defining relative clause replaced by a participle construction (n = 1) relative infinitive clause (n = 1)	14
	adverb		12
	modal verb		9
	pronoun	personal (n = 5) possessive (n = 2) demonstrative (n = 1)	8
	to infinitive		7
	conditional clause	type 2 (n = 3)	3
	article		4
	determiner		2
	adjective		1
	causative + tense		1
	non-finite clause		1
	syntagmatic pattern		1

UoE Aspect	Specific aspect	Subcategory	Σ
Vocabulary (n = 81)	denotation		56
	prepositional verb		8
	polyword		6
	idiom		5
	lexical collocation		2
	phrase		2
	phrasal constraints		1
	phrasal prepositional verb		1
Lexicogrammar (n = 92)	word formation	suffixation (n = 57)	78
		prefixation (n = 7)	
		suffixation + inflection (n = 7)	
		conversion (n = 3)	
		conversion + inflection (n = 2)	
		prefixation + suffixation (n = 2)	
	phrasal verb		5
	syntagmatic pattern		5
	phrasal constraints + verb form		1
	phrasal verb + -ing form		1
	sense fit + grammatical form		1
	gerund + lexical collocation		1
Total			297

*Excluding the 23 correct lines / Error identification tasks

**see Vocabulary for prepositions used in prepositional verbs

4.

A Glimpse into the English Classroom: Results of an Observation Study in Hungarian Secondary Schools

Brigitta Dóczi
doczi.brigitta@btk.elte.hu

Kata Csizér
wein.kata@btk.elte.hu

Abstract

The aim of the present article is to provide some insights into English lessons in Hungary to explore classroom work and classroom characteristics in depth. The importance of our research lies in the fact that there are very few studies in Hungary that use real-time classroom observation guides to collect data, and we are not aware of any such current investigation. To fulfil the aim of our study, we designed and piloted an observation scheme and collected data from 47 English lessons across the country during the 2019/2020 school year. Our findings indicate that teachers used a great variety of input, while students' output was more limited in quantity and variety. In terms of modality, we observed similar levels of frontal, individual, and pair work with group work being less typical. English dominated in the classes with the mother tongue as a helping tool for the teacher and a possible avoidance strategy for the students. One of the most important outcomes of the present study is the development and piloting of the observation protocol, which might serve as a useful tool for gaining a deeper understanding of classroom processes in second language teaching.

Keywords: classroom observation, task-based language teaching, input tasks, output tasks

A Glimpse into the English Classroom: Results of an Observation Study in Hungarian Secondary Schools

Research in applied linguistics and language pedagogy in Hungary has long been motivated by quality assurance issues as often quoted European statistics reveal that the foreign language knowledge of the Hungarian population is lagging behind those of other European countries (European Commission, 2012; European Statistics, 2016). Language policy makers have tried to contribute to the enhancement of the quality of education in Hungary to increase the foreign language knowledge of Hungarian learners of English. First, there is no lack of modernization efforts in foreign language education in the country, and these efforts are often backed up by empirical evidence, which often shows positive results in terms of language knowledge (e.g., Einhorn, 2007; Kapitánffy, 2001; Medgyes & Öveges, 2005; Öveges & Csizér, 2018; Öveges & Kuti, 2018; Vágó, 1999). Second, content regulations have been updated several times in the last 20 years (e.g., the National Curriculum has had five versions), based on the Common European Framework of Reference for Languages (Council of Europe, 2001), and foreign language instruction has been treated as a competence of key importance.

The apparent discrepancy between favourable educational frameworks and initiatives to reinforce school language education and the relatively low level of foreign language proficiency of the Hungarian population raises the need of an in-depth exploration of classroom work. One possible way of collecting data on language classrooms is direct observation. Although observation schemes are widely used data collection instruments in international education research (e.g., Ahmed et al., 2018; Campbell, 2017; Elmendorf & Song, 2015; Fletcher, 2018; Stallings & Mohlman, 1988), they are relatively rare in applied linguistics and language pedagogy.

In Hungary, there is a lack of up-to-date observational data and knowledge of what is happening in English classrooms, as even well-funded national studies have been reluctant to include observational data on teachers (e.g., Dörnyei et al., 2006; Öveges & Csizér, 2018). Therefore, the aim of this study was to collect data in real time on classroom processes to describe the instructional characteristics of English language teaching. To fulfil this aim, an observational tool was designed to measure teacher input, student output, forms of instruction, flow of information and language use in the classroom. The purpose of this article is to provide an overview of both the pilot process and the data collected with the finalized instrument. Unfortunately, due to the COVID situation in 2020, we were not able to collect the volume of data we had planned, so nationwide generalization is not possible, therefore, the results presented here should be treated as preliminary research information.

Background to the Study

In this part of our chapter, we first summarize the role of classroom observation studies in applied linguistics. Next, we provide some details about the available classroom observation schemes. In the second part, we outline the classroom-related concepts that informed the development of our observation scheme.

Observation Schemes as Research Tools in Applied Linguistics

According to Moyles (2002), observations are "permanent and systematic records of social interactions and physical settings" (p. 174), such as the classroom. Secondly, they are useful for triangulation because they can "enrich and supplement data" (p. 174) collected by other instruments thus increasing the reliability of the research. Thirdly, observation schemes are versatile research tools that can be used to collect either qualitative or quantitative data (Dörnyei, 2007). Highly contextualized qualitative data can

be collected by unstructured observation schemes usually used in ethnographic studies, or pre-piloted, structured observation schemes collect quantitative or quantifiable data that are often subjected to quantitative analysis with an aim of providing generalizable results. The former is outside the scope of the present investigation, so here we focus on outlining some structured observation schemes.

As mentioned in the introduction, structured observation schemes are often used in general educational research for various purposes, including teachers' appraisal, in-service teacher development, as well as a learning tool for pre-service teachers (for such examples, see Ahmed et al., 2018; Campbell, 2017; Elmendorf & Song, 2015; Fletcher, 2018; Stallings & Mohlman, 1988). In addition, Stallings and Mohlman (1988) delineated four aspects of structured observation: the setting, observational schedule, unit of time, and method for recording the data. Moreover, the observation scheme includes information on the social environment, preparation and planning, teaching and instruction, assessment, and self-reflection (Elmendorf & Song, 2015). Another example of observation tools is the use of peer-observation schemes in pre-service teacher education to provide feedback to individual teachers to enhance their understanding of curricula, their planning skills, their teaching styles/techniques, their interpersonal skills, and their skills of self-evaluation (Ahmed et al., 2018). As these tools are usually focused on teaching, they cannot be directly transposed and applied in foreign language classrooms; therefore, applied linguists have designed a number of additional schemes, which are described here briefly to justify the need for an observation scheme developed specifically for the purpose of this project.

One of the first observation schemes used in applied linguistics is the Foreign Language Interaction Analysis (Allwright & Bailey, 1991; Chaudron, 1988; Moskowitz, 1971). In this observation scheme, the coding work needs to happen in real-time, by using very short intervals. The tool contains three categories (teacher talk, student talk and miscellaneous) and 22 subcategories. The tool has been used in various contexts around the world,

including Hungary, and it was a source of motivation for Nikolov and her colleagues to do an observation project in Hungary (Nikolov, 1999). In this project, 118 English classes were observed in 55 secondary schools in Hungary, and the observation scheme was used in conjunction with an instrument that collected data on teachers' perception of classrooms. The units that were observed in this study included different types of activities (speaking, writing, listening and reading) and 22 possible texts. Nikolov's conclusion was discouraging at the time: Classes were mostly teacher-centred without pair- and group-work, and Hungarian was used excessively. The lessons were mainly questions and answers, translation, reading aloud and grammar exercises.

Another instrument focusing on classroom-level processes was developed in Canada by Spada and Fröhlich (1985) and it measured the communication orientation of language teaching. The first part of the tool collected data in real-time (at 1-minute intervals) about classroom events based on a pre-defined list of 48 possible categories, including characteristics of classroom organization, content, student modality and materials. The second part focused on the communicative characteristics of each activity and was to be completed based on the recordings of the lesson. Similarly to the previous part, the pre-defined categories allowed the results to be quantified. The impact of this tool was attributed to not only its potential empirical use, but also to the fact that it served as a basis for collecting observation data on a particular individual difference variable. Guilloteaux and Dörnyei's (2008) observation scheme was developed to explore teachers' motivational practices, while the perception of students' motivated behaviour was also measured. As this tool collected data on both teachers and students within the same classroom context, the strength of the instrument was clearly that it produced comparable data, and thus the effect of teaching could be assessed. The instrument developed by Guilloteaux and Dörnyei (2008) was also used in a large-scale Swedish project (Henry et al., 2019) to collect information on the motivational practices of teachers

described as successful motivators. Based on the collected data, the authors outlined a sustainable motivational practice for teachers that was based on self-reflection and collegial discussion.

The scarcity of the use of observation schemes in applied linguistics, and more specifically in our immediate field of study of individual differences, is attested to by the fact that in their comprehensive overview of individual difference (ID) research, Dörnyei and Ryan (2015) presented only one study in which an observation tool was used, and that was the aforementioned Motivational Orientation of Language Teaching (MOLT) (Guilloteaux & Dörnyei, 2008). The lack of observational data raises quality-control issues pertaining to the validity of classroom-related research in general, as most data collected in the field relies on self-reported information with possible participant biases in both sampling procedures and empirical data. However, we can think of several reasons to explain why researchers often fail to include an observation tool to their studies. First, most investigations in applied linguistics, and particularly individual differences research focus on the learner (Csizér, 2020; Dörnyei, 2019), and, thus, less attention is paid to the work of teachers, hence direct classroom observations are often not considered relevant. Second, even if teachers are the focus of a study (e.g., Öveges & Csizér, 2018), classroom observations can be seen as a highly intrusive and threatening procedure for teachers. Finally, the development, testing and use of classroom observation tools is time-consuming and hardly cost-effective (Dörnyei, 2007), and, therefore, their use is usually avoided even in well-funded research projects.

Concepts Related to Classroom Processes

Despite the challenges of classroom observation, it is a valuable tool; particularly when the focus of enquiry is clear and observable, such as the nature and building blocks of teaching, which is the topic of this section. In what follows, we define the key concepts of input, output and form-focused instruction, and then discuss their relevance in second language teaching.

The Role of Input and Output in L2 Acquisition and Instruction

There has long been a debate about the role of input and output in second language learning. Krashen's Input Hypothesis (1981) was a pioneering idea, which claimed that to acquire a language, learners need exposure to sufficient comprehensible input, comprising the language elements that the learner already knows, as well as a necessary amount of new information (i+1). There were several unresolved issues with Krashen's theory, such as the vagueness of the terms comprehensible and sufficient, which led to some fine-tuning of the theory and the emergence of the Interaction Hypothesis (Long, 1985). Long's theory posits that second language proficiency is facilitated by comprehensible input that has been simplified but also argues that it is equally important to aim for negotiation of meaning, which is the basis of two-way communication.

The move from overreliance on input towards interaction also highlighted the importance of output and the need for productive knowledge in language learning. Swain's Output Hypothesis (1985) claimed that in challenging linguistic situations, learners make more effort to get their meaning across. The discussion then focused on output, namely the mismatch between two very different learning environments: the classroom, where input is modified or simplified for the sake of the learner (thus output may be different or less) and naturalistic settings, where all the participants need to negotiate for meaning in order to achieve successful communication (Pica, 1987). Due to the imbalance of roles in a classroom setting, it is more difficult to achieve a balanced share of communicative goals. Ellis (1994), who proposed changes to the hypothesis, accepted the usefulness of comprehensible input but questioned whether it was sufficient for second language acquisition, arguing instead that language learning is promoted by interaction situations where learners understand the input they receive and are ready to adapt and modify their own output if necessary. The pedagogical approach of the interaction hypothesis has been characterized as

Focus-on-Form instruction (Long, 1998), where communicative tasks first focus on meaning and then draw attention to form.

Today, researchers and theoreticians argue for combined input- and output-based instruction (Song & Suh, 2008; Swain, 1985, 2005; Uggen, 2012), but the extent to which each should be present in teaching is still under investigation. In recent years, the role of various types of input and output in terms of different proficiency levels has also attracted the attention of researchers. As far as the findings are concerned, young learners have been shown to benefit more from input-based instruction (Hu, 2012; Shintani, 2011, 2013), which has been explained by their ability to retain words faster, but also by the fact that their lower level of proficiency may hinder their production. With increasing proficiency, however, output tasks have an increasingly important impact on developing language knowledge (Song & Suh, 2008; Uggen, 2012), but this might be a challenge because in comparison to processing input, learners need to make more mental effort to produce output (Huang et al., 2012).

An important theory from research on vocabulary acquisition is Laufer and Hulstijn's *Involvement Load Hypothesis* (2001), which argues that successful acquisition and retention of unfamiliar words depends on the learner's involvement in processing the words. In various studies, both enriched input (repeating target items to ensure several encounters) and enhanced input (emphasizing words in some way such as bolding, glossing or underlining) were shown to be more effective for L2 learners (Sonbul & Schmitt, 2010, 2013) than only teaching the definitions of words. Vocabulary research has shown us that rich engagement, repeated exposure, and the use of deeper-level processing lead to explicit learning, enhancing both the recognition and recall of words considerably (Dóczi & Kormos, 2016).

Output tasks have also been shown to effectively promote language learning, and compared to input tasks, those with a higher involvement load (i.e., writing tasks as opposed to simple gap-fill exercises) resulted in more

development (Huang et al., 2012; Song & Suh, 2012). Another important concept is the pushed output task, in which learners are required to productively use the language and create content or output with given grammatical or lexical items, thus putting their receptive knowledge into productive use. According to recent studies, having opportunities to produce output have resulted in better language performance than conditions relying on input (Song & Suh, 2008; Uggen, 2008).

Focus-on-Form (Task-based) vs. Focus-on-Forms Instruction

In terms of classroom task approaches, we can differentiate between Focus-on-Forms instruction (FonFs), which is characterized by the teaching of linguistic forms, and Focus on Form (FonF), where the initial focus is on meaning, and learners pay attention to form in subsequent communicative activities. While the former is guided by the principle of presentation-practice-production (PPP) and is often referred to as an isolated form-focused instruction or explicit, intentional learning, the latter is inherent in task-based teaching and is defined as integrated form-focused instruction (FFI), which is a characteristic of incidental learning (Laufer, 2006, 2009; Long, 1998; Shintani, 2013). Research is abundant about the usefulness of each learning condition, and the results are contradictory. However, recent findings have pointed to the long-term benefits of the task-based approach, in vocabulary research in particular (e.g., de la Fuente, 2006; Laufer, 2006; Laufer & Girsai, 2008). The authors highlighted the importance of negotiation of meaning and argued that focus on form instruction contributed to the production of more output than more traditional focus on forms (PPP) type of instruction.

Despite the benefits of task-based teaching and focus-on-form instruction, and the support they have received from researchers and educators alike, we cannot completely ignore the criticisms they have made. As Ellis (2017, 2018) pointed out, a major problem is that research on the topic has predominantly focused on tasks, task complexity and effectiveness,

and not actually on task-based teaching. Secondly, the problems of task authenticity and what counts as a task are still unresolved. Thirdly, while it might seem suitable for learners with clearly identified needs, task-based teaching has been less widely incorporated into state education and there is still much debate about pedagogical issues and how to integrate it into a class syllabus, let alone the national curriculum.

As a modified version of task-based teaching, which require tasks to be the basis of a syllabus, the concept of task-supported teaching has been introduced (Ellis, 2014), allowing for a more traditional linguistic syllabus, where the pre-defined linguistic items are practiced with tasks. Since in the Hungarian context there is an absolute lack of research on the concepts of input and output, as well as on Focus-on-form vs. focus-on-forms instruction or task-based-teaching, we wanted to focus on these aspects in our observation.

Methods

The Research Question and the Design of the Study

Based on the aim of the study and the literature reviewed above, we set out to answer the following research question: What are the instructional characteristics of the English classrooms in the Hungarian secondary schools we observed? To be more specific, we focused on the following sub-questions:

1. To what extent are features of task-based or task-supported teaching used?
2. What work modes are used in the classrooms?
3. What types of (input and output) tasks and activities are used in these language classes?

4. What characterizes language use in the classrooms?
5. To what extent is differentiation used in the classrooms?

In our research, a formal approach was taken with non-participatory researchers/observers using a systematic observation tool as a means of gathering data. The benefit of a systematic observation tool is that it allows data to be compared and systematically analysed. At the same time, it is rather challenging to create such a tool and to learn how to use it effectively. The instrument must be sophisticated but not overcomplicated, especially if they are to be used by several observers, as in our case.

To answer the research question, we employed an observation research strategy with a piloted instrument in a mixed-methods study. The instrument was developed for the purpose of the present study, and it was used to collect quantifiable qualitative data on classroom processes but also contained open-ended questions for the observer. We decided to include a list of certain characteristics, such as types of input and output, work modes, flow of information and language use, which meant that the focus of observation was predetermined. The quantifiable data were produced in a table format, where the number, frequency or timing of certain events were noted and counted. All categories were agreed on in advance by the research team, but the observation tool was semi-structured so that additional comments, explanations or details could be recorded.

Participants of the Study

The first step of the sampling was the purposeful selection of the participating schools in Budapest and in rural areas of Hungary. This article contains data collected from five schools in Budapest and three schools in rural areas, one from the Eastern and two from the Western part of the country. Altogether, 47 lessons of 19 teachers in Year 9–12 were observed by

six observers with a total of 553 participating students in 42 groups (see Table 1 for a detailed description of the years, the number of students and their levels, as well as the number of lessons per week).

Table 1

Information on Participants' Characteristics and the Observed Lessons

Year of study	Number of students	Number of lessons observed	Average number of students per group	Level of participants in the groups observed	Average number of lessons per week
9	168 (in 11 groups)	13	15	A1–A2; 5 groups A2–B1; 6 groups	5
10	223 (in 16 groups)	19	14	A2–B1; 2 groups B1–B2; 13 groups C1; 1 group	4
11-12	162 (in 15 groups)	15	11	A2; 3 groups B1–B2; 5 groups B2+; 7 groups	4
Total	553	47	13	A1–A; 10 groups B1–B2; 24 groups B2+; 7 groups C1; 1 group	4.5

As far as the levels are concerned, there were 10 groups at A1-A2 level (mostly 9th graders) and most students studied at B1–B2 level (24 groups).

We also saw that there were more advanced groups (mostly B2+ but in one case C1) in grades 11 and 12 (this accounts for 20% of the groups), which is a welcome trend. The average number of lessons is 5 for 9th graders and 4 for 10th, 11th and 12th graders. There were 15 students in a group in grade 9, 14 in grade 10, and 11 in grades 11 and 12. This decreasing tendency can be explained by the fact that an increasing number of students are taking language exams and achievement tests in order to avoid having to attend classes regularly.

The Instrument and its Pilot

The instrument used in our study was designed for the purpose of the present investigation. We set out to design a structured observation scheme that aimed to collect data on the characteristics of input from the teacher, student output (written/spoken, number of solutions, pushed output), form of instruction (frontal, pair- or group-work), flow of information (student giving/receiving information), and the language of instruction.

The instrument was piloted in several steps. The first drafts of the instrument were subjected to expert judgment, and several discussion sessions were held by the six participating researchers to finalize the pre-pilot version of the instrument (first version). Once we agreed on the preliminary format and content of the instrument, we observed four lessons, with two researchers each, to pilot the preliminary version of the instrument which became the second version after some changes and corrections were implemented.

It seemed that the original instrument was trying to cover too many aspects, was too long and difficult to fill in, so it had to be narrowed down and simplified. One aim was to describe the type of tasks (presence of input and output and its forms); another important aspect was researchability: quantitative data had to be supplemented with qualitative notes to make the results interpretable and triangulated. In light of the pilot results, both the

format and content of the instrument were fine-tuned by including clear rubrics for data to help the quantification process in data analysis as well as increase inter-rater reliability. The final version of the instrument was piloted by two researchers, who observed two more lessons. In this phase, only a few minor language-related changes were made as well as some changes in the display of pre-, during, and post-task activities, and the way notes were taken, which were also improved after a joint consultation, giving the tool its final form. The final version of the observation guide can be found in the Appendix.

It is important to point out that these lesson observations were not meant to evaluate or assess the teacher in any form, nor was it possible to document the learning process as such. What we tried to achieve was to report on the visible teaching processes, the types of tasks (both input and output) used, the forms and modes of instruction and language use, as well as some general observations about student–teacher and student–student interaction and group atmosphere, as perceived by the observer and analysed by the researcher. As we observed regular lessons embedded in teachers' timetables, it was not feasible to have an in-depth interview after the observed lessons, but we did take some time for a short informal discussion.

Although observer bias cannot be completely avoided, it is necessary to remain as objective as possible. According to Moyles (2002), observer bias occurs due to (p. 179):

1) selective attention: we often notice what we want to;

2) selective encoding: our expectations might make us judge a situation unconsciously and force us to take notes of these and not something else;

3) selective memory: if notes are used, the most memorable events will be remembered more vividly;

4) interpersonal factors: the presence of an outside observer raises the issue of detachment and personal likes and dislikes might hinder objectivity.

Since these problems might compromise the outcome, to address these issues, we reduced the number of features to focus on during the observations, trained observers to take notes during lessons, and instructed them to write what they saw without judgement.

Procedure and Data Analysis

Data collection took place in the school year of 2019/2020 from September until mid-March, the start of the nation-wide lockdown. Once permission was granted by the heads of schools, teachers and parents, intact language classes were visited. The training of observers ensured the quality of data collection. Observation was planned to be unobtrusive, and observers sat at the back of the classroom. The instrument was filled in in real time, and the summative comments were added at the end of the class.

Although it may seem obvious that data analysis is based on the intended research objectives, the sheer volume of data collected may cause digressions, which may lead to other potential findings. However, in our case, during the long piloting process, we tried to exclude overlaps in categories and narrowed down our concepts as much as possible. Nevertheless, there was a great amount of information gathered, which required systematic analysis. We worked manually with colour codes to categorize and order the data collected. Then, focusing on the predetermined and piloted concepts and categories, Atlas.ti was used as a computer-based analytical tool to organize the information better.

Results and Discussion

In what follows, the main results will be presented regarding instruction, modes of work, flow of information, input and output tasks, language use, and differentiation. As there is only one previous study in the Hungarian research context, we will rely on that (Nikolov, 1999) for comparison where possible.

Focus-on-form versus Focus-on-forms Instruction

One of the greatest challenges of analysing the observation data we collected was the division and comparison of the elements used by the teachers in these classes. To define task in our analysis, we applied Nunan's (1989) definition, which combines real-world tasks with pedagogical consideration with an aim for situational authenticity (Ellis, 2014). Nunan (1989) considered a task as "a piece of classroom work which involves learners in comprehending, manipulating, producing or interacting in the target language while their attention is primarily focused on meaning rather than form" (p. 10).

For comparability between the activities in the lessons observed, an activity/task was defined as a unit if (1) it was a real *task* (as in task-based learning, based on Nunan's (1989) definition), or (2) it focused on a topic or language area with activities to support the understanding of meaning (i.e., task-supported learning), or (3) a purely linguistic exercise (i.e., a grammatical or lexical explanation followed by a set of exercises). For example, creating a mind map was counted as a task, including the preparation, vocabulary development, and the actual execution. In another class, a grammar practice session with activities aiming to teach the conditional structure (starting with revision, some explanation followed by exercises) was also considered as a unit. In the results, we tried to calculate how many different such units were present in one lesson. As we can see, the average number was three in grade 9 and two for grades 10, 11, and 12. This

indicates that as students get more advanced theoretically and older, they spend more time on one type of activity.

Our findings indicate that although there were some lessons with solely linguistic agendas, we could see a shift towards meaning-focused instruction, where tasks are not only used for practicing pre-determined linguistic elements but also focus on meaning. This was true for almost half of the classes and, as expected, there were few examples of real-life tasks: eight out of the 47 lessons observed (e.g., preparing for, giving and assessing a TED talk; preparing to talk to a doctor about a health problem; negotiating information while creating a mind map on a topic).

Table 2

Number of Tasks in the Observed Lessons

Year of study	Number of lessons observed	Number of units (tasks or sessions of activities) altogether	Number of units/ lesson	Number of lessons based on focus on forms	Number of lessons based on task-supported learning	Number of lessons based on task-based learning
9	13	36	3	4	5	4
10	19	37	2	6	11	2
11-12	15	32	2	6	7	2
Total	47	105		14	22	8

Work Modes and Flow of Information

Table 3 presents the work modalities in the observed lessons. We can see that the most frequently used work mode was frontal, with the teacher taking the leading role. However, it must be noted that because the notes were taken in real time, every such incident was counted (even if it was an

instruction, a clarification or an explanation). This means that 50% of these instances were often very brief and lasted maybe a minute or two. Considering the duration, it can be concluded that frontal, individual and pair work modes were used to a similar extent. Compared to Nikolov's findings (1999), this is a welcome change. On the other hand, group work was still quite rare in the observed lessons, which might also have been due to the layout (difficult to carry out in the given room) or lack of time.

Table 3

Work Modes Used in the Observed Lessons

Year of study	Frontal	Individual	Pairwork	Groupwork (3+)
9	23	13	9	6
10	25	14	17	8
11 & 12	41	12	13	1
Total	89	39	39	15

As shown in Table 4, students in all grades gave and received information to a similar extent, and there were also opportunities for two-way information exchanges. This is not present in Nikolov's (1999) findings and can be seen as a positive feature because it is an important element of real-life communication, characteristic of communicative language teaching.

Table 4

Flow of Information

Year of study	Student gives information	Student receives information	Both
9	16	18	17
10	20	27	23
11 & 12	27	25	19
Total	63	70	59

Input

In our observation sheet, we wanted to find out what types of input were used. The results can be seen in Table 5. Most of the input was given by the teachers either verbally, written on the board, or using extra materials (mostly in the form of handouts with practice tasks). Unsurprisingly, the second most frequently used input was the course book, and there were only a few occurrences of student-initiated input. It seems that teachers felt more in control when they provided the content for the lessons.

Table 5

Types of Input Used

Year of study	Teacher-initiated	Student-initiated	Content-based	
			Coursebook or workbook	Extra material
9	12	6	16	6
10	10	5	17	16
11 & 12	10	4	25	24
Total	32	15	58	46

Output

In Table 6, various types of information are presented with regard to the output students produced in the observed lessons. First, in terms of spoken and written output, we found that as students' language proficiency increases, oral performance increases and written performance decreases accordingly (44% in grade 9, 41% in grade 10 and 27% in grades 11 and 12). On the positive side, students are doing more speaking tasks at a higher level as this may indicate that they have opportunities to become even more fluent. However, it would be interesting to further explore the ratio of

written and spoken output in more classes and at different levels as well as the consequences of reducing the amount of written output.

Most of the output generated by the students consisted of items that had only one correct answer or solution (100 instances), followed by tasks with open-ended solutions (55 instances). Interestingly, there were very few instances where the group or pair had to come to agreement (9), even though this is an important feature of task-based teaching: Using communication to agree or disagree is something that would be useful in real-life language exchange. In fact, agreement is an important aspect of communication tasks as it is supposed to enhance negotiation of meaning (Pica et al., 1993).

Table 6

Output

	Type of output		Number of solutions to a task			Content is		
Year of study	spoken	written	one	more	group has to agree	pre-defi-ned	based on personal experien-ce	based on pushed output
9	56% (24)	44% (19)	27	10	1	21	9	2
10	59% (29)	41% (20)	43	16	7	45	14	8
11 & 12	73% (48)	27% (18)	30	29	1	41	14	1
Total	64%	36%	100	55	9	107	37	11

There were some interesting findings regarding the type of content students created: In the vast majority of cases content was predefined, which means that there was little room for creativity, variation or deviation. There

was less opportunity to rely on personal experience, which is inherent in task-based teaching and real-life language use. An important shortcoming identified was that there were even fewer occurrences of pushed output, where the teacher asked the students to use certain lexical items or grammatical structures to produce output. The reason for this phenomenon needs further examination. Based on our experience in these classrooms, the type of output that students produced was rather monotonous and pre-defined, and this could be one of the reasons why some learners suffer in real-life situations and are (or feel) less able to communicate.

Language Use

In contrast to Nikolov's (1999) earlier findings, in the lessons we observed, teachers used English, and much of the communication between teacher and students was also in English. The majority of students used the target language on task; however, naturally, there were instances of Hungarian and some code-switching taking place. Based on our observations, when necessary, the mother tongue was used to clarify information and meaning, to make corrections and check solutions as well as to provide feedback. Students often resorted to their mother tongue in group discussions when they did not know how to express something or when the teacher did not seem to notice it. Several times, teachers encouraged their students to try to say what they wanted in English. This finding harmonizes with current tendencies which highlight the relevance and predominance of target language use while, at the same time, acknowledge the pedagogical and strategic role of the mother tongue in second language learning (Tammenga-Helmantel et al., 2020).

Differentiation

Differentiation was rare in the classes we observed: in all three age groups, we observed two cases. In two cases, the teacher explained a

language point to a student who had not previously known it, and in another, two boys (who were more advanced than the other members of the group) read graded readers while the rest of the group was working on something else. In the most mature age groups (grades 11 and 12), there were two occasions when students were allowed to select the material they were working on (a TED talk), and one occasion, a student was encouraged to start working on a new task while the others were still working. Despite the infrequency of differentiation, a few teachers said at the end of their lessons that they used differentiation from time to time when necessary. Another explanation given by the teachers was that differentiation was not necessary because they had spent a lot of time and energy creating homogeneous groups either at the beginning of their high school studies or at the start of the year. The differentiation aspect would also be interesting to observe in greater detail in the future.

Conclusion

This paper has reported on a lesson observation project with the aim of investigating various processes that take place in the language class. Our findings indicate that teachers used a great variety of input: mostly language-based activities and some real life communication tasks in English lessons. In terms of pedagogic approach, Focus on Forms type of instruction, where language forms rather than meaning are in focus, dominated over meaning-based teaching; however, in some cases, task-based or task-supported language learning was also used. In terms of modality, frontal, individual and pair work were detected to a similar extent, while group work was less common. The target language was dominant in the classes we observed with the mother tongue being used as a helping tool for the teacher and as an avoidance strategy for the students. In most classes, there was no differentiation, but this may also be explained by the participation of relatively homogeneous groups. Obviously, the findings raise several questions as to why certain phenomena were observable and how teachers

themselves view these processes. It also remains to be seen how more elements of task-based (or task-supported) teaching could be incorporated into secondary-level language teaching.

This study has several limitations. First of all, no generalizations can be made because only a small number of lessons were observed, and although we tried to observe more classes of each teacher participating in our study, two or three lessons were not enough to give a comprehensive picture of their teaching. Secondly, despite our commitment to improving the validity and reliability of our research tool and its use, we cannot disregard the fact that observation:

1) may change how a teacher would normally behave (even though most of our participants reported that the lessons observed were representative of what they normally do);

2) can only give an insight into teaching and cannot provide us with information about how much learning has taken place;

3) is also dependent on the observer and there are always bound to be differences between what various observers notice and note down, as well as how they note these events down.

A final issue is that of the researchers who cannot but interpret the data in light of their own experience and assumptions. We have tried to exclude this possibility as much as possible by adhering to the existing categories and made sure that both the observers and researchers clarified any misunderstandings to avoid possible misinterpretations. Nevertheless, no matter how much we tried to account for these potential flaws during the piloting and data analysis processes, these issues should be better acknowledged and tackled in the future.

In terms of future research directions, we think we need more observation data, which should be compared to student- and teacher-related data as well as with data on perceptions of classroom tasks and processes.

Teachers should be asked why they do what they do in follow-up interviews, and students could also inform the research processes by adding their viewpoints about how they see classroom processes.

That said, we think that our lesson observations have proven to be extremely fruitful and an important element of triangulation in our large-scale national study and have resulted in a valuable analytical observation tool that can be used by researchers and teachers alike. We are grateful to all the teachers who let us into their classrooms and allowed us to look into the processes of language learning even if it was only a snapshot. In the future, we hope to further develop our instrument and continue to investigate what actually happens in English classrooms with a special focus on tasks. It would also be important to collect data on how, when and why they are used, so that we can incorporate their importance into teacher training to ensure that our trainees become more familiar with both the theory and practice of classroom learning.

Notes

1. The present study was supported by the NKFIH–129149 research grant.

2. We sincerely thank the work of our colleagues participating in this project: Ágnes Albert, Ádám Lajtai, Katalin Piniel, Dávid Smid and Anna Zólyomi.

3. We would like to thank the school principals, teachers, and students for their participation in our project, especially for letting us into their classrooms. We would not have been able to do our work without their help.

4. Finally, we really appreciate the feedback of our reviewers that helped us to finalize our article.

References

Ahmed, E., Nordin, Z. S., Shah, S. R., & Channa, M. A. (2018). Peer observation: A professional learning tool for English language teachers in an EFL institute. *World Journal of Education*, *8*(2), 73–87. https://doi.org/10.5430/wje.v8n2p73

Allwright, D., & Bailey, K. (1991). *Focus on the language classroom: An introduction to language classroom research for language teachers.* Cambridge University Press.

Campbell, C. M. (2017). An inside view: The utility of quantitative observation in understanding college educational experiences. *Journal of College Student Development*, *58*(2), 290–299. https://doi.org/10.1353/csd.2017.0021

Chaudron, C. (1988). *Second language classrooms: Research on teaching and learning.* Cambridge University Press.

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

Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching, assessment.* Council of Europe. http://www.coe.int/t/dg4/linguistic/source/framework_en.pdf

De la Fuente, M. J. (2006). Classroom L2 vocabulary acquisition: Investigating the role of pedagogical tasks and form-focused instruction. *Language Teaching Research*, *10*, 263–295. https://doi.org/10.1191/1362168806lr196oa

Dóczi, B., & Kormos, J. (2016). *Longitudinal developments in vocabulary knowledge and lexical organization.* Oxford University Press.

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

Dörnyei, Z. (2019). Psychology and language learning: The past, the present and the future. *Journal for the Psychology of Language Learning*, *1*, 27–41.

Dörnyei, Z., Csizér, K., & Németh, N. (2006). *Motivation, language attitudes and globalisation: A Hungarian perspective*. Multilingual Matters.

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

Einhorn, Á. (2007). Az idegen nyelvi érettségi vizsga reformja. In I. Vágó (Ed.), Fókuszban a nyelvtanulás (pp. 73–105). Oktatáskutató és Fejlesztő Intézet.

Ellis, R. (1994). *The study of second language acquisition.* Oxford University Press.

Ellis, R. (2014). Taking the critics to task: The case for task-based teaching. *Proceedings of CLaSIC*, 103–117.

Ellis, R. (2017). Position paper: Moving task-based language teaching forward. *Language Teaching*, *50*, 507–526. https://doi.org/10.1017/S0261444817000179.

Ellis, R. (2018). *Reflections on task-based language teaching.* Multilingual Matters.

Elmendorf, D. C., & Song, L. (2015). Developing indicators for a classroom observation tool on pedagogy and technology integration: A Delphi study. *Computers in the Schools*, *32*(1), 1–19. https://www.learntechlib.org/p/160438/

European Commission (2012). *Europeans and their languages: Special Eurobarometer 386.* http://ec.europa.eu/public_opinion/archives/ebs/ebs_386_en.pdf. European Commission EUSTAT.

European Statistics (2016). *Education and training in Europe: Facts and figures* [Data set]. Eurostat. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Foreign_language_skills_statistics#Number_of_foreign_languages_known. European Commission EUSTAT.

Fletcher, J. A. (2018). Peer observation of teaching: A practical tool in higher education. *The Journal of Faculty Development*, *32*, 51–64. https://doi.org/10.13140/RG.2.2.19455.82084

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

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

Hu, C. F. (2012). Fast-mapping and deliberate word-learning by EFL children. *The Modern Language Journal*, *96*, 439–453. https://doi.org/10.1111/j.1540-4781.2012.01357.x

Huang, S., Eslami, Z. R., & Willson, V. (2012). The effects of task-involvement load on L2 incidental vocabulary learning: A meta-analytic study. *The Modern Language Journal*, *96*, 544–557. https://doi.org/10.1111/j.1540-4781.2012.01394.x

Kapitánffy, J. (2001). Az idegen nyelvek oktatásának fejlesztése a közoktatásban. Az oktatási Minisztérium fejlesztési stratégiája. *Iskolakultúra*, *8*, 71–74.

Krashen, S. (1981). *Second language acquisition and second language learning*. Pergamon.

Laufer, B. (2006). Comparing focus on form and focus on FormS in second language vocabulary learning. *Canadian Modern Language Review*, *63*, 149–166. https://doi.org/ 10.3138/cmlr.63.1.149

Laufer, B. (2009). Second language vocabulary acquisition from language input and from form-focused activities: Research timeline. *Language Teaching*, *42*, 341–354. https://doi.org/10.1017/S0261444809005771

Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics*, *29*, 694–716. https://doi.org/10.1093/applin/amn018

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language. *Applied Linguistics*, *22*, 1–26. https://doi.org/10.1093/applin/22.1.1

Long, M. H. (1985). Input and second language acquisition theory. In S. Gass & C. Madden (Eds.). *Input in second language acquisition* (pp. 377–393). Newbury House.

Long, M. H. (1998). Focus on form in task-based language learning. *University of Hawai'i Working Papers in ESL*, *16*(2), 35–49.

Medgyes, P., & Öveges, E. (2005). Paved with good intentions: Foreign language policy in Hungary. In C. Dobos, Á. Kis, Y. Lengyel, G. Székely, & S. Tóth (Eds.), *Mindent fordítunk és mindenki fordít–Értékek teremtése és közvetítése a nyelvészetben* (pp. 279–290). Bicske: SZAK.

Moskowitz, G. (1971). Interaction analysis: A new modern language for supervisors. *Foreign Language Annals*, *5*, 211–221. https://doi.org/10.1111/j.1944-9720.1971.tb00682.x

Moyles, J. (2002). Observation as a research tool. In M. Coleman & A. R. J. Briggs (Eds.), *Research methods in educational leadership and management* (pp. 172–195). Sage.

Nikolov, M. (1999). Classroom observation project. In H. Fekete, É. Major, & M. Nikolov (Eds.), *English language education in Hungary: A baseline study* (pp. 221–246). British Council.

Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge University Press.

Öveges, E., & Csizér, K. (2018). *Foreign language learning education in Hungary: A study into primary and secondary schools* [Vizsgálat a köznevelésben folyó idegennyelv-oktatás kereteiről és hatékonyságáról. Kutatási jelentés]. Oktatási Hivatal.

Öveges, E., & Kuti, Zs. (2016). *What are the problems of language teaching*? [Mi a baj a nyelvoktatással? Konferencia a hazai nyelvoktatás problémáiról]. Nyelvtudásért Egyesület.

Pica, T. (1987). Second-language acquisition, social interaction, and the classroom. *Applied Linguistics*, *8*(1), 3–21. https://doi.org/10.1093/applin/8.1.3

Pica, T., Kanagy, R., & Falodun, J. (1993). Choosing and using communication tasks for second language research and instruction. In G. Crookes & S. M. Gass (Eds.), *Tasks and second language learning* (pp. 9–34). Multilingual Matters.

Shintani, N. (2011). A comparative study of the effects of input-based and production-based instruction on vocabulary acquisition by young EFL learners. *Language Teaching Research*, *15*, 137–158. https://doi.org/10.1177/1362168810388692

Shintani, N. (2013). The effect of focus on form and focus on forms instruction on the acquisition of productive knowledge of L2 vocabulary by young beginner learners. *TESOL Quarterly*, *47*, 36–62. https://doi.org/10.1002/tesq.54

Sonbul, S., & Schmitt, N. (2010). Direct teaching of vocabulary after reading: Is it worth the effort? *English Language Teaching Journal*, *64*, 253–260. https://doi.org/10.1093/elt/ccp059

Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning*, *63*, 121–159. https://doi.org/10.1111/j.1467-9922.2012.00730.x

Song, M-J., & Suh, B-R. (2008). The effects of output task types on noticing and learning of the English past counterfactual conditional. *System*, *36*(2), 295–312. https://doi.org/10.1016/j.system.2007.09.006

Spada, N., & Fröhlich, M. (1985). *COLT Observation Scheme*. National Centre for English Language Teaching and Research.

Stallings, J. A., & Mohlman, G. G. (1988). Classroom observation techniques. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: An international handbook*. (pp. 469–474). Pergamon Press.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass, & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Newbury House.

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471–483). Routledge.

Tammenga-Helmantel, M., Mossing Holsteijn, L., & Bloemert, J. (2020). Target language use of Dutch EFL student teachers: Three longitudinal case studies. *Language Teaching Research*. https://doi.org/10.1177/1362168820911195

Uggen, M. S. (2012). Reinvestigating the noticing function of output. *Language Learning*, *62*(2), 505–540. https://doi.org/10.1111/j.1467-9922.2012.00693.x

Vágó, I. (1999). Az élő idegen nyelvek oktatása–Egy modernizációs sikertörténet. In I. Vágó (Ed.), *Tartalmi változások a közoktatásban a 90-es években* (pp. 135–173). Okker Kiadó.

Appendix: The final observation guide

School & grade: _____

Number of lessons per week: _____

Level of proficiency: _____

Date, time, which lesson: _____

Course book: _____

Number of students: _____

Task						Notes [detailed description of the task]: *teacher is marked if no course book or extra material is used
Time						
input	listening text	course book				
		extra material				
	reading text	course book				
		extra material				
	oral task	teacher*				
		student				
output1	spoken					
	written					
work mode	frontal					
	individual					
	pair work (2 students)					
	group work (3 + members)					
output2	number of solutions: 1/more					
	group has to agree					
information flow	student receives information					
	student gives information					
	student both receives and gives information					
	there is time for preparation (Y/N)					
output3	pre-defined content					
	own experience					
	pushed output					
language	language of instruction: E/H					
	language of execution: E/H					

NOTES ON THE LESSON:

The majority of students pay attention: Y/N, remarks:

The atmosphere is good: Y/N, remarks:

Teacher has a positive attitude towards the students: Y/N, remarks:

Teacher differentiates [i.e., tasks are adjusted to students' abilities]: Y/N, remarks:

Teacher is encouraging [i.e., students are encouraged to make more effort]: Y/N, remarks:

Question for the teacher after the lesson: How typical was this lesson?

1. Frontal: teacher-student interaction; Individual: Student executes task/exercise alone.

2. output3 explanation: "pre-defined content"–pics, texts given by the teachers and students need to work with this; "own experience"–While doing the task, students rely on their own personal experience, "pushed output"–students are free to create a text, but they need to include certain words/expressions or other linguistic/grammatical points.

5.

An Interview Study on Hungarian High School English Teachers' Beliefs about Learning English

Dávid Smid
smiddav@yahoo.com

Anna Zólyomi
zolyomi.anna@btk.elte.hu

Abstract

Investigating deep-rooted beliefs is increasingly gaining prominence in the field of applied linguistics. After the change of focus from teacher-centeredness to learner-centeredness, several studies examined language learners' beliefs about themselves (lay theories); however, studying the beliefs of the language teacher, an important stakeholder in the language learning process, was rather relegated to the background. Consequently, the aim of this study is to broaden our understanding of language teachers' beliefs about language learning. To this end, we conducted an exploratory qualitative study with 10 Hungarian high school English teachers. We used qualitative content analysis and the constant comparative method to arrive at emerging themes and recurring patterns. Our main findings suggest that, besides learner-internal factors, the social context also has a vital role in language learning. More specifically, based on our data, we distinguished four specific social contextual layers that were believed to have an impact on English language learning. Pedagogical implications are discussed with respect to pre-service foreign language (L2) teacher education. In both data collection and data analysis, we paid special attention to ethical considerations and quality control.

Keywords: teachers' beliefs, beliefs about English language learning, exploratory study, interview, Hungarian English teachers

An Interview Study on Hungarian High School English Teachers' Beliefs about Learning English

Exploring individual differences (IDs) has a long research tradition in the field of applied linguistics (Dörnyei & Ryan, 2015). As to the study of IDs, a considerable number of inquiries focus on foreign language (L2) learning-related beliefs subsuming L2 learners' beliefs about L2 learning from various aspects (e.g., Kalaja & Barcelos, 2003). However, little research is centered on investigating L2 teachers' beliefs about L2 learning. More importantly, to our knowledge, there has been no inquiry into Hungarian English teachers' beliefs about English language learning (ELL). Studying English teachers' beliefs is relevant and conducive to success in L2 learning and teaching because, as previous studies have shown, teachers' beliefs can shape their pedagogical practices (Borg, 2003). Consequently, the aim of the present study is to understand English teachers' beliefs about ELL in Hungary. The emphasis is on the English language owing to the globalization and spread of English as a lingua franca (Csizér & Illés, 2020; Jenkins, 2007), a result of which is that learning English is a different process to that of any other languages.

In this paper, first, we provide a theoretical framework to L2 teachers' beliefs. After that, we present the details of the qualitative research design employed in an attempt to answer our research question. Following this, we elaborate on the findings in light of previous studies on L2 teachers' beliefs. Pedagogical implications as well as ideas for further research are also discussed at the end of our paper.

Theoretical Background

Within the field of foreign language teaching, the study of teacher beliefs is most commonly associated with a broader area of inquiry, namely, teacher cognition (Borg, 2003). The latter notion stands for "the unobservable cognitive dimension of teaching–what teachers know, believe, and think"

(Borg, 2003, p. 81). The cited definition suggests that the various cognitive processes involved in teachers' professional lives are interconnected. An implication of the intricate nature of L2 teachers' cognitive system is that researchers need to distinguish their target construct from other seemingly similar ones as much as possible.

Concerning the central construct of our study, *belief*, it can be defined as "an individual's judgment of the truth or falsity of a proposition" (Pajares, 1992, p. 316). Typically, it is the psychological construct, *knowledge*, to which belief is compared. According to Pajares (1992), "belief is based on evaluation and judgment; knowledge is based on objective fact" (p. 313). Therefore, beliefs are more subjective, disputable, emotionally loaded, and less dynamic than knowledge (Abelson, 1979). Given the deep-rooted nature of beliefs, they are assumed to exert an influence on an individual's ways of knowing; what is more, the impact of beliefs can even manifest itself in how one interprets the information coming from one's environment, that is, in one's perceptual processes (Pajares, 1992). In sum, beliefs are powerful mental events that underlie psychological functioning and, thus, human behavior (Bandura, 1986; Pajares, 1992). Consequently, it can be argued that teachers' beliefs about ELL encompass their deep-rooted convictions about the overall L2 learning process.

Based on our review of the literature, it seems that there has been limited attention directed at L2 teachers' beliefs about L2 learning, our research focus. The results of the studies in question reveal several aspects of the L2 learning process to which L2 teachers ascribe value. These can be grouped into learner-internal and learner-external dimensions. As to the former, a recurring result across different contexts has been that the roles of learner autonomy and motivation are indispensable in L2 learning (Borg & Al-Busaidi, 2011; Borg & Alshumaimeri, 2019; Gkonou et al., 2016). In addition, many of the participants in Gkonou et al.'s (2016) research believed that willingness to communicate and emotions were also important psychological processes involved in L2 learning. These results are in line

with some of the most prominent ID variables that have emerged in connection with the study of L2 acquisition (Dörnyei & Ryan, 2015).

With respect to L2 teachers' beliefs about the learner-external aspects of L2 learning, studies have generally pointed to the role of the pedagogues themselves. This has manifested itself via a belief in the importance of classroom management (Borg & Alshumaimeri, 2019) and various motivational strategies (Dörnyei & Csizér, 1998), such as being a role model, creating a positive classroom atmosphere, and making classes interesting with respect to the attainment of success in L2 learning. Although there have been some theoretical models in connection with L2 learning motivation that have highlighted the role of L2 teachers in the L2 learning process (for an overview, see Dörnyei & Ushioda, 2013), there have been relatively few attempts at investigating the relationship between L2 learner and L2 teacher psychology (Dörnyei & Ushioda, 2013; Gkonou et al., 2016).

Besides the role of teachers, some other contextual aspects have also emerged in relation to L2 teachers' beliefs about L2 learning. In particular, there were participants across the reviewed studies who believed that the influence of the school setting (Borg & Al-Busaidi, 2011; Borg & Alshumaimeri, 2019) and the national culture (Borg & Alshumaimeri, 2019; Gkonou et al., 2016) were also defining in the quality of the L2 learning outcome. These results suggest that L2 teachers' beliefs are strongly linked to their perceptions of the broader social context in which L2 learning is embedded. What is more, in light of complex dynamic systems theory (CDST; Lowie & Verspoor, 2007), we can assume that in-service L2 teachers' awareness of the impact of the environment on L2 learning might even be more elaborate and multi-layered. Thus, given the complexity and context-sensitivity of L2 teachers' beliefs, their understanding warrants the utilization of in-depth accounts (Dörnyei, 2007). However, it seems that there is a dearth of studies on the given research topic that have relied solely on the qualitative research design. As such, more research with an exploratory focus is needed.

Research Context

Given the context-sensitivity of L2 teacher cognition (Borg, 2003), we find it necessary to provide a brief introduction to the institutional L2 learning context in Hungary. In terms of policies, each public educational institution has its own local curricula for each of the school subjects taught. These include information about the different aspects of instruction, for example, the content of a course, the course requirements, and assessment methods. By definition, all schools must follow governmental regulations [see *Decree no. 5/2020.* (I.31.), 2020] upon creating their local curricula. By way of illustration, *Decree no. 5/2020 (I.31.)* (2020) prescribes that the goal of L2 instruction must be the development of competences associated with L2 learning, such as communicative competence, learner autonomy, and intercultural competence. This competence-based approach to formal L2 learning corresponds to the ideas promoted by the Council of Europe (2001) regarding the L2 learner/user.

Unfortunately, we could not find any recent research on the degree to which the criteria set by policymakers [Decree no. 5/2020. (I.31.), 2020] are observed in English language instruction in Hungarian public schools. This can be attributed to the fact that it has only been a year since higher-level policy changes took place. Nevertheless, the results of a large-scale investigation about L2 instructional practices in Hungary conducted in 2017 (Öveges & Csizér, 2018) might still be worth considering. They shed light on a number of issues. One is the L2 class size: Öveges (2018) found that a typical L2 class consisted of 11–15 pupils in grades seven and 11, which corresponds to the given governmental decree; however, several reports (Illés & Csizér, 2018; Kálmán & Tiboldi, 2018; Tartsayné Németh et al., 2018) showed that the various stakeholders involved in instructed L2 learning in Hungary (i.e., school leaders, L2 teachers, L2 consultants) tended to be of the opinion that the size of L2 classes should be smaller. Second, there was a relatively large proportion of schools in Öveges's research where pupils were grouped in alphabetical order. Based on Scrivener (2011), failing to stream

students according to their L2 proficiency level can detract from the efficiency of L2 instruction. Finally, another issue consistently highlighted by many of the school leaders, L2 consultants, and L2 teachers in a number of inquiries (Illés & Csizér, 2018; Kálmán & Tiboldi, 2018; Tartsayné Németh et al., 2018) was the lack of information and communications technology (ICT) equipment of schools. This is unfortunate given that one of the aims of instructed L2 learning is supposed to be the development of pupils' digital competence [Decree no. 5/2020. (I.31.) (2020)]. Based on the studies reviewed above, we can assume that Hungarian L2 teachers' perceptions of these instructional conditions are likely to be intricately linked to their beliefs about L2 learning, the focus of our study.

Research Methods

In light of the research niche presented, we have formulated the following research question: What beliefs do Hungarian high school English teachers hold about English language learning? In this study, the qualitative method is applied because it is the most suitable to make sense of the context-sensitivity (Dörnyei, 2007) of psychological processes. In addition, the qualitative research design allows us to broaden our understanding of the research focus (Dörnyei, 2007), which has been a neglected area within Hungary.

Participants

The participants, altogether 10 Hungarian high school English teachers from Budapest (n = 6), Tatabánya (n = 2), Miskolc (n = 1), and Nagykanizsa (n = 1), were selected through convenience and snowball sampling (Dörnyei, 2007). At the time of the data collection, they were teaching in six different high schools. Female participants are overrepresented (nine females and one male), and their age ranged between 27 and 64 (M = 46.10, SD = 11.23). On average, the participants have 20.80

years of teaching experience with a high standard deviation (*SD* = 9.50) and 20.20 years of English language teaching experience (*SD* = 9.30). This slight difference might be attributed to the fact that some teachers did not start with teaching English specifically. At the time of data collection, they were not teaching any other foreign languages besides English. They all have university degrees, and one person obtained a PhD degree as well. All of the interviewees have taught in high schools, seven have taught in language schools in addition to their high school teaching, but only three of them have the experience of having taught in a primary school. At the time of the data collection, most of them were teaching all grades of secondary education; that is, they reported that they have experience in teaching students between 14 and 20 years of age.

Research Instrument

It must be highlighted that this study is part of a fairly large nationwide project, which is intended to examine IDs from various perspectives (Csizér et al., 2021). For this specific study, we opted to choose one aspect; the one that seemed not only to be suitable for investigation but also key in understanding English teachers' beliefs in the Hungarian context. A semi-structured interview guide format was selected due to its practicality in directing the line of thought with pre-set questions but still giving freedom to the interviewee to add anything that they consider relevant (Wallace, 1998). The interview guide had been piloted after which slight modifications were implemented and further questions were added. Besides the first part where we asked about background information, the interview guide consists of three main areas, namely, experiences related to language classes, beliefs about the language learner and the language teacher, and beliefs about out-of-class L2 acquisition. For the final Hungarian version of the interview guide, see our webpage (http://nyelvtanulas.elte.hu/).

Data Collection Procedures

Data was collected online as well as face-to-face before the COVID-19 pandemic with the help of our research group in the mother tongue of the participants (i.e., Hungarian) from October 2019 to February 2020. Originally, these in-depth interviews are from the aforementioned larger nationwide project, and for this specific study the aim was to achieve maximum variation in the data involving so-called "extreme" cases as well; in other words, we scanned the interview transcripts and selected teachers with various reported beliefs. For this purpose, we have chosen 10 interview texts with which we reached data saturation for the aim of this study. The interviews were 47.60 minutes long on average (*SD* = 15.46). The research group recorded the interviews with the participants' consent using Dictaphones and smartphones after which the recordings were transcribed verbatim manually by the second author. The 10 interviews comprised a rich corpus of approximately 57,360 words excluding the interviewers' contribution.

Data Analysis

In the first coding phase, we used initial coding independently. We used thematic content analysis (Dörnyei, 2007; Elo et al., 2014) in which we were searching for underlying patterns and recurring themes in the data. We employed three coding methods, labelled by Saldaña (2013) as process, emotion, and values coding. With the first coding method, we were looking for practices reported by L2 teachers; with the second, we targeted their reported affective reactions; and with values coding, we sought to identify their beliefs. The rationale behind choosing to do so was to gain a clearer picture of the complexity of the teachers' mental processes and behaviour. Following this, we finalized the coding together. After discussion, we arrived at approximately 368 codes, but it must be noted that one excerpt of interview text usually received more than one code. Finally, we grouped the codes according to their common features and arrived at definitions for each

category with the use of the constant comparative method (Dörnyei, 2007). It must be noted that the quoted excerpts below are the authors' translations.

Research Ethics and Quality Control

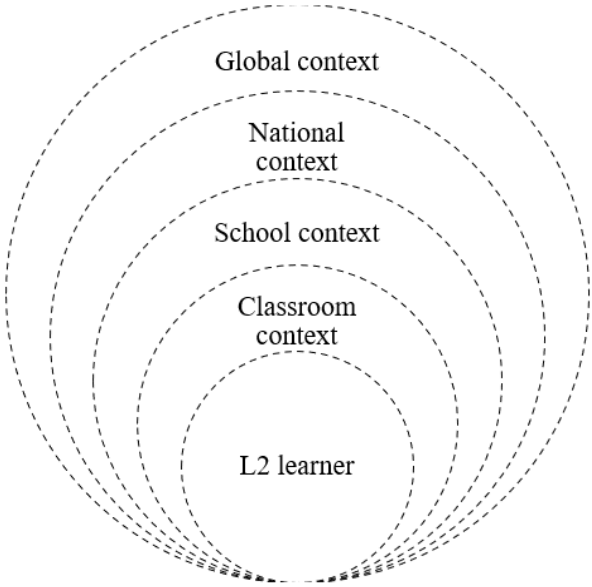
Throughout the study, we paid special attention to ethical considerations and quality control measures. Participation in this research project was voluntary, we asked interview participants for their consent to record the interviews, we retained their anonymity (using pseudonyms), and kept the data confidential. The interview guide had been piloted and was finalized after minor modifications. We used peer debriefing (i.e., analytic triangulation) to ensure good quality analysis (Dörnyei, 2007).

Research Findings and Discussion

Figure 1 shows the visual framework which we created during the data analysis. It helped us to answer our research question as well as to understand our findings better. As can be seen, we conceptualize the participants' beliefs about ELL as a multi-layered system, in which the different layers represent the L2 learner and the surrounding social context. Such an understanding of the composite nature of L2 learning is in line with the framework of the Douglas Fir Group (2016). In the following subsections, we discuss the emerging themes (i.e., the system layers) one by one, starting with our definitions of them, and then we provide quotes with pseudonyms to let the participants' voices come through.

Figure 1

The Five Layers of the Interviewed L2 Teachers' Belief System Concerning ELL



First Layer: Beliefs about the L2 Learner

Based on our findings, we conceive of the core of the participants' belief system as a layer consisting of the teachers' beliefs about the L2 learner. More precisely, we identified their convictions regarding the attributes of a successful L2 learner to belong here. The overwhelming majority of our sample reported that L2 ability has some role in L2 learning achievement; however, a L2 learner also needs to invest effort in learning in order to succeed. Such a belief was especially apparent when several teachers emphasized that learning a L2 does not stop when one obtains an advanced language proficiency exam but requires continuous self-development or when they shared how they react to students who are not so hard-working: "when I see that there are some [pupils] who are not working, I tell them that 'you won't get ahead if you don't try'" (Daniella). The emergence of the IDs

in question is in line with research in the field of L2 acquisition, according to which L2 aptitude and L2 learning motivation (referred to as L2 ability and effort investment above, respectively) are the most important determinants of success in L2 learning (Dörnyei & Ryan, 2015).

Besides L2 aptitude and L2 learning motivation, we identified a broad range of other learner characteristics that the interviewees believed had a role in successful L2 learning. First, learner autonomy manifested itself in statements such as Botond's: "the successful L2 learner is someone who is able to [...] do something [...] by oneself [...] not necessarily related to classwork." This quotation underscores the L2 learner's ability to seek new opportunities to practice the L2, a conceptualization that corresponds to Holec's (1981). Second, some teachers acknowledged the importance of self-regulation (Dörnyei & Ryan, 2015). As Daniella argued, such an ID was indispensable in learning situations characterized by strong external demands such as exams. Next, linguistic self-confidence also emerged as a potentially significant L2 learner attribute. Janka's conceptualization is in line with Clément's (1980) definition of linguistic self-confidence in that she links it to communicative competence and a lack of anxiety: "in my opinion, the successful and lucky one is someone who dares to speak up, who, let's say, is not afraid of making grammar mistakes, who doesn't care about one's pronunciation not being perfect enough." Fourth, several participants' statements reflected the presence of an underlying belief that the emotional dimension of L2 learning should not be underestimated regarding L2 learning achievement. To quote Amália, "This promotes them or impedes them from going forward. I think they can already see that this [ELL] is a necessary thing, so the best would be to do it with enjoyment." Such an understanding regarding the role of emotions in L2 learning supports prior empirical research. For example, MacIntyre and Vincze (2017) found positive emotions and motivation-related variables to be positively correlated, while negative emotions and the different aspects of L2 learning motivation to co-occur negatively in the case of L2 learners. In sum, it can be argued that our

sample members' convictions about the attributes of the successful L2 learner correspond both to research advancements within the field of L2 acquisition (Dörnyei & Ryan, 2015) as well as to the beliefs of in-service L2 teachers in other foreign contexts reviewed above (e.g., Borg & Al-Busaidi, 2011; Borg & Alshumaineri, 2019; Gkonou et al., 2016).

Second Layer: Beliefs about the Influence of the Classroom Context

The second layer of our framework concerns the influence of the classroom on ELL. More specifically, it includes the interviewed teachers' beliefs regarding the role of the teacher in ELL. This layer emerged with three patterns. Most interviewees mentioned that flexibility in planning and in being adept with students' needs is of key importance with respect to L2 learning success:

> So I really need to be flexible [...], so I tend to take such preferences into account as well, so let's say that I'm explaining a difficult grammar aspect, then, to grab their attention–just like when we give medicine with something yummy–then let's say that in the sports group, we practice it with a topic related to sports, so what they are interested in. (Edit)

We interpreted such statements as indicators of an underlying belief that learners' investment in L2 learning is contingent on the instructor's use of motivational strategies, a finding that is echoed by Dörnyei and Csizér's (1998) study.

Another notable aspect in relation to the teachers' role in L2 learning concerns failures/mistakes. The interviewed teachers are convinced that making mistakes should be seen as the pathway towards mastering the language and not an obstacle hindering the process. To quote Daniella:

> Well, I think they learn more easily if they have positive experiences. If they are constantly afraid during the lessons or this blocks them in

> some way, they will not be able to move forward, they will not be able to improve. So if I were to communicate that we shall not commit mistakes here, or if we do make mistakes and say: "Good Lord, are you saying that you made a mistake like that?", then from then on, it would be so disruptive for them that they could probably not improve much in my lessons. I always try to affirm that "good, you made mistakes, do make mistakes, I also make mistakes because I'm not a native speaker." (Daniella)

The quoted excerpt does not only reiterate the aforementioned facilitating effect of positive emotions and the debilitating effect of negative emotions on the learning process, but it also suggests a conviction that teachers can have a mediating role in L2 learner emotions by their handling of failures/mistakes. Based on the beliefs of the participants, if teachers focused on mistakes in a negative or debilitating way (as is portrayed by Lou & Noels, 2019), it would hinder the learning process. Thus, as illustrated in Daniella's account, emphasizing that making mistakes is accepted in the teachers' eyes is of key importance in fostering students' L2 learning achievement.

Finally, the third emerging aspect of the classroom layer in our framework relates to the role of L2 teacher self-efficacy in ELL. This pattern was most apparent in Margit's account, who is aware that using technology and digital features in the L2 learning classroom as a means of implicit learning (i.e., unconscious learning; for a review, see Rebuschat, 2015) would be beneficial for students. However, apparently, she has low self-efficacy in respect to digital skills, a concomitant of which is that she experiences anxiety during her classroom work:

> It's obvious that they would like tasks where they could use their phones, my colleagues do this… I don't dare to start this because I don't know technology very well, and I always become embarrassed as I often can't even find my pendrive and I become anxious. I shouldn't, I know that I must confess that "kids, I'm already… it is

> already praiseworthy that I got it this way, but somehow it really bothers me." So, I can no longer do things with the phone that–I don't know–tests and applications. No, no… I don't think we have ever done that, no, no. Yet, it is surely good, and it surely has an impact. (Margit)

The interviewees' belief that L2 teacher self-efficacy is linked to L2 acquisition is in line with prior empirical research in applied linguistics. More specifically, in Mifsud's (2011) multi-perspective study it was found that the higher the L2 teacher's self-efficacy, the stronger the L2 learners' motivation to learn the L2. Finally, Margit's case also underlines the need for developing in-service L2 teachers' pedagogical expertise in Hungary; the same call has been voiced by several recent research reports (Illés & Csizér, 2018; Kálmán & Tiboldi, 2018; Tartsayné Németh et al., 2018).

Third Layer: Beliefs about the Influence of the School Context

In the third layer of our framework, we grouped teachers' beliefs related to the school context. This subsumes beliefs about the influence the school has on the L2 learning process. A frequent pattern associated with this emerging theme was the role of grouping practices in students' L2 learning. The method of student grouping in L2 classes, as Öveges's (2018) report highlights, can vary from school to school in Hungary, hence our inclusion of it in this layer.

Many teachers treated (or wished to treat) the whole classroom as one homogeneous group when it came to the pupils' level of English proficiency. Statements such as Margit's were recurring in our data set: "the ideal would be banded teaching, then the groups would not be so mixed, we all struggle with this, we struggle terribly, because honestly, every day, bringing extra tasks always to this and that student is basically impossible" (Margit). Consequently, as illustrated by the quoted excerpt, most L2 teachers in the selected sample perceive differentiation as a challenging action and believe

that homogeneous groups are beneficial. Such a belief is in line with Scrivener (2011), who argued that grouping based on proficiency levels is indispensable for efficient language learning. The fact that differentiation is seen by L2 teachers as rather demanding is also concluded in Öveges and Csizér's (2018) research report. Further, our finding lends support to Öveges's (2018), who found that, in relatively many cases, pupils are not grouped according to placement tests in L2 classes in Hungary.

Fourth Layer: Beliefs about the Influence of the National Context

Based on our data, we detected the existence of a fourth layer in the belief system of our participants. This relates to the influence of the national context on ELL. Our respondents' accounts reflected the conviction that the educational system as well as educational policies have a role in ELL. We came to this observation based on their statements about the effectiveness of the different facets of the current Hungarian public school system.

This theme emerged with two patterns, both of which had a negative connotation. More precisely, the interviewed participants tended to complain about the large amount of content a teacher should teach according to the National Core Curriculum and the relatively few contact hours. These issues appeared through the respondents referring to them as constraints in their work by using the modal verbs, "have to" and "should". Thus, it is not surprising that the teachers also reported experiencing negative emotions in their work. Anikó's account illustrates the struggle in question well: "I often tell the kids that 'Don't believe that we enjoy what we do, I mean, us, teachers. I mean, we see the faults of the whole system very well, and we cannot do at all what we would really want–to make you like the subjects'." The main source of her struggle seems to be a strong conviction that L2 learners can succeed better provided the right macro-conditions are given. As such, this quotation also seems to suggest that the teachers' beliefs about the impact of the national educational context were intricately linked to their

beliefs about pedagogical practices as well as about the nature of L2 learning ability. The reported finding regarding L2 teacher frustration is not unique to our research setting as it has been reported in connection with socio-political constraints in other contexts as well (e.g., Nagamine, 2018). Moreover, our research outcomes are in line with Tartsayné Németh et al.'s (2018) in which school-leaders also reported problems with the number of lessons and the taught content (i.e., textbooks) when it comes to the effectiveness of instructed L2 learning in Hungary.

Fifth Layer: Beliefs about the Influence of the Global Context

There emerged another theme in our data, which we labeled global context. It entails the participants' beliefs about the influence of English as a global language on learning English. Given the ubiquity of the language, we conceive the global context as the broadest layer within the teachers' belief system concerning ELL. Two main patterns emerged in relation to this layer: One concerns the role of implicit learning environments, the other the roles of positive emotions and motivation related to the use of digital media.

Nearly all members of our sample argued that L2 learners should take advantage of the input available to them in their surroundings. Interestingly, this does not only entail the instructional environment, but also what we have labeled as an implicit learning environment. We refer to it as the engagement in an activity involving the use of the L2 outside school and learning something new in English without consciously realizing it. Naturally, this is linked to the ubiquity of the English language, especially to the use of digital media. Many participants, including Anett, talked about it in connection with L2 learners' personal interests, thereby indicating their belief that being in an implicit learning environment is most likely to be effective when it is accompanied by the L2 learners' experiencing positive emotions:

> We always encourage them–or I encourage them–to switch to English television, there are already lots of things [...] or since they're already watching YouTube videos and they're watching […] the makeup influencer, then they should watch it in English because they pick up a lot of things from that as well. So that's it, so that's how they should try to get as much input as possible for themselves. (Anett)

Strongly connected to the previously mentioned implicit learning environment, the role of positive emotions and particularly enjoyment in enhancing motivation was also apparent as highlighted by a L2 teacher:

> Why is it important? Well, because this is what keeps them going further or it hinders them from going forward on that path. I think they all see it already that this is a necessary thing, and then it would be good if they did it with enjoyment. And maybe nowadays most kids have the inner drive that they have to do this, and by having so many things available on the net, they can watch movies, videos, they have the motivation. And it's important, of course, because the more motivated they are, the more successful they are. (Amália)

Most interviewees in this study mentioned that they feel lucky that they are teaching English due to the role of English language use in L2 learning motivation in today's world. This may be linked to our second layer, the classroom context, because if teachers feel lucky to teach English, they are presumably more motivated to teach it. By this token, students also seem to be more motivated to learn English regarding its practical use in real life:

> Well, I think they are mostly enthusiastic about English, so they are interested in it, they see the practical benefits of it, they can practically immediately use what they have learned, and I think this has a serious motivating power. (Jolán)

The importance of English as a global language is also highlighted by Csizér and Illés (2020). They claimed that due to the ubiquity of English as a lingua franca, the focus is on language use rather than simply on L2 learning.

Conclusion

In this paper, we have presented our findings regarding Hungarian high school English teachers' beliefs about learning English, which we gained by employing a qualitative research design. The main outcome of our study is a framework (Figure 1). We created it based on our emerging themes, and it is in line with the Douglas Fir Group's (2016) understanding of the composite nature of L2 learning and L2 teaching. Our data indicated that the interviewed participants' beliefs formed a multi-layered system, with beliefs about the L2 learner being at the core, followed by the classroom context, the school context, the national educational context, and finally the global context. We have argued that the teachers' beliefs about the L2 learner (the core of the system) and the teachers' beliefs regarding ELL in the different contexts (classroom, school, national, global) are in an intricate relationship. As to the content of the beliefs in question, we could relate our findings to other studies conducted in Hungary (e.g., Csizér & Illés, 2020; Öveges & Csizér, 2018; Tartsayné Németh et al., 2018) as well as abroad (e.g., Borg & Al-Busaidi, 2011; Borg & Alshumaimeri, 2019; Gkonou et al., 2016; Nagamine, 2018). Therefore, it can be argued that adopting a context-sensitive approach and distinguishing among the different layers of L2 teachers' belief system is an optimal way to identify critical issues (e.g., low L2 teacher self-efficacy, L2 teacher frustration, unfavorable student grouping practices) related to L2 learning and L2 teaching, and thus to further research within our field.

The pedagogical relevance of our findings is seen primarily in pre-service L2 teacher education. Based on what we have found, it can be stated that teachers perceive differentiation as a challenging action; consequently, workshops for teacher trainees targeting efficient differentiation methods could alleviate their low self-efficacy beliefs directed at differentiation. A second implication concerns the role of taking IDs into account to address student diversity, and thus to provide a possibility to enhance the efficiency of ELL. Professional developments could also introduce pre-service L2

teachers to ways of exploiting the whole range of digital opportunities in the 21st century. In addition, as is apparent from the interviewed teachers' accounts, treating mistakes as steps towards mastering the language plays a pivotal role in the success of ELL.

Finally, it is important to stress that this study also has its limitations. For a more complete understanding of the L2 teachers' belief system, their enacted beliefs could also be examined. However, due to the COVID-19 pandemic, doing follow-up research with our participants in the form of classroom observations was not an option. As such, our findings could have been affected by the respondents' social desirability bias (Dörnyei, 2007). Another shortcoming of our research can be that we examined L2 teacher beliefs at one point in time only, which cannot account for their dynamic character suggested by some researchers (e.g., Borg, 2017; Kalaja & Barcelos, 2003). Thus, for further research, we recommend examining L2 teachers' ELL-related beliefs through a CDST framework (Lowie & Verspoor, 2007) with a longitudinal research design. An additional topic on the research agenda could be examining possible discrepancies within the belief system of an individual via a case study.

Acknowledgements

We would like to thank our research group for their help with collecting the data, Kata Csizér for always encouraging us to keep moving forward with our research project, as well as the two anonymous reviewers.

Note: This study was supported by the National Research, Development and Innovation Office in Hungary (NKFI-6-K-129149).

References

Abelson, R. P. (1979). Differences between belief and knowledge systems. *Cognitive Science*, *3*(4), 355–366. https://doi.org/10.1207/s15516709cog0304_4

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

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

Borg, S. (2017). Teachers' beliefs and classroom practices. In P. Garrett & J. M. Cots (Eds.), *The Routledge handbook of language awareness* (pp. 75–91). Routledge.

Borg, S., & Al-Busaidi, S. (2011). Teachers' beliefs and practices regarding learner autonomy. *ELT Journal*, *66*(3), 283–292. https://doi.org/10.1093/elt/ccr065

Borg, S., & Alshumaimeri, Y. (2019). Language learner autonomy in a tertiary context: Teachers' beliefs and practices. *Language Teaching Research*, *23*(1), 9–38. https://doi.org/10.1177/1362168817725759

Clément, R. (1980). Ethnicity, contact and communicative competence in a second language. In H. Giles, W. P. Robinson, & P. M. Smith (Eds.), *Social psychology and language* (pp. 147–154). Pergamon.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.

Csizér, K., Albert, Á., Smid, D., & Zólyomi, A. (2021). *Motivation, autonomy, and emotions in foreign language learning: A nation-wide multi-perspective investigation in Hungary* [Manuscript in preparation]. Multilingual Matters.

Csizér, K., & Illés, É. (2020). Helping to maximize learners' motivation for second language learning. *Language Teaching Research Quarterly*, *19*, 19–31. https://doi.org/10.32038/ltrq.2020.19.02

Decree no. 5/2020. (I.31.) on national core curriculum. (2020). *Magyar Közlöny*, *17*, 290–446.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies.* 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. https://doi.org/10.1177/136216889800200303

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

Dörnyei, Z., & Ushioda, E. (2013). *Teaching and researching motivation* (2nd ed.). Routledge.

Douglas Fir Group. (2016). A transdisciplinary framework for SLA in a multilingual world. *The Modern Language Journal*, *100*(Supplement), 19–47. https://doi.org/10.1111/modl.12301

Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K., & Kyngäs, H. (2014). Qualitative content analysis: A focus on trustworthiness. *SAGE Open*, *4*(1), 1–10. https://doi.org/10.1177/2158244014522633

Gkonou, C., Mercer, S., & Daubney, M. (2016). Teacher perspectives on language learning psychology. *The Language Learning Journal*, *46*(4), 501–513. https://doi.org/10.1080/09571736.2016.1172330

Holec, H. (1981). *Autonomy and foreign language learning*. Pergamon.

Illés, É., & Csizér, K. (2018). A nyelvtanárok válaszai [The language teachers' responses]. In E. Öveges & K. Csizér (Eds.), *Vizsgálat a köznevelésben folyó idegennyelv-oktatás kereteiről és hatékonyságáról: Kutatási jelentés* (pp. 161–188). Hungarian Educational Authority.

Jenkins, J. (2007). *English as a lingua franca: Attitude and identity.* Oxford University Press.

Kalaja, P., & Barcelos, A. M. F. (Eds.). (2003). *Beliefs about SLA: New research approaches*. Springer.

Kálmán, Cs., & Tiboldi, T. (2018). A szaktanácsadók válaszai [The consultants' responses]. In E. Öveges & K. Csizér (Eds.), *Vizsgálat a köznevelésben folyó idegennyelv-oktatás kereteiről és hatékonyságáról: Kutatási jelentés* (pp. 189–220). Hungarian Educational Authority.

Lou, N. M., & Noels, K. A. (2019). Language mindsets, meaning-making, and motivation. In M. Lamb, K. Csizér, A. Henry, & S. Ryan (Eds.), *The Palgrave handbook of motivation for language learning* (pp. 537–559). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-28380-3_26

Lowie, M., & Verspoor, M. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition*, *10*(1), 7–21. https://doi.org/10.1017/S1366728906002732

MacIntyre, P. D., & Vincze, L. (2017). Positive and negative emotions underlie motivation for L2 learning. *Studies in Second Language Learning and Teaching*, *7*(1), 61–88. https://doi.org/10.14746/ssllt.2017.7.1.4

Mifsud, M. (2011). *The relationship of teachers' and students' motivation in ELT in Malta: A mixed methods study* [Unpublished doctoral dissertation]. University of Nottingham.

Nagamine, T. (2018). L2 teachers' professional burnout and emotional stress: Facing frustration and demotivation toward one's profession in a Japanese EFL context. In J. de Dios Martínez Agudo (Ed.), *Emotions in second language teaching: Theory, research and teacher education* (pp. 259–275). Springer.

Öveges, E. (2018). Az iskolai nyelvoktatás keretei [The frameworks of institutional language learning]. In E. Öveges & K. Csizér (Eds.), *Vizsgálat a köznevelésben folyó idegennyelv-oktatás kereteiről és hatékonyságáról: Kutatási jelentés* (pp. 14–29). Hungarian Educational Authority.

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.

Pajares, M. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, *62*(3), 307–332. https://doi.org/10.3102/00346543062003307

Rebuschat, P. (Ed.) (2015). *Implicit and explicit learning of languages.* John Benjamins Publishing Company. https://doi.org/10.1075/sibil.48

Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed.). Sage.

Scrivener, J. (2011). *Learning teaching: The essential guide to English language teaching* (3rd ed.). Macmillan.

Tartsayné Németh, N., Tiboldi, T., & Katona, L. (2018). Az intézményvezetők válaszai [The school leaders' responses]. In E. Öveges & K. Csizér (Eds.), *Vizsgálat a köznevelésben folyó idegennyelv-oktatás kereteiről és hatékonyságáról: Kutatási jelentés* (pp. 29–51). Hungarian Educational Authority.

Wallace, M. (1998). *Action research for language teachers.* Cambridge University Press.

6.

The Teacher's Role in Motivating Adult Learners of English: A Questionnaire Study in Hungarian Corporate Contexts

Csaba Kálmán
kalman.csaba@btk.elte.hu

Abstract

The last decade of L2 motivation research has witnessed a surge in publications related to the pedagogical aspects of motivation (Lamb, 2017). The idea of the teacher's role being more salient than has been demonstrated so far is further evidenced by an increasing number of studies conducted in recent years on the teacher's motivational influence (see e.g., Henry & Thorsen, 2018; Wong, 2014). If we narrow down the scope of enquiry to adult L2 learners in corporate contexts, empirical research is even more scarce. This paper fills this gap by presenting the results of a quantitative questionnaire study on the role of the teacher in motivating adult learners of English in Hungarian corporate contexts. The study conducted with 232 adult learners compared the strength of ten dimensions related to the teacher's motivating agency. The relationships between these constructs as well as two additional criterion measure scales (intrinsic and extrinsic motivation) were also investigated. Results show that the personality and the preparedness of the teacher play the most salient role in motivating adult learners of English in Hungarian corporate contexts. Regression analyses attested that the teacher's preparedness and appearance contributed to the participants' intrinsic motivation, whereas their extrinsic motivation could be predicted by the personal branding of the teacher, as well as the teacher's endeavour to incorporate English for specific purposes (ESP) in the syllabus.

Keywords: L2 motivation, teacher's role, adult education, corporate contexts, Hungary

The Teacher's Role in Motivating Adult Learners of English: A Questionnaire Study in Hungarian Corporate Contexts

The majority of businesses are knowledge organizations today, which means that their products and services are inseparable from the knowledge acquired and accumulated by their employees. Three quarters of Hungarian companies already regarded knowledge as a strategic tool in the past decade (KPMG, 2014). Workplaces are becoming increasingly knowledge-intensive, and require continuous learning in the form of non-formal, informal, or formal learning, which is becoming part of our everyday lives (Kovács & Kálmán, 2020). Due to globalization and the internationalization of businesses today, foreign language learning also falls into this category (Moron & Mujtaba, 2018).

As a consequence of the political, social and economic changes in Hungary in the late 1990s, which led to the opening of borders, English language skills have become indispensable in the Hungarian labour market (Földi et al., 2013). The use of English is crucial in professional and corporate life (Sturcz, 2010). Even though recruitment and hiring necessitate the knowledge of foreign languages (Horváth-Csikós, 2018), and B2 level English is a basic requirement in the Hungarian labour market (Sturcz, 2010), most adult employees do not speak languages at the level required by knowledge organizations (Szabó & Mátó, 2019). Therefore, it is not surprising that companies organise onsite or online language training courses for their employees (Kálmán, 2015a; 2019).

The effectiveness and success of these courses depend on a number of factors including the urgency of usable language skills, the profitability of the business, the involvement of the organization in conducting business internationally, the corporate culture and human resources policy of the organization, the attitude of human resources managers and employees to language learning, corporate language teachers, and language schools alike (Kálmán, 2016; Kálmán & Gutierrez, 2015). Regardless of the above

conditions, there is extensive research to prove that motivation plays a pivotal role in the long-term success of language learners (see e.g., Dörnyei & Ryan, 2015; Lamb, 2017).

Even though L2 motivation has been widely researched in the field of applied linguistics, the role of the teacher as a motivating agent has received considerably less attention. If we narrow down the scope of enquiry to adult L2 learners in corporate contexts, empirical research is even harder to come by. This paper fills this niche by presenting the results of a quantitative questionnaire study on the role of the teacher in motivating adult learners of English in Hungarian corporate contexts. The study conducted with 232 adult learners compared the strength of ten teacher-specific dimensions exerting their influence on L2 motivation. The study also measured the participants' intrinsic and extrinsic motivation with two criterion measure scales. In addition to establishing the comparative strength of the ten teacher-related dimensions, the aim of the study was to find out which teacher-specific dimensions function as predictors of intrinsic or extrinsic motivation.

Literature Review

Motivation has long been researched and acknowledged as a predictor of performance and achievement both in corporate and academic contexts (see e.g., Horváth & Kálmán, 2020; Pink, 2009; Winke, 2013). For the theoretical background of this paper, I review the concepts and theories pertinent to the study, such as Pink's (2009) conceptualisation of motivation in the workplace, the role of the teacher in L2 motivation, and the role of English for specific purposes (ESP) in L2 motivation.

A Conceptualization of Motivation in the Workplace

A conceptualisation of motivation in both educational and professional contexts was theorised by Pink (2009), who claims that *autonomy*, *mastery*, and *purpose* are indispensable for humans to be motivated,

presuming that individuals would like to learn, develop, create, and make the world a better place. The core constituents of Pink's (2009) theory are based on Deci and Ryan's (1985) self-determination theory (SDT) and Csikszentmihályi's (1988) concept of *flow*, both of which are concerned with intrinsic motivation, that is, the inherent satisfaction that comes from doing something.

While people are often motivated by external rewards such as money, prizes, and acclaim known as *extrinsic motivation*, self-determination theory focuses primarily on internal sources of motivation such as a need to gain knowledge or independence known as *intrinsic motivation* (Deci & Ryan, 1985). SDT draws on existential, humanistic, and organismic psychologies (Noels et al, 2019) and assumes that individuals' innate psychological needs must be met if they are to develop and flourish (Deci & Ryan, 1985; Ryan & Deci, 2017). The existential character of SDT encapsulates the notion that individuals strive to have meaningful lives, while its humanistic feature postulates that people are content when they are true to their authentic self. Finally, the organismic nature of SDT is indicative of individuals' innate curiosity, and their relentless desire to develop by their constant urge to acquire new skills and explore their environment. Deci and Ryan (1985) and Ryan and Deci (2017) also posit that autonomy, competence, and relatedness are necessary prerequisites of self-actualization, well-being, and optimal functioning.

In Pink's (2009) conceptualisation of motivation, *autonomy* does not mean complete independence, rather, it provides an individual with choice. The second constituent of the model, *mastery*–which is the counterpart of *competence* in Deci and Ryan's (1985) taxonomy–requires openness and curiosity (Dutton & Wrezniewsky, 2001). *Mastery* is not finite, as it can be continually perfected. It also involves grit, effort and deliberate practice, which makes development attractive, frustrating, and challenging at the same time. The third component of Pink's model, *purpose*, is also crucial as it embeds *autonomy* and *mastery* in a context. An intersection of the second and

third constituents of the model (mastery and purpose) is *personal branding*. An individual's personal brand represents the professional values and principles that they can pass on to people they serve. Llopis (2013) summarises this notion with the following words:

> Personal branding does not mean self-promotion–that you should be creating awareness for your brand by showcasing your achievements and success stories. Managing your personal brand requires you to be a great role model, mentor, and / or a voice that others can depend upon. (p. 1)

The Teacher's Role in L2 Motivation

Research on L2 motivation has focused on the learner and the learning environment and has been primarily characterized by analysing the components of L2 motivation and the context until recently. Interestingly, research on teachers' L2 motivating agency has been marginalized. In the social psychological period of L2 motivation research represented by the work of Gardner and Lambert (1972), the teacher's role in motivation was only marginally mentioned in their model as part of the *attitudes towards learning the language.* This negligence can partly be attributed to the fact that Gardner (1985) conducted his research in an English as a Second Language (ESL) context (where the teacher's role can be assumed to be less significant due to immersion in the native language context), and partly to the prevalent language educational approaches in North America at the time, as both audiolingualism and the situational approach were informed by behavioural psychology (Celce-Murcia, 1991), and neither put emphasis on the role of the teacher in the language learning process.

In the subsequent phase of L2 motivation research, the cognitive-situated period, Williams and Burden (1997) described components of L2 motivation emphasising the role of contextual influences, including that of the teacher. In Dörnyei's (1994) three-level framework of L2 motivation, the

learning situation level contains teacher-specific motivational components: the teacher's personality, behaviour, teaching style and practice, and the way tasks are presented and feedback is applied. Dörnyei and Csizér's (1998) study with Hungarian EFL teachers established motivating techniques, which were later expanded by Dörnyei (2001) into 102 motivational strategies (*micro-strategies*) within 35 *macro-strategies*. The cognitive-situated period of L2 motivation research also highlighted the reciprocity of (lack of) teacher motivation and enthusiasm and (lack of) student motivation and enthusiasm. As Csikszentmihályi (1997) highlighted, learners are implicitly motivated by their teachers' enthusiasm. Subsequent studies on the role of the teacher in motivation have confirmed the above findings on teacher enthusiasm: a study by Ghanizaded and Moafian (2010) involving 826 EFL learners found that interpersonal relationships, the teacher's happiness, enthusiasm, support, and empathy had the highest correlations with learners' success.

The next phase of L2 motivation research, the process-oriented period around the 2000s brought the temporal nature of motivation into focus. A model to describe the temporal aspect of L2 motivation was developed by Dörnyei and Ottó (1998), who organised motivational influences into two main dimensions: action sequence and motivational influences, along *pre-actional*, *actional*, and *post-actional* phases. In Dörnyei and Ottó's (1998) model, there is no reference to the teacher, even though it is assumed that the teacher assists the learner to set realistic goals (*pre-actional stage*), supports the learner during the *actional stage*, and in the *post-actional stage*, helps the learner to reflect on and assess the *actional stage*.

The socio-dynamic period of L2 motivation research is characterised by interactions between the context and the self. The influence of *self*-theories was most apparent in the L2 Motivational Self System (L2MSS; Dörnyei, 2005), drawing on Markus and Nurius' (1986) *ideal self* and *ought-to self*, which represent what learners would like to become and what they believe others would like them to become. These two selves, in combination with the *L2 learning experience*, make up Dörnyei's (2005) L2 Motivational Self System.

In the L2MSS, teachers explicitly form part of the third constituent of the model: *the L2 learning experience;* indirectly, however, they can also shape the ideal L2 selves of learners, and through expectations, the ought-to L2 selves.

The last decade of L2 motivation research has witnessed a "surge in publications related to the pedagogical aspects of motivation" (Lamb, 2017). Lamb underlined such novel phenomena as the changing nature of teacher-learner relationships, in which "learners are less accepting of a submissive role in class" (p. 3). The second novelty in his writing is the outright classification of the teacher as a motivator. The idea of the teacher's role being more salient is further evidenced by an increasing number of studies conducted in recent years on the teacher's motivational influence (see e.g., Guilloteaux, 2013; Wong, 2014). Several of these studies conclude that "some individual teachers express more agency than others in developing their learners' motivation, despite working in similar contexts" (Lamb, 2017, p. 14), which is indicative of the teacher's personality and behaviour as a crucial factor in motivating learners. While some teachers demonstrate the quality to motivate more effectively, others might not. Lamb et al. (2016) described this quality as a persistent willingness and ability to empathise with learners, while Lamb (2017) refers to it as *responsiveness*, defined as "the personal quality of empathy […] built up over years of practice, which defines the successful motivator" (p. 15). Another characteristic of a motivating teacher is that he or she is first and foremost motivated as a professional. This area of L2 motivation research–teacher motivation–has been extensively investigated by Csizér recently (2019, 2020) as being motivated in the teaching profession is assumed to be a prerequisite of exerting motivating agency.

ESP in L2 Motivation

Several scholars have confirmed that ESP is more likely to generate higher levels of motivation than EGP (English for general purposes) as it is more pertinent to learners' needs and interests (e.g., Basturkmen, 2010;

Dudley-Evans & St John, 1998). Relevance makes ESP courses more effective, as these courses take learners' needs as a starting point (Krajka, 2018). Additionally, Strevens (1988) highlights the cost-effective nature of ESP courses. Dudley-Evans and St John (1998) claim that "the focused nature of the teaching, its relevance and cost-effectiveness ensure that its aims are widely accepted by learners" (p. 10). Basturkmen (2010) concludes that "it can be argued that because ESP courses cater to students' interests and needs, they are more likely to engender high levels of motivation. It can be assumed that students will be more interested in topics and texts related to their work or study areas" (p. 11).

Empirical evidence substantiating the claims above is still hard to come by (Altalib, 2019). One exception is Altalib's (2019) survey conducted at four Saudi universities involving 4,043 students enrolled in ESP and EGP courses. The findings of his study confirm that there is a significant relationship between the learners' motivation and the type of course they are enrolled in. The ESP group exhibited higher ideal L2 selves and more positive attitudes towards the L2 learning experience than the EGP group. The need for practicality in motivating adult learners of English was also confirmed by Kálmán's (2019) interview study with 18 L2 teachers working in Hungarian corporate contexts and was also a salient theme in interviews conducted with Human Resources Managers of 18 organizations operating in Hungary (Kálmán, 2020). Another study (Dostal, 2016) based on the spoken corpus of Business English (BE) students, in addition to simulating business discussions to demonstrate the importance of relevant course material in developing foreign language communicative competence (FLCC) for English business discussions, concluded that teacher corrective feedback improves FLCC effectiveness in terms of fluency, accuracy, sophistication, control, range, length and variety in interaction; thus highlighting the role of the ESP teacher.

Research Method

In order to establish the relative importance of the ten teacher-specific dimensions in motivating adult learners of English in Hungarian corporate contexts, a questionnaire survey was designed. As the purpose of the study was to test hypotheses based on previous findings by the author (Kálmán, 2015a, 2015b, 2016, 2019, 2020; Kálmán & Gutierrez, 2015), a quantitative questionnaire study was the best choice. Another reason for opting for a quantitative study was to reach a larger number of participants, and to obtain generalizable results. Furthermore, to be able to study causality between the investigated dimensions and the constructs of extrinsic and intrinsic motivation, a quantitative questionnaire was the only possible instrument. Based on the objectives of this study and the theoretical background reviewed, the following research questions were formulated:

> Research question 1: What are the most motivating aspects of a teacher's personality, behaviour, and classroom practices for adult learners of English in a corporate environment?
>
> Research question 2: Which of the teacher-specific dimensions function as predictors of learners' intrinsic and extrinsic motivation?

Participants

All participants of the study were Hungarian employees from 18 organisations with more than 250 employees representing a wide range of industries including the automotive; baby, women and family care; banking; construction; electricity; fast moving consumer goods (FMCG); gas; information technology; insurance; nuclear; pharmaceutical; telecommunications; tobacco; and trading industries; as well as public administration. The organisations involved in the study were contacted through convenience sampling. Six of the organisations were Hungarian state-owned, 12 were public limited companies including two American, two German and one Austrian multinationals. Fourteen of them were based in

Budapest, four elsewhere in the country. Most of the participants had a university degree and worked in sedentary, knowledge-based jobs with the exception of five respondents who worked on a production line or in technical jobs in power plants. In total, 232 employees, 119 females and 113 males completed the questionnaire. 31 of them were working in managerial positions, 201 in non-managerial positions at the time of the survey. The average age of the participants was 37, ranging from 22 to 65 with a standard deviation of 9.25. According to the participants' self-report, the level of learners' proficiency ranged from B1 to C1 on the Common European Framework of Reference (CEFR) (Council of Europe, 2001). Of the participants, 31 rated their proficiency as B1, 137 as B2, and 63 as C1 (data missing for one learner).

The Instrument

Dörnyei's (2007) 5-step validation model was used to develop the instrument. Initially, an *item pool* was drawn up (*Step 1*) with as many potential items for each scale as possible. All the items were developed by the author partly based on the theoretical background (Csikszentmihályi, 1997; Deci & Ryan, 1985; Dörnyei, 1994, 2001, 2007; Dörnyei & Csizér, 1998; Gardner, 1985; Ghanizaded & Moafian, 2010; Heitzmann, 2008; Williams & Burden, 1997; Dörnyei & Ottó, 1998), and partly on the results of previous studies conducted in the field by the author (Kálmán, 2015a, 2016, 2018, 2019) which identified themes that HR managers, teachers, and learners considered key to motivating adult learners. After getting feedback from senior researchers (*Step 2*), the pilot questionnaire was put together (*Step 3*). The verification of the responses to the pilot instrument (*Step* 4) was followed by post-hoc analysis conducted on the scales (*Step* 5), which resulted in the following ten constructs with six items in each:

1. Preparedness: to what extent learners find the professionalism of a teacher motivating. Example: My language teacher motivates me if she or he has a lot of experience in language teaching. (Cronbach's α = .72)

2. Focus on the present: how motivating it is for learners to see that their teachers are focused on the present by addressing current learner needs; by using current technologies, up to date textbooks, updated handouts. Example: My language teacher motivates me if she or he continuously updates the handouts for the lessons. (Cronbach's α = .71)

3. Appearance: how motivating it is for the learners if the language teacher is well-groomed, well-dressed. Example: My language teacher motivates me if she or he looks good. (Cronbach's α = .88)

4. Free choice of topic: how motivating it is for the learners to be given the opportunity to speak about anything they feel like sharing with the teacher. Example: My language teacher motivates me if I can discuss my problems at work with her or him if I feel like it. (Cronbach's α = .82)

5. Incorporating ESP in the syllabus: how motivating it is for the learners if the teacher incorporates specific vocabulary and tasks related to their work in the lessons. Example: My language teacher motivates me if we learn things in the lesson that I can use in my work. (Cronbach's α = .84)

6. Personalised teaching: how motivating it is for the learners if the lessons are personalised to cater for their needs. Example: My language teacher motivates me if she or he adapts to my pace of learning. (Cronbach's α = .76)

7. Personality and behaviour: how motivating it is for learners to see certain personality traits and behaviours of the teacher that have been highlighted in the author's previous studies. Example: My language teacher motivates me if she or he starts and finishes the lesson on time. (Cronbach's α = .71)

8. Getting to know the learner: how motivating it is for the learners if the teacher approaches them holistically, not only as a learner. Example: My language teacher motivates me if she or he is sincerely interested in who I am. (Cronbach's α = .77)

9. Personal branding: how motivating it is for the learners if their teacher is well-known and reputable. Example: My language teacher motivates me if she or he has a good reputation. (Cronbach's α = .78)

10. Atmosphere: how motivating it is for the learners if the teacher creates a good atmosphere in the lessons. Example: My language teacher motivates me if she or he creates an informal atmosphere in the lesson. (Cronbach's α = .77)

The ten constructs were supplemented with two criterion measure scales measuring extrinsic and intrinsic motivation respectively. The two scales were taken from Noels et al.'s (2001) study, and measured the following dimensions:

1. Intrinsic motivation (five items): to what extent learners enjoyed learning English for its own sake. Example: I learn English for the "high" feeling I experience while speaking English. (Cronbach's α = .80)

2. Extrinsic motivation (five items): how motivated learners were from the point of view of instrumentality. Example: I learn English to get a better job in the future. (Cronbach's α = .73)

In the last part of the questionnaire, background questions were asked concerning the participants' age, gender, and position in their organisation.

Data Collection and Analysis

The questionnaire was developed in the participants' mother tongue, Hungarian. Respondents were asked to indicate their level of agreement with the statements on a Likert scale of 1 to 10. Applying a 10-point Likert scale made it possible to avoid the ceiling effect and additionally and gave participants the opportunity to differentiate between the answers more subtly (Beall, 2009). The final version of the pen and paper questionnaire was personally delivered or posted to the HR managers of the 18 organisations involved and then collected three weeks later. A total of 232 questionnaires were returned. All completed questionnaires were computer coded and SPSS (Statistical Package for Social Sciences) 20.0 was used to analyse the data. After the descriptive analysis of the scales, comparative and regression analyses were conducted.

Results and Discussion

Comparative analysis of the scales

To answer RQ1, that is, "What are the most motivating aspects of a teacher's personality, behaviour, and classroom practices for adult learners of English in a corporate environment?", descriptive statistics of the scales, their mean values, and standard deviation values are presented in Table 1. *Personality and behaviour*, *Preparedness*, *Incorporating ESP in the syllabus*, and *Focus on the present* showed the highest mean values, approximating the higher end of the scale, which highlights several inferences. The findings related to the motivating influence of the teacher's *Personality and behaviour* confirm the results of previous research (cf., Csikszentmihályi, 1997; Dörnyei 1994, 2005; Dörnyei & Csizér, 1998; Ghanizaded & Moafian, 2010; Kálmán 2015b; Williams & Burden, 1997). The teacher's personality and behavioural traits measured in the construct were thoroughness, enthusiasm, credibility, flexibility, punctuality, and empathy, all of which both teachers (Kálmán, 2019) and HR managers (Kálmán, 2020) reported as important in motivating learners.

Table 1

Descriptive Statistics of the Scales

Constructs	Mean value	Standard deviation
Personality and behaviour	8.63	.93
Preparedness	8.60	.93
Incorporating ESP in the syllabus	8.45	1.16
Focus on the present	8.44	.97
Personalised teaching	8.21	1.06
Getting to know the learner	7.60	1.26
Atmosphere	7.25	1.31
Free choice of topic	6.46	1.63
Personal branding	6.33	1.49
Appearance	5.76	1.70
Intrinsic motivation	7.14	1.60
Extrinsic motivation	6.87	1.70

*The lines indicate significant differences between the scales above and below the line based on paired t-test procedures performed linearly on the scales ranked based on the means. (For paired sample t-test results see Appendix).

Considering the peculiarity of the research context, it can be concluded that the personality of the teacher is even more important, as learning languages in the organisations under investigation is not compulsory, and courses often take lengthy periods of time; therefore, learners would like to have a positive language learning experience. The teacher's personality and behaviour play a crucial role in creating a positive

language learning experience (Henry & Thorsen, 2018), and it has been confirmed by several studies that a positive learning experience fosters L2 motivation (e.g., Dörnyei, 2001; Heitzmann, 2008; Noels et al., 1999; Wlodkowski, 2008).

Interestingly, for the learners, *empathy*, one if the items measured in the *Personality and behaviour scale* seemed to be the least important of the characteristics (with a mean value of 7.76 compared with 9.21, that of thoroughness), while Wlodkowski (2008) and Kálmán (2019), argue that empathy should be at the forefront, as their results suggested it was crucial. This finding might suggest that although all the participants of the survey were adult learners, they might not expect as much understanding and empathy from their teachers as teachers presume. Based on the assumption that *empathetic* is the opposite of *authoritarian*, we may conclude that this finding coincides with the claim of Chafee et al. (2014), according to which, authoritarian teachers can sustain motivation by controlling the situation and controlling themselves, which indicates that learners might be happy with authoritarian teachers as well.

We can see that the mean value of *Preparedness*–the second highest scoring construct–came a mere 0.03 short of *Personality and behaviour*. As can be seen in Table 1, a paired sample t-test procedure conducted on the two scales does not indicate a significant difference; moreover, the standard deviation values of the two scales are identical. As a result, it can be stated that the teacher's *Preparedness* is as important in motivating learners as the teacher's *Personality and behaviour*.

The key to understanding the underlying mechanism behind the importance of this dimension can be explained by three parallel processes. First, as has been confirmed by Donovan et al. (1999), deep understanding of a subject transforms mere information into usable knowledge. In the context investigated, synthesising knowledge into usable knowledge is indispensable, as *practicality* emerged as one of the most motivating aspects

of on-site language courses (Kálmán, 2019, 2020). Second, teachers' commitment to readiness and professionalism enhances their confidence giving them excellent access to their best talents and memories, which in turn improves their motivating abilities (Zull, 2002). Third, we cannot ignore the relationship between teacher and student motivation (e.g., Carbonneau et al., 2008; Martin, 2006; Roth et al., 2007). If we accept Pink's (2009) theory on motivation in the workplace, which cites mastery (i.e., professionalism) as one of the three prerequisites of a motivated professional (in any profession), we must assume that well-prepared teachers are motivated teachers, and as such, they increase their students' motivation.

The second most important group of dimensions contains *Incorporating ESP in the syllabus*, and *Focus on the present*, with mean values of 8.45 and 8.44, respectively. The difference in the importance of the two constructs is non-significant again; however, the standard deviation value of *Incorporating ESP in the syllabus* (1.16) is somewhat higher than any of the other three dimensions at the top end of the league (0.93–0.97). This indicates that while there was greater agreement among participants on the importance of *Personality and behaviour*, *Preparedness and Focus on the present* in the motivational impact of the teacher, there was less consensus on the *Importance of ESP* in the syllabus. This finding might reflect the wide variety of English courses attended by participants at the time of the survey, some of which were fully ESP-focused, others combined ESP with general English, and some of the courses were teaching only general English. Both *Incorporating ESP in the syllabus* and *Focus on the present* contained items that measured how the *relevance* of the teaching material motivated learners.

The next dimension in line was *Personalised teaching* (M = 8.21, DS = 1.06). There might be three underlying reasons why personalised teaching is so important and, at the same time, self-evident in this context. For a start, language teaching for an organisation, is similar to a whole range of services purchased by the organisation. Tailoring services to the customer's specific needs gives them a competitive advantage over other service providers who

do not personalise their services. Since there are plenty of language schools and language teachers on the market, usually selected through tendering, it is safe to say that in the long run, those who are able to provide personalised services will prevail (Kálmán, 2015a). Secondly, corporate on-site courses can be far more easily personalised than in institutional school settings, as most on-site courses are one-to-one, while learning in groups usually means a maximum of five members (Kálmán, 2015a). While the first two reasons are external to and independent of the learner, the third one explains how tailor-made teaching can contribute to motivation. By tailoring the course in response to the needs and requests expressed, learner autonomy is enhanced, and learning becomes more relevant to the learner's needs and preferences. This autonomy-supportive behaviour has been associated with learners' self-determined motivation and positive feelings about learning (e.g., Assor et al. 2002; Benson, 2007, 2010; Black & Deci, 2000).

Getting to know the learner ($M = 7.60$, $SD = 1.26$), came fourth, which can be explained by two reasons. On the one hand, getting to know the learner is a prerequisite of tailor-made teaching, as the more the teacher finds out about the learner, the more she or he can tailor aspects of teaching to the needs of the learner. On the other hand, by getting to know the learner, the quality of interaction between the teacher and the learner can be improved. The importance of this dimension of motivating has been verified by e.g., Crookes and Schmidt (1991), who claimed that "type of interaction between teacher and students is likely to increase, maintain, or decrease the students' motivation" (p. 483). However, interaction does not necessarily stop at the level of transmitting cognitive information, but opens more dimensions of interacting, which in turn results in a richer human experience both for the learner and the teacher. In this way, one can make a positive difference not only in learning, but also in the lives of one's students. Day (2004) claims that passionate teachers can achieve this:

> Teachers with a passion for teaching are those who are committed, enthusiastic, and intellectually and emotionally energetic in their

> work with children, young people, and adults alike. Yet these overt signs of passion are underpinned by clear moral purposes that go beyond the efficient implementation of set curricula. Passionate teachers are aware of the challenge of the broader social contexts in which they teach, have a clear sense of identity and believe that they can make a difference to the learning and achievement of their pupils . . . For these teachers, teaching is a creative and adventurous profession and passion is not an option. It is essential to high-quality teaching. (p. 2)

Next in line was *Atmosphere* (M = 7.25, SD = 1.31), which measured both the importance of the physical environment, that is, the room where the lessons were held, and the technical infrastructure used in the lessons, as well as the atmosphere created by the teacher. The mean values of the items within the construct show an even distribution, which means that both are equally important in motivating learners. The relatively high mean value of this construct is not surprising, as it has been proved by several studies that a pleasant learning environment contributes to a positive learning experience, which increases motivation (see e.g., Crookes & Schmidt, 1991; Dörnyei, 1994; Dörnyei & Csizér, 1998; Williams & Burden, 1997).

The next construct, *Free choice of topic* is the most controversial one in the survey. Its mean value of 6.46 did not come close to what could have been hypothesised based on the findings of Kálmán (2015b, 2019, 2020). Learners, teachers, and HR managers alike highlighted the importance of free choice of topics in motivation. It should be added that the standard deviation of this construct was the second highest (1.63), which indicates that there is a greater degree of disagreement among participants. Nevertheless, based on the results of previous studies, it is possible that learners do not realise or are unaware of the underlying motivational effect of talking about anything in the lessons. Another possibility is that they find it embarrassing to admit that sometimes they come to lessons specifically to vent their work-related anger

or problems, as it has been expressed by teachers and HR managers in Kálmán (2019, 2020).

The teacher's *Personal branding* came second to last among the ten dimensions measured; however, statistically, it did not prove less important than *Free choice of topic*. Even though the mean value of this construct was relatively low ($M = 6.33$, $SD = 1.63$) compared to the other constructs, *Personal branding* cannot be ignored as its mean value is relatively high on a scale from 1 to 10. A teacher's personal branding can be important both in generating and maintaining the motivation of learners. Due to the mechanisms of social networks (Mercer, 2015), teachers with a good reputation can encourage employees in an organisation to start learning a language and can also play a crucial part in the evolution of corporate language education systems. At the same time, the teacher's reputation can speed up the process of building trust in the learners and, thus, can foster the development of a relaxed learning environment where the learners can put aside their mistrust of the teacher right from the beginning.

Finally, the least motivating dimension proved to be *Appearance* with a mean value of 5.76 and the highest standard deviation of 1.70. The data show that the participants' opinions are most divided about the motivating influence of this dimension. Whether 5.76 is high or low is a matter of opinion. Nevertheless, Howlett et al. (2013) have confirmed that clothing conveys information about the wearer and first impressions can be heavily influenced by the messages conveyed by attire. The study of Howlett and colleagues (2013) revealed that people were more positively rated on the attributes of *confidence*, *success*, *flexibility* and *the ability to earn money* when they wore smarter clothes. If we translate this finding into the classroom, well-dressed teachers can give the impression of being more confident, more successful, and more flexible, and in the discussion on *Personality and behaviour*, *Preparedness* and *Personal Branding* above, we have seen that the teacher's flexibility, confidence, and reputation may increase learner motivation.

Regression analysis between the scales

In order to answer the second research question (Which teacher-specific scales function as predictors of learners' intrinsic and extrinsic motivation?) linear regression analyses were carried out with a stepwise approach separately for the criterion variables of *Intrinsic* and *Extrinsic motivation*. The results are summarised in Table 2 and 3. Of the ten dimensions investigated, only two contributed significantly to *Intrinsic motivation*: *Preparedness* and *Appearance*; and another two contributed significantly to *Extrinsic motivation: Personal branding and Incorporating ESP in the syllabus*.

Table 2

Results of Regression Analysis of the Teacher Scales with Intrinsic Motivation as the Criterion Variable (significance level p <.01)

Variable	*β*	*t*	*p*	
1. Preparedness	.31	4.81	<.001	
2. Appearance	.25	3.93	<.001	
R^2				.22

The data show that the proportion of variance in *Intrinsic motivation* explained by the two independent variables is 22%, and the impact of *Preparedness* (.31) is stronger than the impact of *Appearance* (.25) in the equation. Since the key element of intrinsic motivation is the enjoyment of the activity in general, the data show that about one fifth (22%) of this enjoyment is due to a greater extent to the teacher's preparedness and to a lesser extent to his or her appearance. This causality cannot simply be explained by claiming that learners in this context enjoy learning English because their teacher is well prepared and looks good. The answer may be found in the learners' identification with the teacher. Since most participants of the survey have at least one university degree, and work in a very

competitive environment, it can be assumed that they are diligent, well-prepared professionals in their own field of work. Because of the business environment they work in, probably they pay more attention to their own appearance than average citizens. Therefore, presumably, they enjoy spending time in the company of likeminded people who are as well-prepared and well-groomed as they are, and take pleasure from conditions which are present.

Table 3

Results of Regression Analysis of the Teacher Scales with Extrinsic Motivation as the Criterion Variable (significance level p <.01)

Variable	*β*	*t*	*p*	
1. Personal branding	.39	5.94	<.001	
2. Incorporating ESP	.17	2.55	<.001	
R^2				.23

By comparison, the data obtained from the regression analysis of the teacher scales with *Extrinsic motivation* suggest a nearly equal causal relationship (23%) between personal branding and the inclusion of ESP in the syllabus (see Table 3). Of the ten dimensions, two contributed significantly to *Extrinsic motivation*: *Personal branding* and *Incorporating ESP in the syllabus*. The effect of *Personal branding* (.39) on *Extrinsic motivation* is more than twice as strong as the effect of *Incorporating ESP in the syllabus* (.17). The significance of both scales in contributing to extrinsic motivation can be explained by the fact that the participants' work environment is highly competitive; therefore, instrumentality, which is a core determinant of extrinsic motivation, plays an important role. The participants might want to move up the corporate ladder or earn more, all of which are associated with the instrumental use of learning English. Both *Personal branding* and *Incorporating ESP in the syllabus* are strongly connected to developing one's

professional self. The teacher's personal branding might be important from the point of view of setting an example of building one's professional self, or for the identification mechanism described in the previous paragraph. *Incorporating ESP in the syllabus* is another means that helps corporate language learners fulfil their professional selves and development.

Conclusion and implications

This study attempted to fill a niche in L2 motivation research by highlighting the role teachers play in generating and maintaining the motivation of adult language learners in Hungarian corporate contexts. The quantitative questionnaire was completed by 232 corporate English learners, and measured constructs that have been shown in previous studies to be key in L2 motivation. The previous studies have been conducted with the participation of all stakeholders of corporate language education in Hungary: learners, teachers, and HR managers alike. As the constructs of the current study were based on the results of previously conducted studies, the main purpose of this questionnaire was to find out which aspects of a teacher's personality, behaviour, and classroom practices played the most crucial role in motivating adult learners of English. Apart from the rank order between the investigated dimensions, the study also measured relationships between these aspects. Regression analyses were conducted on the following ten constructs: *Appearance*, *Atmosphere*, *Focus on the present*, *Free choice of topic*, *Getting to know the learner*, *Incorporating ESP in the syllabus*, *Personal branding*, *Personality and behaviour*, *Preparedness*, and *Personalised teaching*, and two additional criterion measure scales of *Intrinsic* and *Extrinsic motivation*.

The findings of the study have several pedagogical implications. *Personality and behaviour* and *Preparedness* proved to be the most motivating dimensions of a language teacher. As the personality and behavioural traits measured were *thoroughness*, *enthusiasm*, *credibility*, *flexibility*, *punctuality*, and *empathy*, it is highly advisable to possess or perfect these characteristics if one

would like to teach in a corporate environment. As the teacher's *Preparedness* is as important in motivating learners as the teacher's *Personality and behaviour*, knowing and speaking the English language, having a lot of experience in language teaching, training oneself regularly, and being able to handle unexpected situations in the classroom are also pivotal.

The second most important set of dimensions in the final results were *Incorporating ESP in the syllabus*, and *Focus on the present.* As a result, a teacher in this context should be familiar with the language of a particular industry. Since the jargon and terminology related to a field is changing at breakneck speed, the best source for the teacher to learn the necessary ESP is the learner. It is a good idea to ask the learner for English texts or emails that they have come into contact with, so that the teacher can get acquainted with the vocabulary and set up exercises and role-plays for the learner. Another aspect of ESP is presentation language and the language of negotiations, both of which are needed in all industries. Therefore, it is wise for a teacher to have some personal experience in presenting and to have the skills to be able to teach it at the same time. As far as the dimension *Focus on the present* is concerned, it is safe to say that corporate language learners are more concerned with the present than the past or the future. They need immediate solutions and prompt, constantly updated answers to their needs that may change on a daily basis. Teaching materials should be updated constantly. Tailoring teaching to the needs of a client is equally important. In fact, because of the modern quality management systems of organisations today, teachers will not only be expected to meet their clients' expectations, but to exceed them. This requires a lot of flexibility for the teacher to accommodate the personality and the needs of the learner in many ways.

For teachers to be able to personalise teaching, they should get to know the learner as much as possible. The interactions do not only provide ample opportunities to practise the language, but open more dimensions for interacting, which, in turn, results in a richer human experience for both the learner and the teacher. Teachers play a pivotal role in creating a relaxed

atmosphere by being natural, accessible, informal, respectful, and naturally by treating learners as their equals. A teacher's personal branding can be important both in generating and maintaining the motivation of learners. For a start, teachers with a good reputation can win better contracts and get into prestigious organisations. Much of building one's reputation is hard work, but one can borrow clever tricks from marketing as well. Finally, appearance does count, especially in a corporate environment. Clothing communicates information about the wearer, and first impressions can be heavily influenced by the messages conveyed by what we wear and how we look.

Despite the circumspect approach taken in designing and implementing this study, a weakness is that about 80% of the participating organisations and participants work, live, and learn English in Budapest; therefore, the Hungarian countryside is underrepresented in the survey. Consequently, a possible future research direction could be to investigate the teacher's role in motivation in that context. Alternatively, it would be exciting to replicate (some of) the studies in other Central and Eastern and/or Western European corporate contexts and carry out a comparative analysis of the results. Additionally, within the framework of longitudinal research, the studies could be repeated in the investigated context in a few years' time to observe changes in corporate language training practices. As for the teacher's motivational influence, it would be worthwhile replicating the final questionnaire study in other institutional contexts to find out which aspects of the teacher's personality, character and practices are found motivating by learners in these contexts. Further to this, it would be interesting to do more research on the motivating influence of the relationship between a learner and a teacher, as well as the reciprocity of this relationship. It is hoped that the results of this study will not only prove useful for corporations, language schools, language teachers, and teacher trainers, but will also fill a niche in research on corporate language training and L2 motivation research and draw attention to the teacher's role in motivation as a more central component in L2 motivation theories.

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

Altalib, A. (2019). L2 motivation in ESP and EGP courses. *Australian Journal of Applied Linguistics*, *2*(1), 1–16. http://dx.doi.org/10.29140/ajal.v2n1.113

Assor, A., Kaplan, H., & Roth, G. (2002). Choice is good, but relevance is excellent: Autonomy-enhancing and suppressing teacher behaviours predicting students' engagement in schoolwork. *British Journal of Educational Psychology*, *72*, 261–278. https://psycnet.apa.org/doi/10.1348/000709902158883

Basturkmen, H. (2010). *Developing courses in English for specific purposes*. Palgrave Macmillan.

Beall, A. (2009). *10-point scale fan club.* Argentum Strategy Group. Retrieved from: https://www.argentumstrategy.com/10point-scale-fan-club/.

Benson, P. (2007). Autonomy in language teaching and learning. *Language Teaching*, *40*, *21–40.* https://doi.org/10.1017/S0261444806003958

Benson, P. (2010). *Teaching and researching autonomy in language learning* (2nd ed.). Longman.

Black, A. E., & Deci, E. L. (2000). The effects of instructors' autonomy support and students' autonomous motivation on learning organic chemistry: A self-determination theory perspective. *Science Education*, *84*, 740–756. https://psycnet.apa.org/doi/10.1002/1098-237X(200011)84:6%3C740:: AID-SCE4%3E3.0.CO;2-3

Carbonneau, N., Vallerand, R. J., Fernet, C., & Guay, F. (2008). The role of passion for teaching in intrapersonal and interpersonal outcomes. *Journal of Educational Psychology*, *100*, 977–987. https://psycnet.apa.org/doi/10.1037/a0012545

Celce-Murcia, M. (1991). *Teaching English as a second or foreign language* (2nd ed.). Heine & Heine Publishers.

Chafee, K. E., Noels, K. A., & McEown, M. S. (2014). Learning from authoritarian teachers: Controlling the situation or controlling yourself can sustain motivation. *Studies in Second Language Learning and Teaching*, *4*(2), 355–387. https://doi.org/10.14746/ssllt.2014.4.2.9

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Retrieved from: http://www.coe.int/t/dg4/linguistic/cadre1_en.asp.

Crookes, G., & Schmidt, R. (1991). Motivation: Reopening the research agenda. *Language Learning*, *41*, 469–512. https://doi.org/10.1111/j.1467-1770.1991.tb00690.x

Csikszentmihalyi, M., (1988). The flow experience and its significance for human psychology. In M. Csikszentmihalyi & I. S. Csikszentmihalyi (Eds.), *Optimal experience: Psychological studies of flow in consciousness* (pp. 15–35). Cambridge University Press. https://doi.org/10.1017/CBO9780511621956.002

Csikszentmihalyi, 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.

Csizér, K. (2019). The language-learning experiences and their perceived impact on teaching: An interview study with English teachers in Hungary. In M. Sato & S. Loewen (Eds.), *Evidence-based second language pedagogy: A collection of instructed second language acquisition studies* (1st ed.) (pp. 314–330). Routledge. https://doi.org/10.4324/9781351190558.

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

Day, C. (2004). *A Passion for teaching*. Routledge Falmer.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum.

Donovan, M. S., Bransford, J. D., & Pellegrino, J. W. (1999). *How people learn: Bridging research and practice*. National Academy Press.

Dostal, M. (2016). Developing foreign language communicative competence for English business meetings using business meeting simulations. *Scripta Manent*, *11*, 2–20.

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

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.

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

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

Dörnyei, Z., & Ottó, I. (1998). Motivation in action: A process model of L2 motivation. *Working Papers in Applied Linguistics (Thames Valley University, London)*, *4*, 43–69.

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

Dudley-Evans, T., & St John, M. J. (1998). *Developments in English for specific purposes: A multi-disciplinary approach*. Cambridge University Press.

Dutton, J. E., & Wrezniewsky, A. (2001). Crafting a job: Revisioning employees as active crafters of their work. *Academy of Management Review*, *26*(2), 179–201. https://doi.org/10.2307/259118

Földi, K., László, É., Szűcs, R. S., & Máté Z. (2013). A munkaerőpiacon szükséges nyelvi kompetenciák feltérképezése kvalitatív eszközökkel [Mapping language competencies necessary in the labour market]. *Szolnoki Tudományos Közlemények*, *17*, 140–151.

Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. Edward Arnold.

Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and motivation in second language learning*. Newbury House.

Ghanizadeh, A., & Moafian F. (2010). The role of EFL teachers' emotional intelligence in their success. *ELT Journal*, *64*(4), 424–435.

Guilloteaux, M.-J. (2013). Motivational strategies for the language classroom: Perceptions of Korean secondary school English teachers. *System 41*(1), 3–14.

Heitzmann, J. (2008). The ups and downs of motivation: A longitudinal study of a group of secondary school learners of English. [Unpublished doctoral dissertation]. Eötvös Loránd University, Budapest.

Henry, A., & Thorsen, C. (2018). Teacher-student relationships and L2 motivation. *The Modern Language Journal*, *102*(1), 218–241. https://doi.org/10.1111/modl.12446

Horváth, I., & Kálmán, Cs. (2020). Motivational disposition of translation and interpreting graduates. *Interpreter and Translator Trainer*. https://doi.org/10.1080/1750399X.2020.1838180

Horváth-Csikós, G. (2018). A nyelvtudás szükségessége a munkaerőpiacon [The significance of language knowledge on the labor market]. *Studia Mundi–Economica*, *5*(2), 2–13.

Howlett, N., Pine, K., Orakçıoğlu, I., & Fletcher, B. (2013). The influence of clothing on first impressions: Rapid and positive responses to minor changes in male attire. *Journal of Fashion Marketing and Management: An International Journal*, *17*(1), 38–48. https://doi.org/10.1108/13612021311305128

Kálmán, Cs. (2015a). Magyarországi nagyvállalatok nyelvoktatási gyakorlata, és nyelviskolákkal, nyelvtanárokkal szemben támasztott elvárásaik 2015-ben [Corporate language education, and corporations' expectations of language schools and language teachers in Hungary in 2015]. *Modern Nyelvoktatás*, *21*(4), 3–23.

Kálmán, Cs. (2015b). The teacher's role in generating and maintaining the motivation of adult learners of English in a corporate setting. *Working Papers of Language Pedagogy*, *9*, 1–21.

Kálmán, Cs. (2016). A nyelvtanár szerepe vállalati környezetben angol nyelvet tanuló felnőttek motivációjában három kutatás tükrében [The teacher's role in motivating adult learners of English in corporate contexts: A summary of three studies]. *Neveléstudomány: Oktatás, kutatás, innováció*, *4*, 46–63. https://doi.org/10.21549%2FNTNY.16.2016.4.4

Kálmán, Cs. (2019). EFL teachers' motivation and motivational impact in corporate contexts. *Journal of Adult Learning, Knowledge and Innovation*, *2*(1), 33–47. https://doi.org/10.1556/2059.02.2018.03

Kálmán, Cs. (2020). HR managers' views on L2 teachers' motivating impact in corporate contexts, and their attributions in retrospective L2 learning experiences. In Cs. Kálmán (Ed.), *DEAL 2020: A Snapshot of Diversity in English Applied Linguistics* (pp. 297–333). Eötvös Kiadó.

Kálmán, Cs., & Gutierrez, E. (2015). Successful language learning in a corporate setting: The role of attribution theory and its relation to intrinsic and extrinsic motivation. *Studies in Second Language Learning and Teaching*, *5*(4), 583–608. https://doi.org/10.14746/ssllt.2015.5.4.4

Kovács, Zs., & Kálmán, Cs. (2020). Challenges in measuring self-regulated learning at the workplace. In H. Andl, A. Híves-Varga, & Zs. Molnár-Kovács (Eds.), *New research in educational sciences* (pp. 45–54). Complex Kiadó.

KPMG (2014). *Szervezeti tudásmegosztás Magyarországon 2013/2014* [Corporate knowledge sharing in Hungary2013/2014]. Retrieved from http://tudatosvezetes.hu/docs/tudasmegosztas.pdf

Krajka, J. (2018). The ESP teacher as a researcher–from needs analysis to materials development. *Scripta Manent*, *13*, 2–25.

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

Lamb, M., Astuti, S. P., & Hadisantosa, N. (2016). 'In their shoes': What successful Indonesian school teachers do to motivate their pupils. In M. Apple, D. Da Silva, & T. Fellner (Eds.), (pp. 195–216). *Language learning motivation in Asia.* Multilingual Matters.

Llopis, G. (2013). Personal branding is a leadership requirement, not a self-promotion campaign. *Forbes*, April 8, 2013. Retrieved http://www.forbes.com/sites/glennllopis/2013/04/08/personal-branding-is-a-leadership-requirement-not-a-self-promotion-campaign/#5cad034b15c0

Markus, H., & Nurius, P. (1986). Possible selves. *American Psychologist*, *41*, 954–969. https://psycnet.apa.org/doi/10.1037/0003-066X.41.9.954

Martin, A. J. (2006). The relationship between teachers' perceptions of student motivation and engagement and teachers' enjoyment of and confidence in teaching. *Asia-Pacific Journal of Teacher Education*, *34*(1), 73–93. https://doi.org/10.1080/13598660500480100

Mercer, S. (2015). Social network analysis and complex dynamic systems. In Z. Dörnyei, P. MacIntyre, & A. Henry (Eds.), *Motivational dynamics in language learning* (pp. 73–83). Multilingual Matters.

Moron, V., & Mujtaba, G. B. (2018). Helping non-English speakers speak English in the workplace. *American Journal of Education and Learning*, *3*(1), 14–26. http://doi.org/10.20448/804.3.1.14.26

Noels, K. A., Clément, R., & Pelletier, L. G. (1999). Perceptions of teacher communicative style and students' intrinsic and extrinsic motivation. *Modern Language Journal*, *83*, 23–34.

Noels, K. A., Clément, R., & Pelletier, L. G. (2001). Intrinsic, extrinsic, and integrative orientations of French Canadian learners of English. *The Canadian Modern Language Review*, *57*(3), 424–442. https://doi.org/10.3138/cmlr.57.3.424

Noels, K. A., Lou, N. M., Lascano, D. I. V., Chaffee, K. E., Dincer, A., Zhang, Y. S. D., & Zhang, X. (2019) Self-determination and motivated engagement in language learning. In M. Lamb, K. Csizér, A. Henry, & S. Ryan (Eds.), *The Palgrave handbook of motivation for language learning* (pp. 95–115). Palgrave McMillan.

Pink, D. (2009). *Drive: The Surprising truth about what motivates us.* Penguin.

Roth, G., Assor, A., Kanat-Maymon, Y., & Kaplan, H. (2007). Autonomous motivation for teaching: How self-determined teaching may lead to self-determined learning. *Journal of Educational Psychology*, *99*, 761–774. https://psycnet.apa.org/doi/10.1037/0022-0663.99.4.761

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness.* Guilford.

Strevens, P. (1988). ESP after twenty years: A re-appraisal. In M. L. Tickoo (Ed.), *ESP: State of the art* (pp. 1–13). SEAMEO Regional Language Centre.

Sturcz, Z. (2010). A munkaadók nyelvi elvárásai a szakértelmiséggel szemben: 2010 [Employers' language requirements towards professionals: 2010]. *Modern Nyelvoktatás*, *16*(4), 7–18.

Szabó, Cs. M., & Mátó, F. (2019). A munkaerőpiac elvárása: Idegennyelvtudás. Hogyan fejleszthetik hallhatóik idegennyelvtudását a felsőoktatási intézmények? In Cs. M. Szabó & F. Mátó (Eds.), *Nyelvi terek*, (pp. 1–14). DUF Press.

Williams, M., & Burden, R. L. (1997). *Psychology for language teachers*. Cambridge University Press.

Winke, P. (2013). An investigation into second language aptitude for advanced Chinese language learning. *The Modern Language Journal*, *97*(1), 109–130. https://doi.org/10.1111/j.1540-4781.2013.01428.x

Wlodkowski, R. J. (2008). *Enhancing adult motivation to learn*. Wiley and Sons.

Wong, R. M. H. (2014). An investigation of strategies for student motivation in the Chinese EFL context. *Innovation in Language Learning and Teaching*, *8*(2), 132–154. https://doi.org/10.1080/17501229.2013.777449

Zull, J. E. (2002). *The art of changing the brain: Enriching the practice of teaching by exploring the biology of learning.* Stylus.

Appendix

Paired Samples T-test Results of the Scales

Paired samples	*t*	*p*
Preparedness–ESP	2.228	.027
Focus on the present–Personalised teaching	3.490	.001
Personalised teaching–Getting to know the learner	9.823	.000
Getting to know the learner–Atmosphere	4.915	.000
Atmosphere–Free choice of topics	9.926	.000
Personal branding–Appearance	6.740	.000
Intrinsic motivation–Extrinsic motivation	-2.163	.032

7.

Fairness in Integrated Academic Writing Ability Assessment: An Empirical Investigation of the Effect of Test-taker Gender on Task Performance

Gyula Tankó
tanko.gyula@btk.elte.hu

Abstract

English for Academic Purposes has gained incontestable prominence due to English being the international language of academic communication. Fairness in assessment is paramount (Bachman & Palmer, 2010) and a potential threat to test validity due to bias. Test taker gender has proved to be one such source of bias (e.g., Du et al., 1996). However, its effect on the direct assessment of academic writing performance has been sporadically investigated and typically in relation to large-scale international tests despite the fact that bias affects each test that has consequences for stakeholders. Few studies analysed the effect of gender bias on tertiary-level small-scale yet high-stakes academic writing tests (e.g., Morris, 1998), and none investigated gender bias in integrated language assessment. The current discourse analysis-based study aimed to contribute original knowledge to this limited body of literature available. It investigated whether an integrated guided summary writing task designed to measure the academic writing performance of 221 first year EFL English majors elicited differential performance from male and female students and whether this was reflected in the scoring decisions made by raters. It followed up an investigation of gender effect on argumentative essay quality and rating (Tankó, 2021) that was similarly motivated by concerns about potential threats to test validity. The results of contrastive analyses of test scores and linguistic indicators generated from student essays and summaries revealed significant gender-specific differences in language use and unjustifiable gendered differences in the essay scores that inform academic writing instruction and assessment.

Keywords: assessing English for Academic Purposes, fairness, bias, gender, guided summary writing

Fairness in Integrated Academic Writing Ability Assessment: An Empirical Investigation of the Effect of Test-taker Gender on Task Performance

As English has become the international language of academic communication, researching, teaching, and assessing English for Academic Purposes (EAP) has also gained considerable prominence. English academic writing is taught to students at the majority of institutions of tertiary education where English is the medium of education. Irrespective of whether the assessment of EAP is carried out internationally in large-scale language examinations (e.g., IELTS, TOEFL, PTE Academic) or locally in examinations developed by universities for internal use, fairness is a common concern.

Fairness is considered an essential characteristic of any high-stakes and low-stakes language tests that have consequences for test takers (Bachman & Palmer, 2010), and it constitutes the basis for the trust of stakeholders (Bachman & Damböck, 2018). The fairness of a test can be jeopardised by numerous factors, one of the lesser researched of these being gender, a denomination that is used in assessment research not in the sense of a socially constructed and continuous concept but as a less loaded term for biological sex. Gender has been shown to be a source of bias for both oral (e.g., O'Sullivan, 2000; Reemann et al., 2013) and written (e.g., Du et al., 1996; Gyagenda & Engelhard, 2009) language tests. Given that any potential source of bias may affect the validity of a test, language test developers and users are expected (Bachman & Palmer, 2010; Messick, 1989) to investigate that the assessment process is free of bias.

It has been proposed that the construct of communicative competence itself is not gender neutral (O'Loughlin, 2002). Furthermore, as demonstrated by several studies (e.g., Du et al., 1996; Gyagenda & Engelhard, 2009; Haswell & Haswell, 1996; Morris, 1998; Tankó, 2021), the assessment of academic written production can be a highly gendered process in which a gender effect may originate from a number of sources, two of which are the language use of the test takers or the rating decisions. Either of these

represents a threat to the validity of the test, so if gender bias occurs, it must be eliminated or controlled. Studies addressing this concern that are not on high-stakes large-scale international language tests are rare (see Gender differences in written language performance section below). However, a previous study conducted on the Academic Skills Test (AST) developed to measure the academic writing ability of English major BA students at a large university in Hungary found that the test takers' gender-specific differential language use could be detected in their argumentative essays written as part of the AST (Tankó, 2021). It also reported that although male students' academic prose production was found to be superior to that of female students, this was not reflected in their test scores. On the contrary, the female students' total scores were significantly higher.

These findings directly motivated the current discourse-based research study, which investigated the academic writing performance of students on the second task administered as part of the AST, the guided summary writing task. Specifically, the aim set for this study was to investigate whether the process of assessing academic writing skills with a reading-into-writing integrated guided summary writing task was affected by the gender of the test takers. As no other study on gender effects in relation to integrated academic writing tasks like the guided summary writing task was conducted, the findings of the study on argumentative essays (Tankó, 2021) and of this follow-up study focusing on guided summaries contribute original knowledge to the body of literature on the effect of gender bias on the academic writing performance of university students.

Gender Differences in Written Language Performance

Gender Differences in Academic Prose in University Settings

In addition to and informed by studies on non-academic prose, investigations were also conducted on gender differences in academic language use. The evidence generated by these studies is fairly consistent, with the majority of the studies asserting the presence, detectability, and effects of gender in academic written production. The study conducted by Rubin and Greene (1992) analysed the lexical, syntactic, and text level features of expressive compositions written by male and female students to an intimate audience and argumentative compositions written to a distant audience. In addition to similarities, they found that by using several hedges and expressions of tentativeness in their argumentation, female students were less confrontational and more affiliative. In contrast, male students used very few hedges.

These findings were reinforced by those of Ellis et al. (2008), who in their synthesis of gender differences that hold across cultures and time reported that male language users were more confident academic writers than female ones and used more judgemental adjectives (e.g., *serious* + N, *egocentric* + N). Female academic writers, on the other hand, were reported to use more qualifying words (e.g., *generally*) and phrases (e.g., *for the most part*) than their male peers.

Martin (1997) analysed the content and language use in the texts of 201 female and male undergraduate students relative to how close they were to their next examination. She found that female students were significantly more anxious than male students because of the impending examinations and grading. According to her, the cautious and more hesitant style of the oral and written academic production of female students was caused by their awareness of and concern about the impending assessment process. This may suggest that the academic prose produced by male and female students

is differentially affected by examination situations and by extension by institutional expectations.

Tankó (2021) identified gender-related differences in the written production of first year English major students and the scores awarded for their argumentative essays written as part of a high-stakes academic skills test. The computer-assisted syntactic and lexical complexity analyses revealed that male writers significantly outperformed their female peers on one syntactic and several lexical variables, notably in the use of complex nominal phrases, sophisticated and varied lexis, and the variety of adverbs. These formal features are all characteristic of high information content registers like academic prose. However, while male students were not rewarded for these formal linguistic complexity features, female students received significantly higher task achievement and total scores for their argumentative essays. The recommendations formulated based on the study were that male writers should be encouraged to adhere closely to the guidelines and instructions provided for written assignments and writing tasks in tests, and both male and female students should be taught to write in a cautious and considered tone that is neither tentative nor bold, resulting in a gender-trait-free academic writing style. Further recommendations stated that the rating scales used to assess academic writing should reflect the key features of academic prose, raters must be better trained to recognise the features of academic prose and to avert the occurrence of halo effect during rating, and test papers should be anonymised.

Contrary to the above studies on academic writing, two studies, both of which were arguably setting-restricted, contended that the university students investigated did not have gender-specific writing styles. Notably, the systematic enquiry was motivated by the fact that distinct gendered patterns could be identified in achievement figures in undergraduate examinations in spite of blind marking. In that study, Francis, Robson, et al. (2001) analysed the tone of the texts written by undergraduate history students by grouping sentences in the three main categories labelled bold,

tentative, or evaluative. It was found that although male students adopted a bold style more frequently than their female peers, overall there were more similarities in the writing style of the two groups. The researchers argued that examiners rewarded the type of academic writing style that was free of gender traits and was characterised by a cautious and considered tone, and that all the students learnt to adopt this circumspect academic writing style. Their findings echo Newman et al.'s (2008) observation about the gender-difference-eliminating effect of constrained language use situations and both are in contrast with Martin's (1997) or Tankó's (2021) findings.

The second study (Francis, Read, et al., 2003) focused on the lecturers' perception of the connection between gender and undergraduate writing. Most lecturers were reported to be unable to correctly identify the authors' gender based only on their writing. However, it was also found that they interpreted the gender dichotomy in terms of abilities and were most likely influenced by widely held gender stereotypes. Female students were described as being better at written communication, fluent, articulate, conscientious, and concerned with presentation, but less confident, and more indecisive. Male students were characterised as being more self-confident, incisive, logical, and focused, but less articulate, more careless, and abrupt.

Based on the evidence generated by the studies conducted on gender-specific differences in academic settings, the argumentative writing of female students was reported to be less confrontational, more affiliative, and characterised by more qualifying language use. In contrast, male students were reported to be more assertive, confrontational, and confident academic writers who used more judgemental language. Conflicting evidence was found about the effect of constrained language use situations in tertiary institutional settings on gender differences in academic written production. However, even the studies that claimed that the academic writing of male and female students was similar were occasioned by the distinct associations observed between gender and examination assessment records. A similar

association, specifically a negative bias in disfavour of male writers and positive bias in favour of female writers, was reported also by Tankó (2021).

Fairness in Language Assessment

Fairness is considered a key characteristic feature of language testing (Bachman & Damböck, 2018; Bachman & Palmer, 2010). The concept is complex given that "fairness is not a single quality, but is a function of many different aspects of not only the assessment process itself, but also the use of assessments" (Bachman & Palmer, 2010, p. 128). Its various facets are incorporated into the core of the approach proposed by Bachman and Palmer (2010) and are operationalised in the Assessment Use Argument (AUA), which serves as the conceptual framework for language test development and use.

Fairness is operationalised in the AUA in two ways. Bachman and Palmer (2010) discussed fairness as a multifaceted sub-argument level quality articulated in warrants. This may be regarded as the primarily assessment development interpretation of the concept. Bachman and Damböck (2018) elucidated the second dimension of fairness when they considered it primarily from an assessment use perspective as an overall AUA-level quality. As such, fairness refers to "how well each link from students' performance on assessment tasks, to assessment records, to decisions, and to consequences can be supported or justified to stakeholders. If any link cannot be justified, then the assessment *use* is not likely to be fair" (p. 28).

Bachman and Palmer (2010) related fairness with two aspects of the assessment process, equitability of treatment of individuals and absence of bias, both of which are operationalised in the AUA. Equitability of treatment means that each test taker can get informed about what the assessment consists of and what assessment procedures and equipment are used so that they can prepare for it like any other test taker. The location and cost of the assessment must not make it inaccessible to any interested test taker.

Furthermore, all test takers must be assessed under the same conditions suitable for the purposes of the assessment so that they have "comparable opportunities to demonstrate their ability" (p. 128). Their results must be made available in time, and communicated understandably and confidentially to them.

Absence of bias in the assessment process means that the performance and results of all the test takers are determined only by the ability which is being assessed. Their assessment records, which can take the form of scores or marks, for example, must not vary because of such variables as age, native language, ethnicity, religion, or gender. Their assessment records must also be free of bias that may be induced by response format, thematic content (e.g., specialised or cultural), or inconsistencies in scoring due to unsystematic assessor or rater behaviour. Instances of both inequity and bias were documented in the assessment literature. In the next section, the effects of gender in language assessment are explored because these can have a significant effect on the performance of test takers (McNamara, 1996).

Gender Bias in Language Assessment

The assessment of speaking and writing ability is especially prone to being affected by assessment bias given that both types of ability are assessed with the use of a scale with which an examiner is expected to evaluate the language sample elicited from the test taker. The inconsistencies in the behaviour of the examiner are likely to have unfavourable or favourable outcomes for test takers, which may be reflected in their assessment records. Detecting bias in the case of speaking and writing assessment is an intricate process (Brown & McNamara, 2004; McNamara, 1996) as the sources of bias in the examiner and test taker interactions are diverse; however, in accordance with the aims of the current study, this section focuses only on unfairness induced by gender effects. Given the relative scarcity of studies on the effect of gender in the assessment of written performance, first a brief

review of oral assessment is presented to foster a better understanding of this factor and of its potential implications for writing assessment.

In an investigation of variance in test taker performance in a test of spoken interaction, O'Sullivan (2000) found a relationship between the gender of the examiner and the test taker's performance. Except for one of 12 male and female test takers, all performed better and were awarded higher scores by a group of trained raters based on their video-recorded interviews when the interviewer was a woman. The analysis of transcribed interactions revealed systematic gender differences in interviewer language. Furthermore, the analysis of test taker responses showed that the grammatical accuracy of the language produced in interactions with a female interviewer was higher. Therefore, the gender of the interviewer as well as the manner in which they interacted with the test takers impacted the test results. Additional, if limited, similar evidence on the effect of the interlocutors' gender on the test takers' performance was found by Lumley and O'Sullivan (2005). Male test takers performed better when the interlocutor was a man, whereas female test takers performed better when the interlocutor was a woman.

In their study on interviewer behaviour in oral proficiency interviews conducted as part of a national examination in the English language, Reemann et al. (2013) found distinct gender-related differences in the behaviour of interviewers that threatened the validity of the test. Whereas both male and female interviewers used accommodative strategies to assist the test takers, neither adhered to the interlocutor framework closely enough. Although female interviewers were found to follow the protocol to a larger extent, they also asked the test takers additional questions, acting as "a compassionate language teacher" (p. 224) rather than as an objective examiner. The authors recommended interviewer training to address general as well as gender-based differential rater behaviour in oral assessment.

Their suggestion was reinforced by a study on IELTS oral examinations in which O'Loughlin (2002) found no gendered differences in relation to candidate performance. He explained the absence of gender effect with the potential gender neutrality of the assessments tasks, the incapacity of the assessment band scale to register the effect of rater gender, and the training the examiners received as language teachers. This latter implies that they were competent to recognise, categorise, and evaluate the linguistic phenomena targeted by the measured construct.

As the above studies show, the assessment of oral performance was found not to be gender neutral, and the detection of gendered patterns can be carried out through the direct observation and analysis of the interaction between test takers and interviewers. Similarly to transcribed oral proficiency interviews, in the assessment of writing the differential performance of male and female test takers on assessment tasks can be analysed with the help of the student scripts; however, in real-life examination situations the effect of the rater can only be investigated indirectly, for example, through the analysis of the assessment records. The interaction of the rater as a reader with a script may well be similarly dynamic as the interaction of an interviewer with a test taker, but in an everyday examination situation this remains hidden in its entirety. Some insights have been gained through carefully designed empirical studies.

Du et al. (1996) designed a study to detect differential facet functioning in written performance assessment. Although the study was not conducted with university students but with sixth, eighth, and 10th graders, it needs to be mentioned here for two reasons. First, the authors hypothesised that if raters "know the identity of each respondent or ... can guess the respondent's gender or ethnicity, rater bias may occur" (p. 10), and with many-faceted Rasch modelling they demonstrated that more raters were both favourably and unfavourably biased for or against female writers than male writers. Thus the study provided evidence that rater bias was present and detectable in written performance assessment. Second, the

authors provided a definition for differential rater functioning, namely "a situation where individual students with the same underlying ability level have an unequal probability of obtaining the same level of ratings by the raters because of their group membership" (p. 10). This definition is in line with Bachman and Palmer's (2010) discussion of the equitability aspect of fairness.

Gyagenda and Engelhard (2009) conducted a Rasch-based FACETS analysis to investigate the effect of rater, domain, and gender on high school students' writing ability. They found both a statistically significant rater effect, as the total and sub-scores awarded depended on which rater read a script, and a rater by gender interaction effect. The latter was found in the case of the scores awarded with the subscales. This finding partly matches those of the study conducted by Tankó (2021) in the tertiary context, in which one of the sub-scores but also the total score awarded by raters to the argumentative essays written by students showed a significant gendered score patterning in favour of female test takers in spite of the significantly superior linguistic performance of male test takers on the same test task.

Also in the higher education context, Haswell and Haswell (1996) examined the ways in which the raters' evaluation was affected when they knew the test taker's sex. When the raters, who were 64 teachers and students equally divided into male and female groups, read the first of two college student essay scripts from which information about the writer's sex was removed, they automatically made assumptions about and tried to infer the test taker's sex, not without success. This is a confirmation of Ellis et al.'s (2008) finding that judges can make correct guesses about a writer's gender on the basis of their writing style. Consequently, these two studies call into question the findings of Francis, Read, et al. (2003) about lecturers' inability to determine a student's gender based on writing style only. Furthermore, when they were told the sex of the test takers, most raters in Haswell and Haswell's (1996) study gave lower scores for a same-sex essay than for an opposite-sex one. They were also found to be biased against the male test

taker in their holistic rating and were more likely to avoid reference to the sex of the female essay writer in discussions about her essay.

In a study that compared female and male college students' compositions in terms of linguistic accuracy, readability, and conformity to assignment guidelines, Morris (1998) reported that in spite of the higher composition scores of the female students she found no major accuracy and readability differences between the compositions. She explained her finding with how female students prepared their assignments and with the evaluation grid: the female students' close adherence to the assignment guidelines was over-rewarded by the evaluation grid. Morris stated that this called into question the validity of the assessment. Indeed, in her study the raters disregarded arguably more pertinent assessment criteria, which affected the male test taker's scores unfavourably.

As in the case of oral performance assessment, rater training was recommended also for the assessment of writing as an a priori solution. Trace et al. (2017) demonstrated that individual rater bias can be reduced effectively if raters negotiate the scores they award, and argued that this contributes to a test's validation argument. Furthermore, they also found that the discussions on scores involving scoring rubrics helped raters better understand the rubric, which they could also use in their classrooms. This provided evidence for a positive washback effect.

The above review demonstrated that gender differences in non-academic and academic prose are systematic enough to be detectable by both judges and software, and that they are stable across time, cultures, as well as within and across registers and genres. Strong evidence was found that readers can guess the writer's gender with high accuracy solely on the basis of their written output. Although a few studies argued that university lecturers cannot detect the gender of a student based on their written products only, and that students write in a uniform and gender-neutral academic style fostered by institutional expectations, there is strong evidence

for gender effects in the assessment of academic writing ability, which threatens the validity of the assessment.

The current study was conducted to investigate whether the process of assessing academic writing skills with a reading-into-writing integrated guided summary writing task was affected by the gender of the test takers and aimed to answer the following research questions:

I. Are there gender-related lexical and syntactic differences in the academic written production of female and male English major students elicited with the guided summary writing task?

II. Are there gender-related differences between the guided summary scores of female and male English major students?

III. How do the scores as well as the lexical and syntactic features of the academic written production of English major students compare across the argumentative essay and guided summary writing task types?

Method

Participants

The participants were the students who constituted the second cohort in the study conducted on argumentative essays (AE; cf. Tankó, 2021). For that study, data was collected at a large university in Hungary from 445 BA English major students ($n_{Female\ AE} = 347$, 78% and $n_{Male\ AE} = 98$, 22%) in two consecutive years ($n_{Year\ 1} = 223$, $n_{Year\ 2} = 222$). As in the case of the investigation of argumentative essays, in this study on guided summaries (GS) the biological gender of the 221 participants (one GS script had to be excluded as explained below) determined based on their names ($n_{Female\ GS} = 176$, 80% and $n_{Male\ GS} = 45$, 20%) was taken into account due to lack of data on gender identity. As in the two-cohort sample, the ratio of female and male students in the current sample was representative of the

ratio in first-year English major student populations at the university where the data was collected.

At the time of the data collection, the university admitted the students into the BA English major programme with B2 level general English language proficiency certificates (Council of Europe, 2020), which equals TOEFL iBT 87-109 and IELTS 5.5-6.5 scores. Prior to the data collection, the students received academic English language instruction in two academic skills courses. Furthermore, they completed most of their first year content courses in English, which provided additional exposure to academic English and opportunities to practice academic writing in assignments.

Data Collection

The GS scripts analysed in this study were written by the participants in a 90-minute academic English test. The test was a high-stakes test given that its completion was a prerequisite for enrolment into second year courses. It consisted of an argumentative essay (cf. Tankó, 2021) and a guided summary task, was compulsory for all the students enrolled in academic skills courses, was centrally administered, and no use of dictionaries or any other reference materials was allowed. The GS task is a reading-into-writing integrated task in which the students read an input text of about 700 words and then extract as well as paraphrase, using their own words as much as possible, only the ideas relevant to the guiding question set for the task. Finally, they write a coherent piece of text in about 130 words (+/-10%; cf. Tankó, 2019 for sample tasks). The input text was on online teaching and learning.

As the study followed up the investigation of gender effects in the assessment of AEs, the essay scripts written by the students whose summary scripts were analysed in this study were also re-analysed. The topic set for the AE was "obligatory attendance of university lectures" (cf. Tankó, 2021 for further details).

Rating Process

The GS task was designed to measure the participants' academic reading, summarisation, paraphrasing, and writing abilities, as well as their general academic English proficiency. For the purposes of this study, academic writing was defined as "a writer-responsible, referential, often argumentative, transactional, and conventionalised type of writing" (Tankó, 2012, p. 11).

Each script was double-marked independently by two raters using an analytic rating scale that contained four criteria. The raters determined the final scores based on their individual scores. The participants received sub-scores for Task Achievement operationalised by three scales (Completeness, Content, Paraphrasing) on each of which they could score 0-3 points; Coherence and Cohesion (0-3 points) measuring organisation, the use of linking devices, and punctuation as a discourse organising component; and Grammar (0-3 points) focusing on range and accuracy. The participants could score in total 15 points. No rater marked their own students' scripts, and the rating was preceded by a benchmarking and a standardisation session. The names of the students were not masked on the scripts.

The AE scripts were marked by the same raters following the same protocol as in the case of the GS scripts. The students could score in total 15 points. The total score was calculated on the basis of sub-scores for Task Achievement (0-6 points), Coherence and Cohesion (organisation, linking devices, and punctuation used to assist the clear communication of ideas, 0-3 points), Grammar (range and accuracy, 0-3 points), and Vocabulary and Style (lexical range and accuracy and neutral-to-formal style, 0-3 points).

Corpora

The script of one student who wrote a commentary on the input text instead of a guided summary was excluded from the analysis. The remaining

221 guided summaries were typed up, and a corpus consisting of altogether 30,189 running words was created. Since the task does not require a title, if a student gave a title to their summary, it was deleted from the corpus. In turn, the argumentative essay corpus created for this study consisted of 39,238 running words and contained 222 scripts. Two versions of both the GS and AE corpora were created in preparation for the syntactic and lexical complexity analyses: a plain text version for the syntactic analysis, and a lemmatised and part-of-speech (POS) tagged version prepared for the lexical analysis with the Tree Tagger software (Schmid, 1995).

Data Analysis

The GS and AE corpora were analysed with a syntactic complexity analyser software (Lu, 2010, 2011), which computes 23 syntactic indices (e.g., number and mean length of clauses and dependent clauses per sentence and T-unit; mean length of simple and complex T-units per sentence; number of coordinate phrases, complex nominals per T-unit and clause). Next, the output was also analysed with the lexical complexity analyser (Lu, 2012), which generates 25 lexical complexity indices and 9 structure counts (e.g., verb sophistication; number of different words; lexical word, noun, verb, adjective, adverb, modifier variation, type/token ratio). The output can be subsumed under the lexical density (LD), lexical variation (LV), and lexical sophistication (LS) categories.

The students' AE and GS scores and the linguistic complexity indices generated with the linguistic complexity analyser software were subjected to statistical analyses. A Shapiro-Wilk test showed that 39 (63%) out of the 62 variables were not normally distributed. An analysis of homogeneity of variance performed with Levene's test confirmed that the variance of the female and male groups was the same in the case of all the variables. As a consequence, the Mann-Whitney U test was used to compare the male and female students' scores, syntactic complexity, and lexical complexity

variables. However, given that some of the variables were normally distributed, an independent samples t-test was also conducted. Both the nonparametric and parametric test detected statistically significant differences between the same variables. The results obtained with the non-parametric test are reported here. Consequently, effect sizes (r) were also calculated with the formula $r=Z/\sqrt{N}$ (Fritz et al., 2011) and interpreted according to the recommendations of Cohen (1988).

Finally, a subcorpus created with the POS tagged male GS summaries was analysed with the Concord tool of the WordSmith Tools software (Scott, 2012) in order to explore some of those linguistic complexity indices where significant differences were found in context in more detail. The concord tool was used to search for specific lexical items in the subcorpus and to sort the instances of the searched lexical items in order to analyse them together with their co-texts.

Results and Discussion

Differences in Guided Summary Writing Performance across Genders

As shown in Table 1, out of the 62 indices analysed—several of which targeted the same syntactic and lexical features with different measures used in the literature (Lu, 2010, 2011) in order to enhance the reliability of the analysis—significant differences were found between five variables, namely four lexical complexity (Verb sophistication, Verb variation, Adjective variation, Modifier variation) and one syntactic complexity variable (Complex nominals per clause).

Table 1

Differences between the Linguistic Indices of Female- and Male-authored Guided Summaries

	Gen-der	N	Ranks	Me-dian	Mann-Whitney U	*Z*	*p*	*r*
Verb sophistication (LEX)	M	45	130.68	.47	3074.500	-2.315	.021	-0.15572
	F	176	105.97	.41				
Verb variation (LEX)	F	176	115.95	.23	3089.500	-2.282	.022	-0.1535
	M	45	91.66	.22				
Adjective variation (LEX)	M	45	136.98	.16	2791.000	-3.066	.002	-0.20625
	F	176	104.36	.14				
Modifier variation (LEX)	M	45	138.86	.20	2706.500	-3.286	.001	-0.22101
	F	176	103.88	.17				
Complex Nominals/ cl. (SYN)	M	45	128.60	1.36	3168.000	-2.070	.038	-0.13924
	F	176	106.50	1.28				

The effect size of the significant differences found were all small (Cohen, 1988) as they fell between the values of 0.1 and 0.3. Nevertheless, these effect sizes are not trivial because as Cohen (1988) emphasized "many relationships pursued in 'soft' behavioural science are of this order of magnitude" (p. 79). As a consequence, the discussion of the nature of the differences is warranted.

Also noteworthy is the fact that significant differences were only found between four lexical and one syntactic index and even in their case only between a few from the large set of variables investigated. This indicates that the GS task elicited fairly homogenous language samples from both female and male students. Except for one, the complexity indices which showed a difference between the written production of female and male students were lexical complexity indicators.

There were a significantly higher number of lower frequency verbs in the GS scripts written by male students. Contrarily, the GS scripts of female students contained a larger variety of verbs, which confirms the findings reported by large-scale corpus studies conducted on non-academic prose (Argamon et al., 2003; Biber et al., 1988; Newman et al., 2008). This was the only variable where according to the complexity indicators female students significantly outperformed their male peers, and this feature indicated that their language use was less nominal and in this respect less like academic prose (cf. Tankó, 2020 for a discussion of the syntactic features of academic prose).

The remaining two indices showed that male summary writers used a significantly larger variety of adjectives and modifiers. As the modifier index is calculated based on the ratio of the sum of adjectives and adverbs used to the total number of lexical words in a text, the male students' GSs also included more adverbs. The in-text analysis carried out on the 379 adverbs in the male GS corpus showed that in addition to adverbs of manner, male students used a large number of linking, evaluative and degree adverbs, which indicates that they marked the organisation of their texts explicitly and were determinedly engaged in the discussion of the content of the summary—a feature not unlike the confidence in the writing style of students found by Ellis et al. (2008). Their scripts also contained a large number of comparative adverbs and adjectives, which indicates a high degree of propositional content structuring.

The one syntactic complexity index where the guided summaries written by male and female students differed significantly was the complex nominals per clause computed as the number of complex nominals divided by number of clauses. The variables include the following components: "(1) nouns plus adjective, possessive, prepositional phrase, adjective clause, participle, or appositive; (2) nominal clauses; and (3) gerunds and infinitives in subject" (Lu, 2011, p. 44). Given that the use of nominal structures is an indicator of propositional density achieved through form (cf. Tankó, 2016 for

a discussion of this), this finding shows that male students communicated information in a more compact form, as is specifically expected in the case of a summary and generally in the case of academic writing.

The differences in the written production of female and male students identified with automated analyses were not reflected in the scores of the test takers as no significant differences were found between either the sub- or the total GS scores. In line with Cohen's (1988) discussion of the operational definition of effect size, the differences in the scripts are unquestionable, but they are most likely not "perceptible on the basis of casual observation" (p. 79) as opposed to fine-tuned instrumental measurement. It is therefore understandable that the raters did not award extra points for verb variation for female, and for verb sophistication, adjective and modifier variation, or complex nominals per clause for male students. Neither group benefited from the differences identified in the performance data. Similar linguistic complexity results were obtained from the AE scripts reanalysed for this study, but there significant differences were also found in the scores.

Differences in Argumentative Essay and Guided Summary Writing Performance across Genders

The re-analysis of the essay corpus containing only the scripts produced by the second cohort of students, namely those whose GS scripts were investigated in this study, together with the scores awarded to the essays revealed almost exactly the same significant differences for the subsample ($n_{Year\ 2} = 222$) as were found in the original analysis of the entire AE corpus ($n_{Year\ 1+2} = 445$). The two minor differences found were that there was no difference in Type/Token ratio in the subsample, and lexical word diversity was indicated by modifier variation, which overlaps partly with the results obtained during the analysis of the entire sample because this variable includes adverb variation.

Altogether nine significant differences were found between the AE scores and complexity indices in the AE subsample. As presented in Table 2, two of these were differences between AE scores, and, as shown in Table 3, seven differences occurred between complexity indices. Therefore, differences were identified between one syntactic complexity index (Complex Nominals/Clause) and six lexical complexity indices. As in the original AE analysis, the syntactic and lexical complexity indicators of the essays in the subsample written by male students were better, whereas the Task Achievement and Total AE scores of female students were significantly higher.

Probably the most notable outcome of the comparison of the AE and GS indices is that unlike in the case of the essays, no differences in scores were found in the case of the summaries. The linguistic complexity indices generated based on the essay scripts clearly indicated that it was the male students' essays that were significantly better in terms of seven linguistic variables; nevertheless, female candidates were awarded higher AE scores. In contrast, the rating of the summaries was not biased in favour of either group of test takers and was probably a more accurate record of the performances if we recall Cohen's (1988) comment about the differences that can be detected by means of observation.

Table 2

Differences between the Argumentative Essay Scores Obtained by Female and Male Writers

	Gender	N	Ranks	Median	Mann-Whitney U	*Z*	*p*	*r*
Task Achievement	F	176	116.57	4	3073.5	-2.723	.006	-0.18
	M	46	89.80	4				
Total Score	F	176	115.38	12	3281	-2.001	.045	-0.13
	M	46	94.34	11				

Table 3

Differences between the Linguistic Indices of Female- and Male-authored Argumentative Essays

	Gen-der	N	Ranks	Me-dian	Mann-Whitney U	*Z*	*p*	*r*
Verb sophistication (LEX)	M	45	130.68	.47	3074.500	-2.315	.021	-0.15572
	F	176	105.97	.41				
Verb variation (LEX)	F	176	115.95	.23	3089.500	-2.282	.022	-0.1535
	M	45	91.66	.22				
Adjective variation (LEX)	M	45	136.98	.16	2791.000	-3.066	.002	-0.20625
	F	176	104.36	.14				
Modifier variation (LEX)	M	45	138.86	.20	2706.500	-3.286	.001	-0.22101
	F	176	103.88	.17				
Complex Nominals/ cl. (SYN)	M	45	128.60	1.36	3168.000	-2.070	.038	-0.13924
	F	176	106.50	1.28				

The explanation for the lack of bias in the case of the GSs may be provided by the constructs that are measured and the way they are operationalised in the rating scale. The GS rating scale is not only more complex than the one that was used to rate the AEs, but it more effectively operationalises the constructs and the interconnected sub-constructs (e.g., both vocabulary and grammar are related to paraphrasing as well as to academic writing). Unlike the scale in Morris' (1998) study, by being focused on the relevant construct dimensions, the GS scale does not make possible the over-rewarding of construct-irrelevant features. Furthermore, although to apply the GS scale to student scripts systematically does require more effort, at the same time it also leaves less room for subjective decisions.

Even if the differences detected by the software are due to such linguistic variations across scripts that may elude the eyes of a rater, they provide valuable information for test developers and users. It is reassuring that fewer differences were found in the case of the summary scripts. The more uniform written production may additionally be due to the effect of the task type, which was found (Tankó, 2020) to elicit academic prose effectively even from low-achiever students. Moreover, the fact that several of the variables in the case of which the male student's written production was significantly better in both the AE and GS corpora overlap confirms the earlier finding (Argamon et al., 2003) that gendered language use is indeed manifested across genres.

The results of this study further reinforce the findings of Tankó (2020) and the suggestion that the GS task is the better option for the assessment of academic writing skills. Both the AE and GS scripts of male students were characterised by greater verb variation and more frequent use of sophisticated verbs. Verb variation is more characteristic of spoken discourse and was found to be especially characteristic of the essays written by low-achiever students (Tankó, 2020). The AE task provides more freedom for writers than the GS task, in the case of which not only are the propositional content and structure given, but the length specification also leaves less room for elaborate and idiosyncratic language use. The more restrictive nature of the GS task is most likely also the reason why the significant gender difference in lexical variation found in AE scripts was not detected in the GS scripts.

Interestingly, although both the AE and the GS tasks can be considered such conventional academic genres used in educational contexts that are highly restricted, for example, in terms of audience, purpose, structure, and academic register, gendered language use occurred in both the AE and GS corpora. This seems to contradict the suggestion that the more conventional a genre is, the fewer gender-specific characteristics it has (Newman et al., 2008). It may have been the case that the students have not

yet managed to master the gender-neutral academic writing style that Francis, Robson, et al. (2001) described in their study.

Finally, adjective variation was more pronounced in the GS scripts of male students as indicated by the 709 adjectives they used. This finding is in accordance with the assertion in Ellis et al.'s (2008) synthesis of gender-specific characteristics of academic writing that male academic writers tend to use more adjectives in their writing. The most frequently used adjectives appeared in structures with discourse organising function (e.g., *several*/*many*/*main* + advantages/disadvantages), *able to* was often used instead of the modal *can*, and there was a range of less frequent adjectives that did not appear in the source text and therefore were not lifted items (e.g., *accessible*, *available*, *personal*, or *widespread*).

Conclusion

This study was conducted as a follow-up to an investigation on one of two tasks in a high-stakes written academic skills test, the argumentative essay writing task, which was found to elicit output that varied with the test takers' gender and that negatively affected score inferences (Tankó, 2021). The current study proposed to investigate whether the second task in the written academic skills test, the guided summary writing task, also elicited differential written academic performance from male and female students and whether this was reflected in the scoring decisions made by raters. Both studies (Tankó, 2021 and the current one) were therefore motivated by concerns about potential threats to the validity of the academic skills test originating from academic language use by gender group and rater by gender group interactions. To this end, the same discourse-based research approach was applied in this study as in the previous one (Tankó, 2021): Linguistic complexity indicators generated from the written output of students were compared to detect if there were any differences in the performance of male and female students. The results of the analysis of discourse variables was then expanded with an analysis of the scores

awarded to the guided summaries by a group of trained raters. First, the results obtained from the analysis of the guided summary data were presented and discussed, and then they were contrasted with the results of the argumentative essay analysis.

Based on the findings, the answers to the research questions can be summarised as follows: Gender-related lexical and syntactic differences were identified in the academic written production of female and male English major students' guided summaries. The written production of the two groups differed significantly in the case of five, namely one syntactic and four lexical, variables. These differences, however, did not affect the guided summary scores of the students. The comparison of the findings on the analysis of argumentative essays and guided summaries revealed that whereas gender-specific differences occurred in both the essays and summaries, the written production of female and male students differed in terms of more variables in the case of the argumentative essays. Male students were found to perform better on all the variables where significant differences were found in the essays and summaries except for one summary variable. In the case of the essay, in spite of the better written production of male students, the task achievement and final scores of female students were significantly higher. No such difference was found in the case of the summary scores.

The findings prompt a number of suggestions for teachers and assessors of English academic writing skills. First, in spite of the popular notion reflected in literacy literature that female writers produce better written output than their male peers at most stages of their education, including tertiary education, this study confirmed once more that English major male university students can write the same quality, and in certain respects better quality, academic prose than female students. This is reassuring for academic writing instructors who despite this inadvertent stereotyping seek to focus on practicing in their courses academic writing that is free of gender traits in order to enable students to write effective

writer-responsible academic prose that fulfils its referential, rhetorical, and transactional functions in full observance of disciplinary conventions.

The guided summary scores revealed that unlike the argumentative essay rating scale, the guided summary rating scale operationalised the targeted constructs in a way that did not allow for construct irrelevant variance to occur and be detectable in the scores. The analyses of the guided summary test scores in relation to the analyses of the linguistic complexity of the students' summaries showed that the differential performance of male and female students did not have a significant impact on the rating process. As a result, the assessment of the academic writing skills of English majors with the guided summary writing task was found to be fair in respect to bias that could have been induced by gender, and therefore it presented no threat to the validity of the test. Nevertheless, as findings in general from research on bias in language assessment and in particular from the study on argumentative essays (Tankó, 2021) suggested, teachers and raters of academic writing must be trained to be aware of key linguistic features of academic prose not merely at an English composition teacher level but at the level of an academic writing instructor who is familiar with EAP literature and who regularly reads, writes, evaluates, and reviews English academic prose. They must also be trained to apply rating criteria systematically during the rating process and encouraged to engage in score resolution discussions with other raters during rater trainings or the rating process itself so as to develop a thorough understanding of the academic writing construct as operationalised in the rating scale. As prior research evidence showed (Trace et al., 2017), this is very likely to have a positive washback effect on their teaching practice.

One main limitation of this study was that it could only group the participants according to their biological group membership as determined at birth. This dichotomous view ignored the continuous aspect of gender as a trait. Furthermore, the not-source-based argumentative essay task that was used in the study may not be the kind of academic writing task that provides

the amount of institutional constraints on language necessary for the production of gender-trait-free writing. Should this be the case, more evidence is now available that the integrated reading-into-writing guided summary task is better suited for practicing and assessing academic prose production. Future studies could investigate the effect of rater gender on the rating process.

References

Argamon, S., Koppel, M., Fine, J., & Shimoni, A. R. (2003). Gender, genre, and writing style in formal written texts. *Text—Interdisciplinary Journal for the Study of Discourse*, *23*(3), 321–346. https://doi.org/10.1515/text.2003.014

Bachman, L. F., & Damböck, B. (2018). *Language assessment for classroom teachers*. Oxford University Press.

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

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.

Brown, A., & McNamara, T. (2004). "The devil is in the detail": Researching gender issues in language assessment. *TESOL Quarterly*, *38*(3), 524–538. https://doi.org/10.2307/3588353

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.

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

Du, Y., Wright, B. D., & Brown, W. L. (1996, April 8-12). *Differential facet functioning detection in direct writing assessment*. [Paper presentation]. Annual Meeting of the American Educational Research Association, New York, NY. https://files.eric.ed.gov/fulltext/ED400293.pdf

Ellis, L., Hershberger, S., Field, E., Wersinger, S., Pellis, S., Geary, D., Palmer, C., Hoyenga, K., Hetsroni, A., & Karadi, K. (2008). *Sex differences: Summarizing more than a century of scientific research.* Psychology Press.

Francis, B., Read, B., Melling, L., & Robson, J. (2003). University lecturers' perceptions of gender and undergraduate writing. *British Journal of Sociology of Education*, *24*(3), 357–373. https://doi.org/10.1080/01425690301891

Francis, B., Robson, J., & Read, B. (2001). An analysis of undergraduate writing styles in the context of gender and achievement. *Studies in Higher Education*, *26*(3), 313–326. https://doi.org/10.1080/03075070120076282

Fritz, C. O., Morris, P. E., & Richler, J. J. (2011). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*, 2–18. https://doi.org/10.1037/a0024338

Gyagenda, I. S., & Engelhard, G., Jr. (2009). Using classical and modern measurement theories to explore rater, domain, and gender influences on student writing ability. *Journal of Applied Measurement*, *10*(3), 225–246.

Haswell, R. H., & Haswell, J. T. (1996). Gender bias and critique of student writing. *Assessing Writing*, *3*(1), 31–83. https://doi.org/10.1016/S1075-2935(96)90004-5

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://www.jstor.org/stable/41307615

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *Modern Language Journal*, *96*(2), 190–208. https://doi.org/10.1111/j.1540-4781.2011.01232_1.x

Lumley, T., & O'Sullivan, B. (2005). The effect of test-taker gender, audience and topic on task performance in tape-mediated assessment of speaking. *Language Testing*, *22*(4), 415–437. https://doi.org/10.1191/0265532205lt303oa

Martin, M. (1997). Emotional and cognitive effects of examination proximity in female and male students. *Oxford Review of Education*, *23*(4), 479–486. https://doi.org/10.1080/0305498970230404

McNamara, T. (1996). *Measuring second language performance*. Longman.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–104). American Council on Education and Macmillan.

Morris, L. (1998). Differences in men's and women's ESL writing at the junior college level: Consequences for research on feedback. *Canadian Modern Language Review*, *55*(2), 219–238. https://doi.org/10.3138/cmlr.55.2.219

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

O'Loughlin, K. (2002). The impact of gender in oral proficiency testing. *Language Testing*, *19*(2), 169–192. https://doi.org/10.1191/0265532202lt226oa

O'Sullivan, B. (2000). Exploring gender and oral proficiency interview performance. *System*, *28*, 373–386. https://doi.org/10.1016/S0346-251X(00)00018-X

Reemann, E., Alas, E., & Liiv, S. (2013). Interviewer behaviour during oral proficiency interviews: A gender perspective. In H. Metslang, M. Langemets, & M.-M. Sepper (Eds.), *Eesti Rakenduslingvistika Ühingu aastaraamat 9* [Estonian Papers in Applied Linguistics 9] (pp. 209–226). Eesti Rakenduslingvistika Ühing. https://doi.org/10.5128/ERYa9.14

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

Schmid, H. (1995). TreeTagger (Version 3.2.3). [Computer software]

Scott, M. (2012). WordSmith Tools (Version 6). [Computer software]

Tankó, Gy. (2012). *Professional writing: The academic context* (Rev. ed.). Eötvös University Press.

Tankó, Gy. (2016). Written summarisation for academic writing skills development: A corpus based contrastive investigation of EFL student writing. In N. Dobrić, E.-M. Graf, & A. Onysko (Eds.), *Corpora in applied linguistics: Current approaches* (pp. 53–78). Cambridge Scholars Publishing.

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

Tankó, Gy. (2020). Eliciting written academic performance: An investigation of task effect on the written production of EFL students. In Cs. Kálmán (Ed.), *DEAL 2020: A snapshot of diversity in English applied linguistics* (pp. 1–37). Eötvös Loránd University.

Tankó, Gy. (2021). Gender-related differences in performing a test-task in academic writing: Insights from performance data on an argumentative essay task. In É. Illés & J. Kenyeres (Eds.), *Changing perspectives: Studies in English at Eötvös Loránd University* (pp. 76–94). Eötvös Loránd University.

Trace, G. J., Janssen, G., & Meier, V. (2017). Measuring the impact of rater negotiation in writing performance assessment. *Language Testing*, *34*(1), 1–20. https://doi.org/10.1177/0265532215594830

8.

Co-development of Self-Assessment and Second Language Writing from a Complex Dynamic Systems Theory Perspective: A Single Case Study

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

Abstract

Self-assessment (SA) has been found to promote second language (L2) writing development (Birjandi & Tamjid, 2012). Previous studies on SA and L2 writing have predominantly adopted a pre-test-post-test research design, providing a static insight into the development of self-assessment abilities (e.g., Zheng et al., 2012). However, the Complex Dynamic Systems Theory (CDST) suggests that development is dynamic and nonlinear (de Bot et al., 2007). Therefore, this study adopted a time-series analysis, a frequently used technique in CDST studies, to trace the co-development of the accuracy of self-assessment and linguistic complexity over one academic semester. One participant composed 10 short argumentative essays during an Advanced Writing course at a Hungarian university. After the composition of the essay, the participant was asked to self-assess her essay using an International English Language Testing System (IELTS)-type writing rubric. This study found that both the accuracy of self-assessment and linguistic complexity indices showed upward trends over one semester. However, the Mean length of sentence index displayed a downward trend, indicating that the participant tended to construct shorter sentences. Both the trajectories of SA and linguistic complexity indices fluctuated over time, which was in line with previous CDST studies on L2 writing. This study yields useful pedagogical implications for writing courses by showing how self-assessment might facilitate L2 writing development.

Keywords: self-assessment, second language writing, Complex Dynamic Systems Theory

Co-development of Self-Assessment and Second Language Writing from a Complex Dynamic Systems Theory Perspective: A Single Case Study

Self-assessment (SA) has been a popular area of research in applied linguistics since the early 2000s as it was found that SA helps learners judge their own language abilities and set new learning goals (Chapelle & Brindley, 2010). Therefore, SA has been used as an instrument to gain insight into language proficiency in the Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2001), the European Language Portfolio (EL), and the Bergen "can-do" project (Hasselgreen, 2000). Furthermore, SA has been positively linked to the development of language skills (Liu & Brantmeier, 2019). Although of the four language skills writing yielded the weakest correlation (Li & Zhang, 2021), SA has been positively associated with second language (L2) writing development (Birjandi & Tamjid, 2012; Zheng et al., 2012).

In the field of language assessment, SA is usually operationalized as the accuracy or consistency of self-assessment (Andrade, 2019). The accuracy of SA is usually calculated by the percentage agreement score between SA and teacher assessment (TA) (Derrick, 2016). L2 writers who are able to more accurately self-assess their own writing abilities show greater improvements in L2 writing (e.g., Birjandi & Tamjid, 2012). To the best of my knowledge, only two studies have investigated the longitudinal development of SA and L2 writing. Both studies collected samples from relatively large populations, rendering it possible to make generalizations about the possible relationship between SA and L2 writing development. However, both studies adopted a pre-test-post-test research design, which provided a static picture about the relationship between SA and L2 writing. Therefore, this preliminary study looks at the dynamic changes in the accuracy of self-assessment and L2 writing by tracing the development of one particular learner over one academic semester. In this study, 10 written samples were collected from one participant. The participant and two expert raters scored the 10 essays by

using a rubric. The scores were correlated, while the essays were analysed by using computational tools.

Self-assessment and Complex Dynamic Systems Theory

Self-assessment

Self-assessment might be considered as a type of feedback (Andrade, 2019). The main goal of SA is the generation of feedback by the writer for the writer to promote learning and make improvements in performance (Andrade, 2019). The main promote learning and make improvements in performance (Andrade, 2019). In the literature of SA, a distinction is usually made between self-assessment and self-evaluation (Andrade & Du, 2007). Self-evaluation refers to the act of making summative judgements about one's own performance usually resulting in a final grade. In contrast, self-assessment refers to learners' appraisal of their own performance without it resulting in a final grade. Andrade (2019) pointed out that SA should be formative to provide opportunity for learners to make adjustments and corrections. Therefore, summative SA can be considered pointless and useless relative to formative SA, which may result in learner development.

Self-assessment has become a popular type of assessment since the shift from the teacher-centred to learner-centred curriculum (Liu et al., 2006). It has been closely linked to learner autonomy (Illés, 2012), authentic assessment, and metacognitive development (Lew et al., 2010). Authentic assessment refers to the learners' evaluation of their methods of learning, which might lead to further autonomy and control over their learning processes (Kraayenoord & Paris, 1997; Lew et al. 2010). SA has also been demonstrated to promote self-awareness (Babaii et al., 2016) and successful performance.

SA has also been found to promote self-regulated learning (Butler, 2018). Since self-regulation plays an important role in L2 writing

development (Kormos, 2012), SA might be a crucial ability in L2 writing development. In an investigation on English majors' self-regulatory control strategy use in academic writing, Csizér and Tankó (2017) found that only a third of the 222 first-year English majors possessed the ability and willingness to monitor their writing processes. However, the participants were found to be motivated to develop their professional writing abilities. The findings of Csizér and Tankó (2017) are important since this study was also conducted at the same institution of tertiary education.

SA has often been criticised due to its lack of validity and reliability (Ashton, 2014; Patri, 2002). A possible way to overcome this concern is to investigate the relationships between SA and objective language performance (Butler, 2018). Numerous studies have looked at the relationship between SA and objective language performance by calculating correlations between SA and professional raters' scores or teacher assessment (TA) scores. Li and Zhang's (2021) meta-analysis revealed that writing abilities had the weakest correlation (.381), preceded by listening (.486), reading (.451), and speaking (.442). Li and Zhang's findings are not surprising since writing is the most complex of the four skills and it requires the largest amount of informed metacognition as well as a wealth of knowledge to structure and implement the message in writing (Gy. Tankó, personal communication, July 16, 2021).

The majority of previous studies on the relationship between SA and writing performance have adopted a cross-sectional research design, providing a static picture regarding the SA-TA correlations (Liu & Brantmeier, 2019; Summers et al., 2019). There have been only two longitudinal studies which investigated whether learners can self-assess their writing performance better after instruction than before instruction. Birjandi and Tamjid (2012) explored the role of SA and peer-assessment (PA) in English as a foreign language (EFL) learners' writing performance. Altogether 157 intermediate EFL students were placed into five different groups. Journal writing as a reflective self-assessment technique was used in

the first experimental group, while in the second one only self-assessment was used. Peer-assessment and both peer- and self-assessment were used in the third and fourth groups, respectively. Finally, in the fifth group only the teacher assessed the students' writing, so the fifth group was the control group. The 157 participants took the Preliminary English Test (PET, 2004) at the outset and at the end of the study. However, Birjandi and Tamjid (2012) did not specify which parts of the PET writing tests the participants completed. Moreover, Birjandi and Tamjid (2012) wrote: "two teacher-made writing tests, at the beginning and at the end of the semester, were given to assess the students' writing performance" (p. 517). Therefore, it is not entirely clear what type of writing activities the participants performed in their study. Another problem with Birjandi and Tamjid's (2012) study is that the order of tasks was not counterbalanced. Therefore, it may have happened that the students seemed to have improved their writing because the task at the end of the study was easier than the task at the beginning of the study. Birjandi and Tamjid (2012) found the largest amount of improvement in the second and the third groups in which self- and peer-assessment were used together with teacher-assessment.

Another longitudinal study on the effects of self-assessment on L2 writing development is the one conducted by Zheng et al. (2012). The 189 freshmen and sophomore students were instructed for eight weeks on how to self-assess their writing. Three writing test tasks were used with three different writing topics. The first two topics were selected from the writing part of the College English Test (CET) Band 4 and the CET Band 6, respectively. The third writing test was created by the instructors and the researchers themselves. The researchers used a Chinese version of the scoring rubric for CET 4-6. Zheng et al. (2012) found that the students could self-assess their writing much better at the end of the study. In addition, instructing students on how to use the scoring rubric contributed to the improvement of self-assessment in writing. Once again, the order of the

writing tasks was not counterbalanced, similarly to Birjandi and Tamjid's (2012) study. Therefore, the improvement in writing can be questioned.

In summary, there is a scarcity of longitudinal studies on self-assessment in L2 writing. Birjandi and Tamjid (2012) and Zheng et al. (2012) adopted a pre-test-post-test research design providing a static picture of the development of self-assessment abilities. Moreover, in both Birjandi and Tamjid's (2012) and Zheng et al.'s (2012) studies the writing tasks were not counterbalanced, which might have affected the students' writing performance. Consequently, there is a need to investigate the longitudinal development of the accuracy of self-assessment over time. In order to fill this research gap, this study traced the changes in the accuracy of self-assessment and L2 writing development of one participant over an academic semester.

Complex Dynamic Systems Theory

In order to make the investigation of the longitudinal development of self-assessment and L2 writing possible, a time-series approach was adopted. Time-series analysis is a type of data analysis which deals with time-series data. In the field of applied linguistics, time-series analysis is a common method in studies that adopt the Complex Dynamic Systems Theory (CDST) approach. It was Larsen-Freeman (1997) who first introduced the idea that language development might be seen as a dynamic nonlinear process. Although Larsen-Freeman's (1997) recommendation seemed promising in its adoption of the "Chaos/Complexity science" in second language development (p. 141), no methods and techniques were offered at that time. Verspoor et al. (2004) were the first to use a time-series analysis to trace the longitudinal development of L2 writing. De Bot et al. (2007) pointed out, furthermore, that different approaches to research are necessary for the study of dynamic systems. In 2011, Verspoor et al. (2011) offered an arsenal of methods and techniques to investigate language development from a CDST perspective.

De Bot et al. (2007) claimed that language might be viewed as a complex dynamic system and language acquisition, or as they termed it "development", might be seen as a dynamic process (p. 7). The researchers pointed out that language development matches some of the key characteristics of dynamic systems, such as inter- and intra-individual variability, complete interconnectedness of subsystems, the emergence of attractor states, and sensitive dependence on initial conditions. Inter- and intra-individual variability refers to the variation among individuals, that is, there are no two learners who acquire an L2 at the same pace. Furthermore, there is variation within one language learner's development. For example, a learner may acquire 10 new words one day but only eight the next. The emergence of attractor states refers to the relative stability of systems, and in these periods no or little improvement can be found. There are periods when systems show high degrees of variability. However, there are other periods when systems become relatively more stable. These periods are called attractor states (de Bot et al., 2007). Dörnyei (2009) claimed that conglomerates of specific learner characteristics might create attractor states. One of these learner characteristics is boredom, which might create attractor states in a learner's motivational system (Waninge, 2014) or linguistic system (Wind & Harding, 2020). Systems and subsystems are completely interconnected. A change in one system might trigger a change in another closely related system. For example, a change in a learner's milieu might trigger a change in the learner's motivational system (Csizér et al., 2010). Similarly, a change in a learner's motivational system might cause a change in the learner's linguistic system (Wind, 2018).

From a CDST perspective, "development is defined as the growth or increase in level of more developmentally advanced or complex variables and the decline or decrease of less developmentally advanced variables" (Verspoor & van Dijk, 2011, p. 85). A CDST perspective makes it possible to describe how different components interact over time. In second language development, a language learner's vocabulary acquisition might be

connected to the acquisition of new syntactic patterns. For example, a language learner needs to acquire the past participle form of the verbs in order to use the passive structure. From a CDST perspective, variables are called growers. However, investigating the relationship between different growers makes sense only if there is a meaningful relationship between them (Verspoor & van Dijk, 2011). Growers might have three different types of relationships: (1) supportive, (2) competitive, and (3) conditional. Supportive relationship refers to the parallel development of two or more growers. In contrast, competitive relationship denotes the alternating development of two or more growers. For example, from Time x to Time y one grower increases while the other grower decreases. Finally, a conditional relationship refers to the precursor interaction of two or more growers: one specific grower might start increasing only when another grower reaches a specific level of development. For example, a language learner needs to acquire a certain amount of lexical items in order to create a sentence. In other words, lexical development is a precursor to syntactic development at the beginning of language development.

Research Questions

There have been numerous studies investigating the relationship between self-assessment and L2 writing. However, the majority of studies have been cross-sectional and provided a single snapshot of the interaction between SA and L2 writing. The only two exceptions were Birjandi and Tamjid (2012) and Zheng et al.'s (2012) studies. However, both Birjandi and Tamjid (2012) and Zheng et al. (2012) adopted a pre-test-post-test research design, which rendered it impossible to uncover the dynamics of the interactions between SA and L2 writing development. In contrast, a CDST perspective makes possible the longitudinal investigation of SA and L2 writing. This study looks at the longitudinal interaction between the accuracy of SA and L2 writing over one semester. The following research questions (RQ) are answered in this study:

RQ1: How do the English major participant's self-assessment ability and L2 writing develop over one semester?

RQ2: How do the English major participant's self-assessment ability and L2 writing interact over one semester?

In order to answer the RQs, time-series analyses were used to trace one participant's self-assessment abilities and L2 writing development over one semester.

Methodology

This study adopted a single case study research design to trace the development of self-assessment and L2 writing. In this section, the participant, the research setting, the data collection, and the data analyses will be discussed.

The Participant

The participant was selected from an Advanced Writing (AW) course at a large university in Budapest, Hungary. The participant was a female 23-year-old Hungarian English major student. She was in the third-year of the Undivided Teacher Training Master Program. Prior to attending the AW course, she successfully completed the Academic Skills 1 and 2 courses. At the end of the Advanced Writing course, the participant was contacted by email to ask permission for the use of her written samples produced for the course for research purposes. The participant agreed and gave her consent.

The Advanced Writing Course

The Advanced Writing (AW) course is an obligatory course for English major and minor students. The prerequisite to take the Advanced Writing course is the successful completion of the Academic Skills 1 and 2 courses in the first year of the BA in English and American studies

programme. During the Academic Skills 1 and 2 courses, the students master the paraphrasing, summarising, and synthesising skills in addition to a set of skills related to the completion of single- and multiple-input integrated academic writing tasks (Tankó, 2019). The Advanced Writing courses are built upon these three skills; however, the instructors of the courses have the choice to decide on the syllabus of the AW course.

In this study, the instructor used Wallwork's (2016) *English for Writing Research Papers* to design the syllabus for the AW course. The participants of the AW course were taught how to prepare and structure their essays, increase readability, and reduce the number of errors that they might make in English by writing concisely, with no redundancy and no ambiguity. The students were instructed as to how to highlight their claims, avoid plagiarism, and choose the correct tenses and style.

Data Collection

Altogether 10 argumentative essays were collected from one participant in the Advanced Writing course. The topic of the writing prompts was related to applied linguistics since the course was advertised by the Department of English Applied Linguistics. However, I limited the field of applied linguistics to second language acquisition as it was presumed that each learner had had some experience with language learning. The essays were written as home assignments, and the students were required to submit them via Moodle. The instructor corrected the linguistic errors, and highlighted those sentences which could have been split up, lexical items which were ambiguous, and redundant phrases and clauses.

After the composition of the essays, the participants were required to fill in the self-assessment rubric positioned at the bottom of the MS Word document template in which the students submitted their essays. The categories of the rubric were adopted from the IELTS Academic Writing Task 2. However, the scores of the rubric were converted to the Common

European Framework of Reference (CEFR) scale (see Appendix). In the first Advanced Writing class, the instructor and the participants looked at self-assessment grids and discussed the can-do statements related to writing. The students were asked to check the self-assessment grid every time they fill in the self-assessment rubric after finishing the composition of the essays.

Data Analysis

The argumentative essays were scored by the instructor of the Advanced Writing course. In order to avoid being influenced by the self-assessment rubrics, the instructor first removed the self-assessment rubrics from the submitted documents. Second, the instructor scored the essays and entered the scores into the MS Excel spreadsheet program. The essays were also scored by a colleague of the instructor. The inter-rater reliability score was r = .90 between the two raters and disagreements were resolved via discussion to reach a consensus. The self-assessment scores were also entered into the MS Excel spreadsheet program and a percentage agreement score was calculated for each week (Derrick, 2016). The percentage agreement scores were plotted and moving averages trend lines were added to the trajectories.

Task Response is the first criterion of the four in the IELTS Writing Task 2 Band Descriptors, and it measures the extent to which the test taker answers each part of the essay question. Cohesion refers to the connectedness of different ideas presented in the text (Tankó, 2012), whereas coherence refers to the connectedness of mental representations that readers might construct from a specific text (McNamara et al., 2014). Bulté and Housen (2012) defined complexity as:

> a property or quality of a phenomenon or entity in terms of (1) the number and the nature of the discrete components that the entity consists of, and (2) the number and the nature of the relationships between the constituent components.

Lexical complexity includes collocational and lexemic complexity, whereas syntactic complexity comprises sentence, clausal, and phrasal complexity.

Linguistic complexity indices were also calculated to measure the L2 writing development of the participant. The selection of the linguistic complexity indices was in harmony with the categories of the self-assessment rubric. Since the IELTS Writing Task 2 criterion 'Task response' cannot be measured automatically, it is omitted from the data analysis. Coherence and cohesion were calculated with the All-connectives index computed by Coh-Metrix 3.0 (Graesser et al., 2004; Graesser et al., 2011), while vocabulary was operationalized by the Academic Vocabulary List index calculated by the Word and Phrase tool (Gardner & Davies, 2014). Grammatical range was operationalized as the Mean length of sentence (MLS) index computed by the Web-based L2 Syntactic Complexity Analyzer (L2 SCA) (Ai & Lu, 2013; Lu, 2010, 2011; Lu & Ai, 2015).

The theoretical rationale for selecting these specific indices was the following: Expert L2 writers can create texts that are more cohesive. The All connectives (CNCAll) index measures the incidence of all connectives in a text. Connectives are crucial in creating cohesive links between ideas and clauses. Therefore, connectives provide clues about text organisation (Cain & Nash, 2011; Crismore et al., 1993; Longo, 1994; Sanders & Noordman, 2000; van de Kopple, 1985). As L2 writers acquire the linguistic features of academic writing, they tend to use more academic vocabulary. It was expected that the participant of this study would use more words from the Academic Vocabulary List (AVL; Gardner & Davies, 2014). The AVL is derived from a 120-million-word academic sub-corpus of the Corpus of Contemporary American English (COCA). Because Wallwork (2016) recommended splitting up long sentences, it was expected that the participants of the AW course would create shorter sentences over time. In other words, the Mean length of sentence index would decrease over the course of one semester.

The linguistic complexity indices were plotted, and moving average trend lines were added to trajectories (Verspoor et al., 2011). In order to see the relationship between trajectories of the self-assessment percentage agreement scores and the linguistic complexity indices, moving correlations were calculated and plotted.

Results

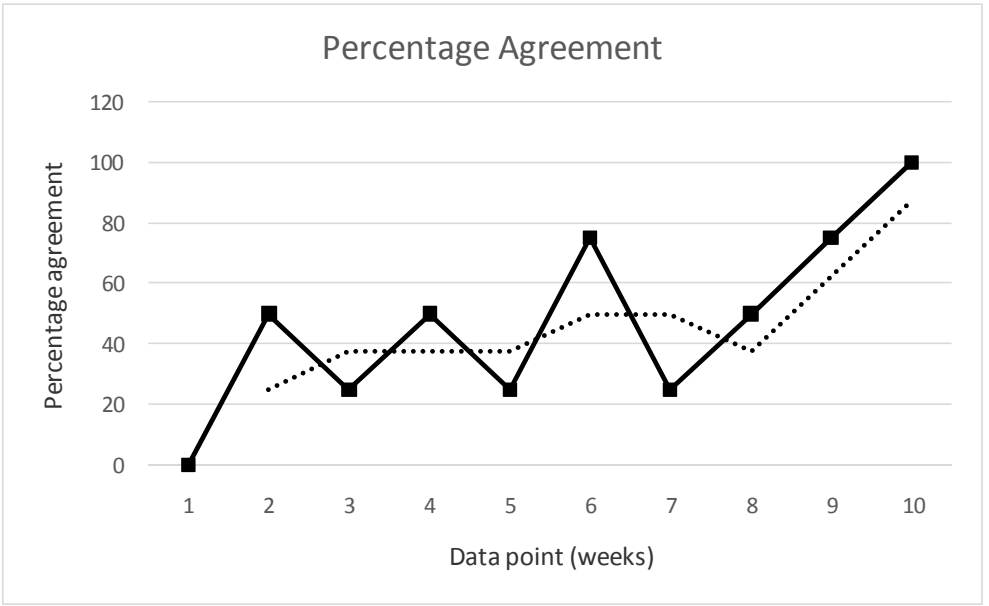
This study investigated the development of one participant's self-assessment ability and L2 writing over one semester.

The Development of Self-assessment

Figure 1 shows the percentage agreement scores over the semester. At data point 1, there was no agreement between the participant and the teacher assessment scores. However, at data point 2 the percentage agreement score was 50%. Both the student and the teacher gave a score 4 for the Task response and the Grammatical range and accuracy categories. However, at data point 3, the percentage agreement score dropped from 50% to 25% because the student and the teacher agreed only in one category (Cohesion and coherence). At data point 4, the percentage agreement scored increased again to 50%. At data point 4, the Coherence and cohesion and the Vocabulary categories received the same scores from the student and the teacher. Interestingly, the same pattern can be seen between data points 4 and 5 as between data points 2 and 3: The percentage agreement score dropped from 50% to 25% again. However, at data point 6 the percentage agreement score increased from 25% to 75%, which meant that the student and the teacher agreed on three (Task response, Vocabulary and Grammar) out of four categories. At data point 7, the percentage agreement score dropped again from 75% to 25%. However, between data points 7 and 10 a gradual increase can be observed from 25% to 100%.

Figure 1

Percentage Agreement Scores



A moving average (2-period) trend line, shown in dotted black line, was added to the percentage agreement trajectory to see the developmental trend over time. The moving average trend line shows a clear upward trend, suggesting the participant tended to self-assess her essays more accurately over the course of the semester. Table 1 shows the percentage agreement scores for the Task response, Coherence and cohesion, and Vocabulary and Grammatical range and accuracy.

Table 1

Percentage Agreement

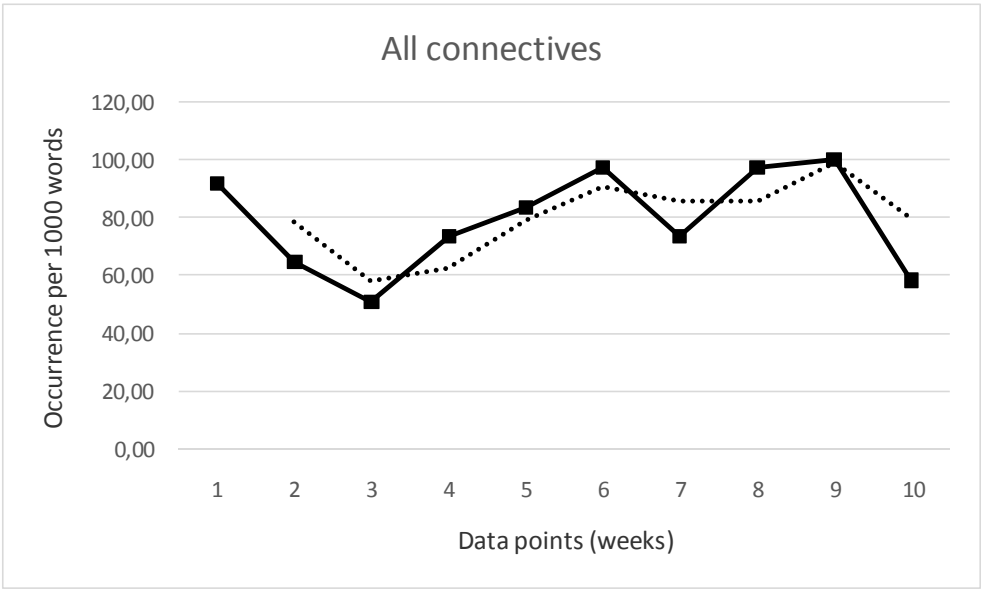
Time	1	2	3	4	5	6	7	8	9	10
Self-assessment										
Task response	5	4	3	3	3	4	3	4	3	4
Coherence and cohesion	4	3	4	4	3	3	3	3	4	4
Vocabulary	4	3	3	4	4	4	4	3	4	4
Grammatical range and accuracy	4	4	3	3	3	4	4	4	4	4
Teacher assessment										
Task response	4	4	5	4	4	4	5	4	4	4
Coherence and cohesion	3	4	4	4	4	4	5	4	4	4
Vocabulary	3	4	4	4	4	4	4	4	4	4
Grammatical range and accuracy	3	4	4	5	4	4	5	4	4	4
Percentage agreement	0	50	25	50	25	75	25	50	75	100

The Development of Cohesion

Figure 2 shows the development of the All connectives (CNCAll) index. The trajectory of the CNCAll index gradually decreased between data points 1 and 3, reaching its lowest point at data point 3 (50.91). However, between data points 3 and 6 a gradual increase can be observed in the trajectory of the CNCAll index. The gradual increase is interrupted by a small drop at data point 7. However, the increase of the CNCAll index continues between data points 8 and 9, reaching its highest point at data point 9 (100.37). Interestingly, at data point 10, the CNCAll index plummeted again.

Figure 2

All Connectives



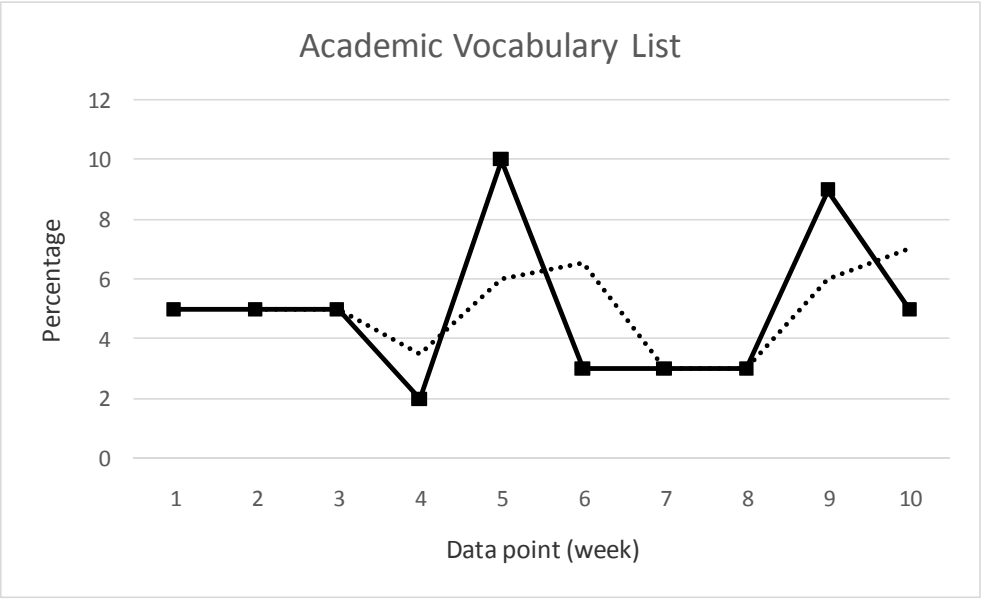
The moving average trend line shows an upward trend between data points 3 and 9, indicating that the participant tended to write essays which were more cohesive and coherent.

The Development of Lexical Complexity

Figure 3 shows the trajectory of the Academic Vocabulary List (AVL). Between data points 1 and 3, the AVL index did not change. As many as 5% of the words in Essays 1, 2 and 3 can be found in the Academic Vocabulary List. However, at data point 4 the AVL index dropped from 5% to 2%, reaching the lowest point of the AVL index. At data point 5, the AVL index rocketed, reaching its highest point (10%). Interestingly, the AVL index plummeted and stabilized between data points 6 and 8. However, the small hump is repeated between data points 8 and 10.

Figure 3

Academic Vocabulary List



The moving average trend line shows a slightly upward trend between data points 4 and 6 and between data points 8 and 10, indicating that the participant started to use more academic vocabulary in her essays. However, the moving average trend line also shows downward trend between data points 6 and 7.

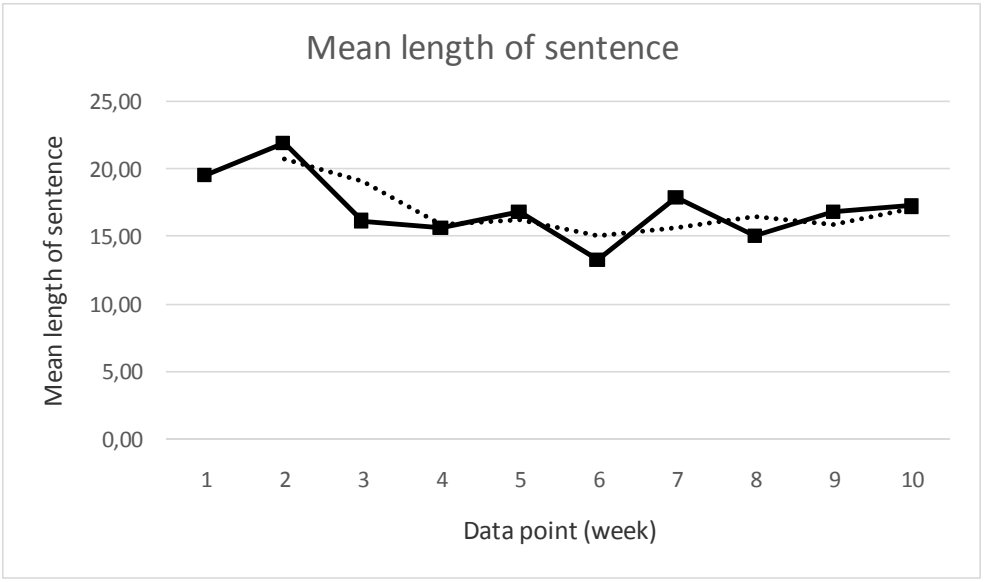
The Development of Syntactic Complexity

Figure 4 shows the development of syntactic complexity operationalized by the Mean length of sentence (MLS) index. The trajectory of the MLS index increased between data points 1 and 2, reaching its highest point at data point 2 (21.92). However, the trajectory of the MLS index gradually declined between data points 2 and 6, except for a small increase at

data point 5. The lowest MLS value (13.30) was measured at data point 6. The trajectory of the MLS index started to increase again at data point 7, but then it dropped once more. Between data points 8 and 10, a gradual increase can be observed.

Figure 4

Mean Length of Sentence



The moving average trend line shows a clear downward trend over the course of the semester. The gradual decline indicates that the participant tended to use shorter sentences over time in line with Wallwork's (2016) recommendation.

Interactions

In order to render the visual inspection of the interactions between the indices possible, the data were normalized, that is, the values were

recalculated to 0 and 1. Figure 5 shows the trajectories of self-assessment and coherence and cohesion over the course of the semester. Between data points 1 and 2, there is an alternating shift. The SA index increased, whereas the CNCAll index dropped. However, between data points 2 and 4 there are parallel shifts. The parallel shifts are interrupted by an alternating shift between data points 4 and 5. However, there are parallel shifts between data points 5 and 9. During this period even the patterns of the two indices are similar. Between data points 9 and 10, there is an alternating shift again. Figure 5 suggests that SA and cohesion develop hand in hand. In dynamic parlance, SA and cohesion might be connected growers.

Figure 5

Self-assessment and Coherence and Cohesion

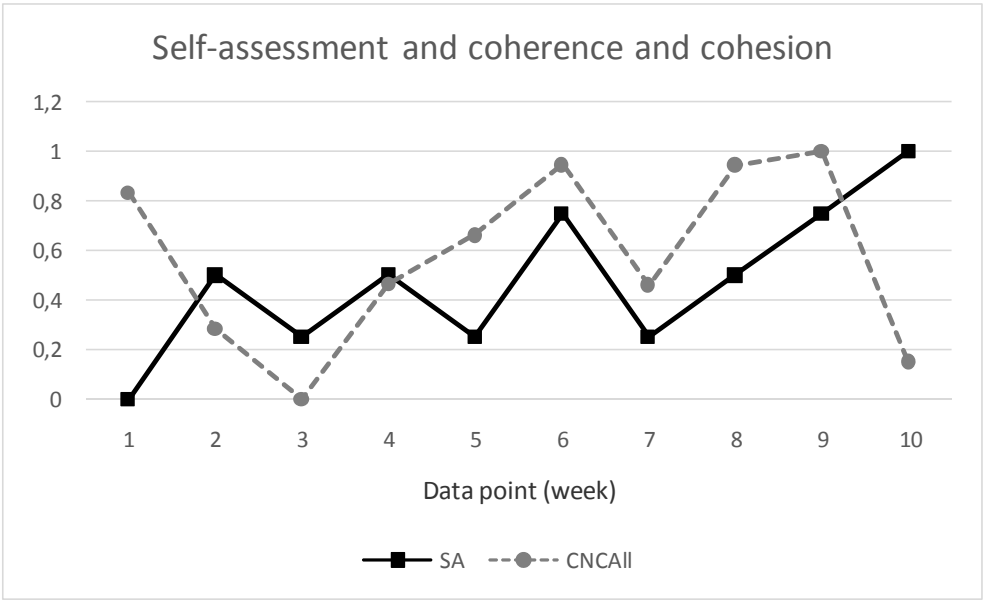


Figure 6 shows the trajectories of the self-assessment and the vocabulary indices. Since there were no changes in the AVL index between data points 1 and 3, the relationship between the two indices is analyzed only between data points 3 and 6. When the SA index increased at data point 4, the AVL index dropped and vice versa, indicating a negative relationship between the two indices. However, a parallel shift can be seen between data points 8 and 9, followed by an alternating shift between data points 9 and 10. Figure 6 indicated a possible negative relationship between SA and vocabulary. In dynamic parlance, SA and vocabulary might be competing growers.

Figure 6

Self-assessment and Vocabulary

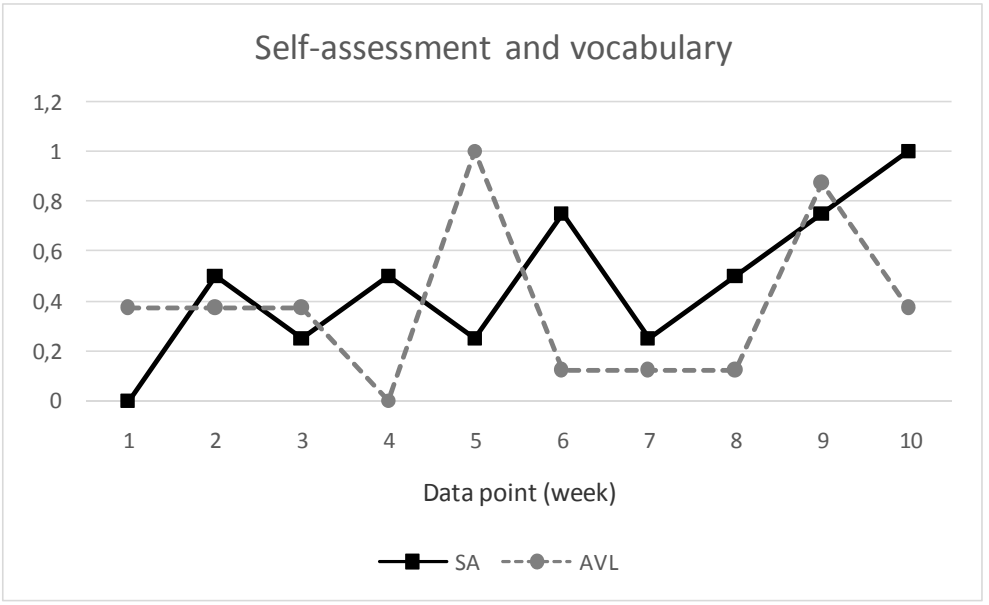
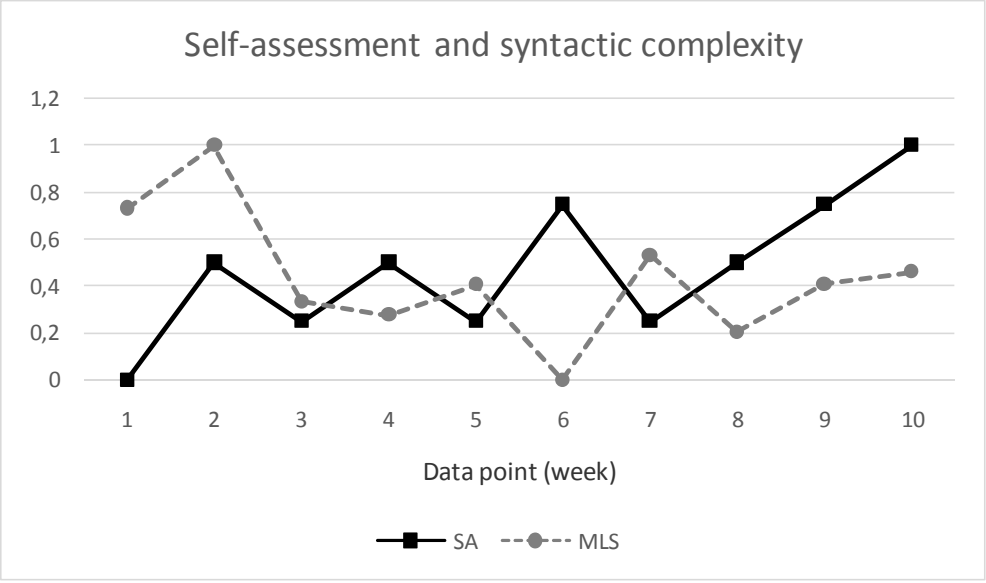


Figure 7 shows the normalized trajectories of self-assessment and syntactic complexity. There are parallel shifts between data points 1 and 3. However, between data points 3 and 8, only alternating shifts can be observed between the SA and the MLS indices. There are clear parallel shifts between data points 8 and 10, indicating a positive relationship between self-assessment and syntactic complexity.

Figure 7

Self-assessment and Syntactic Complexity



In order to explore surface interactions, Spearman correlation coefficients were calculated among the four growers. Table 2 shows the correlation matrix among self-assessment, cohesion and coherence, and vocabulary and grammatical range. The table shows that none of the correlation coefficients was statistically significant. SA correlated positively with cohesion and negatively with vocabulary and grammar. Cohesion

correlated negatively with grammar, but vocabulary correlated positively with grammar. The strongest correlation coefficient was found between vocabulary and grammar.

Table 2

Correlation Matrix

		SA	CNCAll	AVL	MLS
SA	*ρ*	1.000	.187	-.098	-.287
	Sig. (2-tailed)	-	.605	.788	.422
	N	10	10	10	10
CNCAll	*ρ*	.187	1.000	.000	-.358
	Sig. (2-tailed)	.605	-	.100	.310
	N	10	10	10	10
AVL	*ρ*	-.098	.000	1.000	.412
	Sig. (2-tailed)	.788	.100	-	.237
	N	10	10	10	10
MLS	*ρ*	-.287	-.358	.412	1.000
	Sig. (2-tailed)	.422	.310	.237	-
	N	10	10	10	10

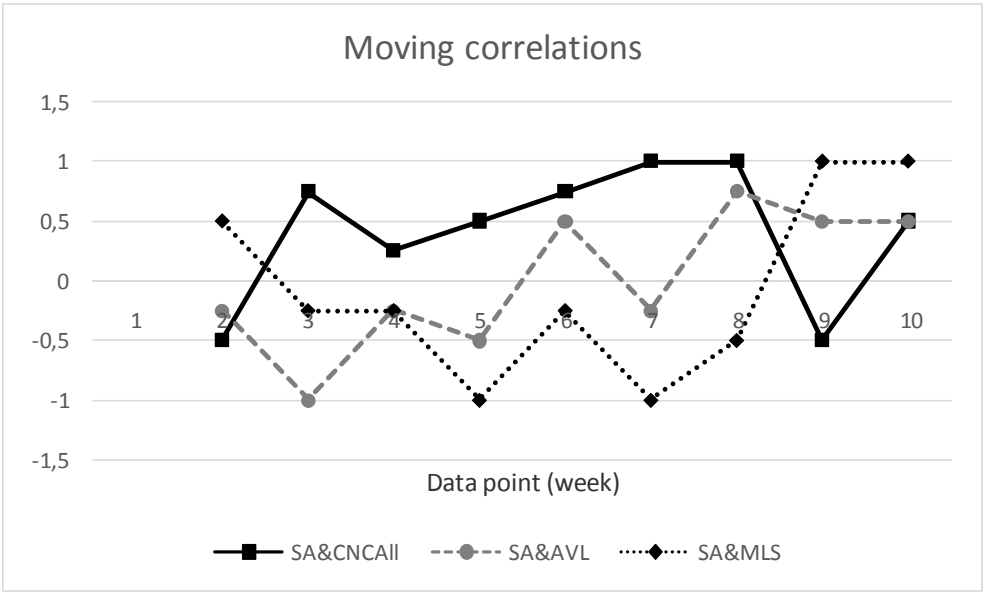
Note. (SA = Self-Assessment, CNCAll = All Connectives, AVL = Academic Vocabulary List, and MLS = Mean Length of Sentence)

In order to explore the underlying interactions, a plot of the moving window of correlations was created. Figure 8 shows the interactions between self-assessment and L2 writing (cohesion, vocabulary, and syntax) as a moving window of 3 measurements. Figure 8 shows that the SA-CNCAll correlation shifted between moderately negative and strongly positive coefficient values between window 2 and 3. However, between window 3

and 10 the correlation was positive, with one peak towards a moderately negative value in the 9th window.

Figure 8

Moving Correlations (Self-assessment-all Connectives, Self-assessment-Academic Vocabulary List, and Self-assessment-mean Length of Sentence)



The SA-AVL correlation was negative between the 2nd and the 5th windows, but then it became positive in the remaining period except for a negative value in the 7th window. The SA-MLS correlation was positive at the beginning and at the end of the semester. However, between the 3rd and the 8th windows the SA-MLS correlation was negative.

Discussion

This study investigated the longitudinal development of self-assessment and L2 writing. More specifically, it looked at the changes in the accuracy of self-assessment and linguistic complexity development over one semester. The study answered two research questions.

How do Self-assessment and L2 Writing Develop over One Semester?

Self-assessment and L2 writing developed nonlinearly over one semester. The trajectories of self-assessment and cohesion and lexical and syntactic complexity indices showed ebbs and flows over time in line with previous studies on L2 writing (Wind, 2014; Wind, 2018; Wind, 2020; Wind & Harding, 2020).

The plot of the moving averages trend line of self-assessment showed an upward trend, indicating improvements in the accuracy of self-assessment. Over one semester, the participant in this study tended to self-assess her essays increasingly more accurately based on the weekly teacher assessments. Birjandi and Tamjid (2012) and Zheng et al. (2012) also found improvements in the accuracy of self-assessment after instruction and practice. Although Birjandi and Tamjid's (2012) and Zheng et al.'s (2012) studies are important in that they show general trends in the development of self-assessment, the underlying dynamics of the development of self-assessment was not explored in them. This single case study clearly suggests that the accuracy of self-assessment does not develop linearly but shows oscillations over a short period of time (one semester).

The moving averages trend line of the cohesion indices (CNCAll) also demonstrated an upward trend. The participant tended to compose essays which were more cohesive and coherent in line with previous studies focusing on cohesion and coherence (Crossley et al., 2016). The moving averages trend line of the mean length of sentence showed a downward trend over one semester. This finding contradicts previous studies which

claim that L2 writers use longer sentences in academic writing (Mazgutova & Kormos, 2015) and more complex nominals (Biber et al., 2011). However, this finding is not surprising if we consider the fact that the participant was instructed from Wallwork's (2016) *English for Writing Research Papers* in which the use of shorter sentences is recommended in academic writing.

How do Self-assessment and L2 Writing Interact over One Semester?

Self-assessment and L2 writing interacted dynamically over one semester. The surface interactions between self-assessment and L2 writing did not show statistically significant correlations. Self-assessment correlated positively with cohesion, indicating that the two constructs developed in unison over one semester. In dynamic parlance, it seems that self-assessment and cohesion are connected growers (Lowie et al., 2011). In contrast, self-assessment correlated negatively with vocabulary and grammar, suggesting that self-assessment and vocabulary and self-assessment and grammar are competing growers (Lowie et al., 2011). The negative competition between self-assessment and vocabulary and between self-assessment and grammar was confirmed by the alternating shifts (Figure 6 and 7).

The polarity and the magnitude of the relationships between self-assessment and L2 writing fluctuated over the semester. None of the interactions remained static over time, conforming to previous studies on the relationship among linguistic complexity indices (Verspoor & van Dijk, 2011).

Conclusion

This study looked at the longitudinal development of self-assessment and L2 writing, more specifically the longitudinal development of the accuracy of self-assessment and linguistic complexity indices. The findings provide some pedagogical implications. Being familiar with the assessment

system and by being exposed to weekly comparisons of self- and teacher assessment, the participant in this study will have tended to focus on specific aspects (cohesion, vocabulary, and grammar) of L2 writing. Using the self-assessment rubric might have contributed to improvements in linguistic complexity in this study. However, it is important to note that the participant received dynamic written corrective feedback (DWCF; Hartshorn et al., 2010). Therefore, improvements in L2 writing might have been triggered by the combination of self-assessment and DWCF. Future studies should explore the effects of DWCF through the use of a time-series research design.

Pedagogical implications suggested in this paper include implementing self-assessment in academic writing courses. Although the findings of this case study cannot be generalized, this study clearly shows that self-assessment and some linguistic complexity indices co-develop over time.

There are several limitations of this study. First, this study adopted a case study research design; therefore, the findings cannot be generalized to the other students of this or subsequent Advanced Writing courses. Second, the rubric used for self-assessment might not be fully harmonized with the specific aims of the course. During the course, the instructor used Wallwork's (2016) *English for Writing Research Papers* in which the participants were advised to pay attention to structuring sentences, paragraphs, breaking up long sentences, being concise and removing redundancy, and avoiding ambiguity, hedging, and criticising. Consequently, future studies should harmonize the rubric with the specific aims of the course and collect data from larger groups, preferably from other Advanced Writing courses.

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

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

Andrade, H. L. (2019). A critical review of research on student self-assessment. *Frontiers in Education*, *4*(87), 1–13. https://doi.org/10.3389/feduc.2019.00087

Andrade, H., & Du, Y. (2007). Student responses to criteria-referenced self-assessment. *Assessment & Evaluation in Higher Education*, *32*(2), 159–181. https://doi.org/10.1080/02602930600801928

Ashton, K. (2014). Using self-assessment to compare learners' reading proficiency in a multilingual assessment framework. *System*, *42*(1), 105–119. https://doi.org/10.1016/j.system.2013.11.006

Babaii, E., Taghaddomi, S., & Pashmforoosh, R. (2016). Speaking self-assessment: Mismatches between learners' and teachers' criteria. *Language Testing*, *33*(3), 411–437. https://doi.org/10.1177/0265532215590847

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

Birjandi, P., & Tamjid, N. H. (2012). The role of self-, peer and teacher assessment in promoting Iranian EFL learners' writing performance. *Assessment & Evaluation in Higher Education*, *37*(5), 512–533. https://doi.org/10.1080/02602938.2010.549204

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

Butler, Y. G. (2018). The role of context in young learners' processes for responding to self-assessment items. *The Modern Language Journal*, *102*(1), 242–261. https://doi.org/10.1111/modl.12459

Cain, K., & Nash, H. M. (2011). The influence of connectives on young readers' processing and comprehension of text. *Journal of Educational Psychology*, *103*(2), 429–441. https://psycnet.apa.org/doi/10.1037/a0022824

Chapelle, C. A., & Brindley, G. (2010). Assessment. In N. Schmitt (Ed.), *An introduction to applied linguistics* (pp. 247–267). Hodder Education.

Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching, assessment.* Cambridge University Press.

Crismore, A., Markkanen, R., & Steffensen, M. S. (1993). Metadiscourse in persuasive writing: A study of texts written by American and Finnish university students. *Written Communication*, *39*, 39–71. https://doi.org/10.1177%2F0741088393010001002

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. *Journal of Second Language Writing*, *32*, 1–16. https://doi.org/10.1016/j.jslw.2016.01.003

Csizér, K., Kormos, J., & Sarkadi, Á. (2010). The dynamics of language learning attitudes and motivation: Lessons from an interview study of dyslexic language learners. *The Modern Language Journal*, *94*(3), 470–487. https://doi.org/10.1111/j.1540-4781.2010.01054.x

Csizér, K., & Tankó, Gy. (2017). English majors' self-regulatory control strategy use in academic writing and its relation to L2 motivation. *Applied Linguistics*, *38*(3), 386–404. https://doi.org/10.1093/applin/amv033

De Bot, K., Lowie, W., & Verspoor, M. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition*, *10*(1), 7–21. https://doi.org/10.1017/S1366728906002732

Derrick, D. J. (2016). Instrument reporting practices in second language research. *TESOL Quarterly*, *50*(1), 132–153. https://doi.org/10.1002/tesq.217

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

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied linguistics*, *35*(3), 305–327. https://doi.org/10.1093/applin/amt015

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, *40*(5), 223–234. http://dx.doi.org/10.3102/0013189X11413260.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, *36*(2), 193–202. http://dx.doi.org/10.3758/BF03195564

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

Hasselgreen, A. (2000). The assessment of the English ability of young learners in Norwegian schools: An innovative approach. *Language Testing*, *17*(2), 261–277. https://doi.org/10.1177%2F026553220001700209

Illés, É. (2012). Learner autonomy revisited. *ELT Journal*, *66*(4), 505–513. https://doi.org/10.1093/elt/ccs044

Kormos, J. (2012). The role of individual differences in L2 writing. *Journal of Second Language Writing*, *21*(4), 390–403. https://doi.org/10.1016/j.jslw.2012.09.003

Kraayenoord, C. E. V., & Paris, S. G. (1997). Australian students' self-appraisal of their work samples and academic progress. *The Elementary School Journal*, *97*(5), 523–537. https://www.jstor.org/stable/1002267

Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics*, *18*(2), 141–165. https://doi.org/10.1093/applin/18.2.141

Lew, M. D. N., Alwis, W. A. M., & Schmidt, H. G. (2010). Accuracy of students' self-assessment and their beliefs about its utility. A*ssessment & Evaluation in Higher Education*, *35*(2), 135–156. https://doi.org/10.1080/02602930802687737

Li, M., & Zhang, X. (2021). A meta-analysis of self-assessment and language performance in language testing and assessment. *Language Testing*, *38*(2), 189–218. https://doi.org/10.1177/0265532220932481

Liu, H., & Brantmeier, C. (2019). "I know English": Self-assessment of foreign language reading and writing abilities among young Chinese learners of English. *System*, *80*, 60–72. https://doi.org/10.1016/j.system.2018.10.013

Liu, R., Qiao, X., & Liu, Y. (2006). A paradigm shift of learner-centered teaching style: Reality or illusion? *Journal of Second Language Acquisition and Teaching*, *13*(0), 77–91.

Longo, B. (1994). The role of metadiscourse in persuasion. *Technical Communication*, *41*, 348–352. http://www.jstor.org/stable/43090348

Lowie, W., Caspi, T., van Geert, P., & Steenbeek, H. (2011). Modeling development and change. In M. Verspoor, K. de Bot, & W. Lowie (Eds.), *A dynamic approach to second language development: Methods and techniques* (pp. 99–122). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.29.06low

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

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

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

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

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press. https://doi.org/10.1017/CBO9780511894664

Patri, M. (2002). The influence of peer feedback on self- and peer-assessment of oral Skills. *Language Testing*, *19*(2), 109–131. https://doi.org/10.1191%2F0265532202lt224oa

Sanders, T. J. M., & Noordman, L. G. M. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes*, *29*, 37–60. https://doi.org/10.1207/S15326950dp2901_3

Summers, M. M., Cox, T. L., McMurry, B. L., & Dewey, D. P. (2019). Investigating the use of the ACTFL can-do statements in a self-assessment for student placement in an intensive English program. *System*, *80*, 269–287. https://doi.org/10.1016/j.system.2018.12.012

Tankó, Gy. (2012). *Professional writing: The academic context* (2nd rev. ed.). Eötvös University Press.

Tankó, Gy. (2019). *Paraphrasing, summarising and synthesising skills for academic writers: Theory and practice* (2nd rev. ed.). Eötvös University Press.

Van de Kopple, W. J. (1985). Some exploratory discourse on metadiscourse. *College Composition and Communication*, *36*, 82–93. https://doi.org/10.2307/357609

Verspoor, M., de Bot, K., & Lowie, W. (2004). Dynamic systems theory and variation: A case study in L2-writing. In M. Hannay, H. Aertsen, & R. Lyall (Eds.), *Words in their places. A Festschrift for J. Lachlan* (pp. 407–421). Free University Press.

Verspoor, M. H., de Bot, K., & Lowie, W. (2011). *A dynamic approach to second language Development: Methods and techniques.* John Benjamins Publishing Company. https://doi.org/10.1075/lllt.29

Verspoor, M., & van Dijk, M. (2011). Visualizing interactions between variables. In M. Verspoor, K. de Bot, & W. Lowie (Eds.), *A dynamic approach to second language development: Methods and techniques* (pp. 85–98). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.29.05ver

Wallwork, A. (2016). *English for writing research papers.* Springer International Publishing. https://doi.org/10.1007/978-3-319-26094-5

Waninge, F. (2014). Motivation, emotion and cognition: Attractor states in the classroom. In Z. Dörnyei, P. D. MacIntyre, & A. Henry (Eds.), *Motivational dynamics in language learning* (pp. 195–213). Multilingual Matters. https://doi.org/10.21832/9781783092574-016

Wind, A. M. (2014). Second language writing development from a dynamic systems theory perspective. In B. P. Olmos-Lopez, J. Huang, & J. Almeida (Eds.), *Proceedings of the 8th Lancaster University Conference in Linguistics & Language Teaching* (pp. 90–123). University of Lancaster. http://www.lancaster.ac.uk/fass/events/laelpgconference/papers/v08/Attila-M-Wind.pdf

Wind, A. M. (2018). *Second language writing development from a complex dynamic systems theory perspective: A multiple case-study of Hungarian learners of English.* Lancaster University. https://doi.org/10.17635/lancaster/thesis/519

Wind, A. M. (2020). A dynamic usage-based approach to the longitudinal development of verb argument constructions in second language writing: Two case studies. In Cs. Kálmán (Ed.), *DEAL 2020: A snapshot of diversity in applied linguistics* (pp. 39–70). Eötvös University Press. https://www.eltereader.hu/media/2020/12/Kálmán-Csaba_WEB.pdf#page=58

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

Zheng, H., Huang, J., & Chen, Y. (2012). Effects of self-assessment training on Chinese students' performance on college English writing tests. *Polyglossia*, *23*, 33–42.

Appendix

SELF-ASSESSMENT

After completing the writing task, please rate your essay based on the following criteria. *(5=excellent, 4=good, 3=mediocre, 2=poor, 1=bad)*

Task response	B1	B2	C1	C1-C2	C2
Coherence and cohesion	B1	B2	C1	C1-C2	C2
Vocabulary	B1	B2	C1	C1-C2	C2
Grammatical range and accuracy	B1	B2	C1	C1-C2	C2

9.

Changes in Students' Motivation, Autonomy, Self-efficacy, and Emotions across a School year: Examining Possible Effects of Switching to Online Education

Ágnes Albert
albert.agnes@btk.elte.hu

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

Abstract

In our longitudinal study, we set out to investigate changes in secondary school students' motivation, autonomy, self-efficacy, and emotions across a schoolyear with the help of a questionnaire survey. While we collected our first set of data in the autumn of 2019, our next point of data collection took place in the spring of 2020, after the lockdown of schools due to the worldwide pandemic. Therefore, besides offering a longitudinal perspective, the responses of our 187 participants studying in three different secondary schools of Budapest might also provide some information about the aftermath of the sudden changes that took place after schools switched to online education. While analyzing the changes in each variable, we found an increase in autonomy and a decrease in motivation and hope by the end of the schoolyear. As regards the relationships of the variables, the main tendencies suggest that students' autonomous learning behavior was mainly determined by their motivation and feelings of hope although the picture becomes more complex by the end of the schoolyear. The pedagogical implications of the reported findings are somewhat complex as it is uncertain whether they can be attributed to changes that usually occur within a schoolyear or to the sudden switch to online education.

Keywords: motivation, autonomy, self-efficacy, emotions, longitudinal approach

Changes in Students' Motivation, Autonomy, Self-efficacy, and Emotions across a School year: Examining Possible Effects of Switching to Online Education

Learning a new language after someone has mastered their mother tongue is usually a time-consuming and cumbersome task, especially in a primarily monolingual country such as Hungary. Learning a new language is also a very complex process; it is by no means a piecemeal accumulation of bits of knowledge carefully selected by teachers and easily and orderly acquired by learners. (VanPatten & Benati, 2015). Although identifying the individual difference (ID) variables responsible for differences in language learning success has a long research tradition (Dörnyei, 2005; Dörnyei & Ryan, 2015; Skehan, 1989), such variables have typically been investigated in isolation, and researchers frequently aimed to identify the most important determinants of success (Dörnyei, 2005). However, our changing view of the language acquisition process renders this approach increasingly untenable. Therefore, to allow for a more complex view of language learning, recent proposals call for the simultaneous investigation of ID variables on the one hand (Ryan, 2019), and longitudinal studies on the other (Dörnyei, 2009a).

In line with this new trend, we designed a longitudinal study in which we set out to investigate changes in secondary school students' motivation, autonomy, self-efficacy, and emotions across a schoolyear with the help of a questionnaire survey. We were interested in detecting changes in the aforementioned variables and their relationships. The second wave of our data collection incidentally coincided with the outbreak of the worldwide COVID-19 pandemic and the resulting nationwide lockdown of schools. Therefore, to some extent, our results can be attributed to the impact of the sudden shift to online education, although it is not possible at this stage to identify the root cause of the changes with complete certainty. In the following section, we would like to offer a brief overview of the individual variables investigated in our study, starting with motivation and autonomy through self-efficacy beliefs to emotions. This will be followed by the

description of our research methods, as well as the presentation and discussion of the results. We will conclude our paper by pointing out the limitations of our research and offering the pedagogical implications of our findings.

Review of Literature

Language Learning Motivation and Learner Autonomy

When examined in isolation as a single predictor of language learning success, motivation tends to be one of the most influential variables in this respect (Dörnyei, 2005). Although theories of motivation are numerous within our field, one that has recently generated considerable empirical research (see Boo et al., 2015; Csizér, 2019) is Dörnyei's L2 Motivational Self System (Dörnyei, 2005, 2009b); therefore, we chose to adopt it in our study as well. The model comprises two self-related concepts, the ideal L2 Self and the ought-to L2 self, and the L2 learning experience. While the L2 learning experience refers to the actual social and physical context of language learning, in our case the foreign language classroom, the self-related concepts describe possible selves (Markus & Nurius, 1986). The ideal L2 self reflects the image of the competent language user the learner wishes to become, while the ought-to L2 self represents the external pressures and expectations learners think they should comply with. Taken together, these three components are expected to predict learners' motivated learning behavior, which reflects the effort learners are willing to invest in learning a foreign language. The model has proven to be successful in predicting overall motivation and achievement in various settings (Boo et al., 2015; Csizér, 2019) despite the fact that the ought-to L2 self appears to be less relevant in certain contexts than the ideal L2 self or the L2 learning experience (Al-Hoorie, 2018).

Compared to motivation, autonomy is a more recent addition to the list of individual differences, and it can be broadly defined as "the ability to take charge of one's learning" (Holec, 1981, p. 3). Within this broad concept, three aspects of autonomy can be distinguished according to Benson (1996, 2007). The first one is *technical autonomy*, which refers to the situation when learning happens outside an educational institution without a teacher, while the second is *psychological autonomy*, which refers to the psychological capacity that enables the learners to take charge of their own learning; the latter may or may not go hand in hand with technical autonomy. The lack of this psychological aspect, for example, may cause difficulties for students who are suddenly confronted with expectations for technical autonomy due to the pandemic. The third aspect, termed *political autonomy*, is quite difficult to implement within the framework of public education, as it refers to having control over both the processes and content of learning. In our study, we aimed to measure autonomous learning behavior, which occurs when learners take responsibility for their own learning, that is, they act in a way that reflects their autonomy; thus, this construct maps learners' willingness to act autonomously.

As early as in 1995, Dickinson argued for establishing links between motivation and autonomy. By citing cognitive theories of motivation, she proposed the strong interrelatedness of autonomy and motivation (Dickinson, 1995). Spratt et al. (2002) claimed that although motivation is often seen as arising from autonomy, the relationship of the two constructs is more complex, and in some cases motivation actually precedes autonomy (p. 262). Strong correlations between autonomous learning behavior and motivation were also seen in the Hungarian context with three different age groups (Csizér & Kormos, 2014; Kormos & Csizér, 2014). As regards the direction of the relationship, Kormos and Csizér (2014) found evidence for their complex model using structural equation modelling. They claimed that through finding opportunities to use the language outside the classroom, overcoming boredom which appears over the course of language learning,

and better time management, motivated learners also became more autonomous. Although the relationship between these two ID variables seems to be supported by empirical findings, it would be useful to explore their potential links with other individual variables, such as self-efficacy or emotions, which have recently begun to be investigated.

Self-efficacy

In ID research, self-efficacy is a relatively new construct, which is conceptually very closely related to variables such as self-confidence (Gardner et al., 1997) and self-perceptions (MacIntyre et al., 1997). Oftentimes these constructs overlap, but they all encompass a certain self-perception of the linguistic or language learning related abilities learners attribute to themselves. Self-efficacy is defined as "people's judgments of their capabilities to organize and execute courses of action required to attain designated types of performances" (Bandura, 1986, p. 391). Accordingly, it is a cognitive construct, which in the context of language learning refers to a particular group of learner beliefs (Horwitz, 1988) also called "language self-efficacy beliefs" (Wong, 2005, p. 248). These include general beliefs about the resources a learner has to learn a foreign/second language and achieve a certain level of language proficiency and more specific beliefs about the ability to perform foreign language related tasks (Bandura, 1986). Pajares (1996) suggests that high levels of self-efficacy beliefs foster persistence in utilizing the strategic tools learners possess to perform learning tasks; what is more, learners with higher levels of efficacy appraisal tend to demonstrate higher levels of perceived responsibility towards their learning (Tilfarlioglu & Ciftci, 2011; Zimmerman, 1994; Zimmerman & Kitsantas, 2005).

In applied linguistics research, self-efficacy beliefs have been directly and indirectly linked to learner's self-perceptions of language anxiety and language learning motivation. In their study, Tremblay and Gardner (1995) found empirical evidence for a model of ID variables depicting a possible

negative relationship between language anxiety and self-efficacy. Interestingly, Tremblay and Gardner (1995) suggest the direction of the link is from language anxiety towards self-efficacy, while Bandura (1988) suggests the reverse, in other words that higher levels of self-efficacy lower language learners' anxiety. He argues that this is because belief in the successful completion of an activity reduces learners' negative feelings. More recently, Piniel and Csizér (2013) have also looked at the relationship between self-efficacy and language learning motivation in the context of English as a Foreign Language learners in Hungary. The authors found that language learners' motivated learning behavior, language learning experience, self-efficacy beliefs and anxiety were circularly related, and that self-efficacy directly influenced motivation.

To date, we know very little about self-efficacy's relationship with other key ID variables, such as learner autonomy that are also strongly associated with language learning success. Macaro (2008) posits that metacognitive awareness is a key part of autonomy, and self-efficacy, or learners' beliefs about their own efficiency regarding the strategic sources they have in order to complete a language learning task, is an important constituent in this metacognition. Similarly, Ruelens (2019) suggests that autonomy and self-efficacy as concepts are so interrelated that it is practical to merge the two when designing an instrument for use in the tertiary education context of language learning. The main tenet of her argument lies in the fact that learners' perceptions of their efficacy, in other words their abilities, and the tools, the strategies they possess to perform certain actions, are very closely linked to the component of autonomy that involves the ability to manage the learning process. Moreover, they are very much related to the readiness to use these strategies, which is yet another aspect of learner autonomy.

Emotions

Other recent additions to the list of variables that are thought to influence the process of language learning are the much less stable (compared to those mentioned above), ephemeral phenomena called emotions. Emotions, which can be defined as "multifaceted responses to events that we see as challenges or opportunities in our inner or outer world, events that are important to our goals" (Keltner et al., 2014, p. 27), are claimed to have an important role in people adapting to their environment (Reeve, 2009). Since they enable the activation of certain built-in response patterns, they are considered instrumental for coping with fundamental life tasks such as protection, reproduction, or the exploration of the environment (Plutchik, 1980). Emotion researchers (Izard, 2010; Keltner et al., 2014) also tend to emphasize the multi-componential nature of emotions: neurobiological, perceptual-cognitive and phenomenological aspects all contribute to them.

It is customary to categorize emotions; proponents of the dimensional view argue that all emotions can be placed along a small number of dimensions, the most often cited ones are arousal/activation and valence/pleasantness (Larsen & Fredrickson, 1999). Although this approach solves the problems associated with the many overlapping labels used to describe emotions, it clearly ignores the fact that specific emotions tend to have their own predictors, they usually lead to different behavioral outcomes, and they are hypothesized to serve different functions (Izard, 2007). As stated above, in addition to the feelings involved, emotions have important cognitive components. Lazarus (1991) proposed a two-stage process of appraisal to account for the cognitive component, where the automatic, almost reflex-like first phase called primary appraisal, is followed by a second phase where cognitions play an important role. During this phase of secondary appraisal, events are evaluated according to the individual's own concerns, and the cognitions that are inherently attached to the feelings experienced are pivotal in determining the quality of the

emotions and the way they are labelled by the individual. Appraisal processes also have a major role in Pekrun's (2006) control-value theory of emotions, which uses a three-dimensional taxonomy of emotions. The first dimension refers to *valence* (positive/pleasant vs. negative/unpleasant quality), the second to activation (activating vs. deactivating tendency) while the third dimension describes whether the focus of achievement is the *activity* itself (intrinsic emotion) or the outcome (extrinsic emotion). This theory has inspired much research in the field of education (Pekrun et al., 2002; Pekrun et al., 2007; Pekrun et al., 2009; Pekrun et al., 2011;), and it is also starting to gain ground in our field. (Shao et al., 2019).

Although research about emotions in applied linguistics has concentrated mainly on negative emotions, especially anxiety (Horwitz, et al., 1991), the advent of positive psychology has prompted a need to pay more attention to the positive experiences and affective states the leaners experience (MacIntyre & Gregersen, 2012; MacIntyre & Mercer, 2014; Oxford, 2015). This argumentation is based on Fredrickson's (2003, 2008) broaden-and-build theory which claims that while negative emotions usually lead to characteristic thoughts and actions which result in focusing and narrowing, positive emotions "all share the ability to broaden people's momentary thought-action repertoires and build their enduring personal resources, ranging from physical and intellectual resources to social and psychological resources" (Fredrickson, 2003, p. 219). This means that positive emotions have a potential to build different resources; consequently, they are considered rather beneficial in educational contexts.

As far as studying emotions in connection with language learning is concerned, Dewaele and MacIntyre (2014, 2016) were the first who highlighted the importance of the feeling of enjoyment and argued for its positive effects on language learning. In Hungary, both qualitative and quantitative approaches have been used to explore a wider range of emotions other than enjoyment and anxiety. In a qualitative exploratory study, English major university student's emotions were investigated in connection with

different language skills, where anxiety and enjoyment turned out to be the most frequently mentioned affects (Piniel & Albert, 2018). Moreover, in a large-scale nation-wide survey, students' feelings of enjoyment, anxiety, boredom, and apathy in connection with their language classes were also explored (Albert et al., 2018a, 2018b). The findings revealed some differences between the two age groups compared. Not only did 11th-graders enjoy language learning less than 7th-graders, but they were also more prone to experience anxiety and apathy.

Many emotions have been linked to motivation in academic contexts in general (e.g., Turner & Schallert, 2001), and in language learning in particular (e.g., MacIntyre & Vincze, 2017), enjoyment and anxiety being among the most frequently researched emotions (Dewaele & MacIntyre, 2014, 2016). Likewise, positive correlations have been established between autonomy and positive emotional experience (e.g., Resnik & Dewaele, 2021), and more recently theoretical links have been proposed between self-efficacy and emotions in academic settings, including classroom language learning (Shao et al., 2019).

Research questions

Although this brief overview of the variables in our study clearly shows their importance, understanding how they relate to each other in the complex process of language learning can reveal much more. While the interconnectedness of these variables appears plausible from a theoretical point of view, empirical data is needed to lend support to establishing their relationships. Besides taking a snapshot of their relationships using a cross-sectional approach, it would also be important to see how their relationships change over time in a longitudinal study.

Based on the above considerations, our research questions were formulated as follows:

1. What characterizes the motivation, autonomy, self-efficacy, and emotions of secondary school students during a school year in which a sudden shift has taken place from offline to online education due to a global pandemic?
2. What characterizes the relationship among secondary school students' students' motivation, autonomy, self-efficacy, and emotions across a schoolyear, where there was a sudden shift from offline to online education due to a global pandemic?

Methods

To answer the research questions above, we designed a longitudinal study where we used quantitative methods to collect and analyze Hungarian secondary school EFL learners' data regarding these variables. This means that we collected our first set of data in the fall of 2019 (time1), and our second point of data collection took place in the spring of 2020 (time2), after the government decided to close all schools and continue education online, including the teaching of foreign languages. In the following, we will elaborate on the participants, the instruments and the details of our data collection and analysis procedures.

Participants

A total of 207 students took part in the study, but due to attrition, the final database from the two data collection points included 187 language learners who had completed the instrument both times. The participants came from three different secondary schools in Budapest, and at the time of data collection were learning English as a foreign language in their respective institutions. In total, there were 123 female and 64 male learners from grades 9 through 11. Their ages ranged from 14 to 18, with a mean of 15.34, and most of them had started learning English when they were 9 years old.

Since 2003, secondary school students in Hungary have been legally obliged to study two foreign languages (Government decree, 2003). For the majority of participants (n = 105) English was the first foreign language chosen, while for 75 learners it was their second. (We had no information on this for 7 learners.) Our participants were learning English in 3 to 5 lessons a week (n = 84 having 5 lessons a week) at school, and most were planning to take the advanced level school leaving examination (B2 level examination according to the CEFR), although their number changed somewhat from time1 (n_{Time1} = 132) to time2 (n_{Time2} = 101). Before the pandemic, n_{Time1} = 151 students also took private lessons and n = 19 were enrolled in extra-curricular English language courses. These numbers changed dramatically after the governmental decision to close educational institutions. At the time of our second data collection only n_{Time2} = 5 out of the n_{Time1} = 151 continued to work with their private tutors and n_{Time2} = 19 who during time1 attended language school courses, started taking private lessons. Two students indicated that they had started private classes who had not done so during time1.

Questionnaire

The data collection tool was written in the participants' native language, Hungarian, and consisted of two parts. The first part tapped into the individual differences outlined above, while the second consisted of biographical questions. In order to collect data on language learners' motivation, autonomy, self-efficacy and emotions, the self-report tool included items representing 12 constructs. Several scales were based on findings of previous studies (e.g., Piniel & Albert, 2018; Piniel & Csizér, 2013) and many items were adapted from previous research conducted in the Hungarian language learning context (Csizér & Kormos, 2009, 2014; Kormos & Csizér, 2008; Öveges & Csizér, 2018; Albert et al., 2018a, 2018b). The constructs measured were the following:

Motivated learning behavior *(5 items):* The scale describes the amount of effort students are ready to invest in learning English. Example: "I can honestly say that I do everything I can to master the English language."

Autonomous learning behavior *(9 items):* The items measure the extent to which participants learn and practice English on their own. Example: "I spend more time practicing elements in English that I find difficult to understand."

Self-efficacy *(6 items):* The construct taps into learners' perceptions of whether they feel they can complete a particular task. Example: "I am sure that I can do the speaking tasks we are given in our English lessons."

Emotions

Enjoyment (5 items): refers to learners' feelings of enjoyment while taking part in the activities and discussion of topics during language lessons. Example: "I enjoy the topics that we discuss in English lessons."

Hope (6 items): measures how hopeful learners feel about achieving success in learning English at school. Example: "I feel hopeful about overcoming challenges in the process of learning English."

Pride (5 items): taps into the extent learners feel proud of their achievements in language learning. Example: "I am proud of my achievements in language learning."

Curiosity (6 items): the items measure how curious and interested learners feel about learning English, and the topics and activities they encounter during the English lessons. Example: "In English lessons, we deal with topics that arouse my curiosity."

Anxiety (5 items): measures learners' feelings of inhibition experienced in connection with English language activities during lessons. Example: "I get frustrated if I can't understand an English-language text."

Boredom (5 items): the items measure to what extent learners feel bored during the activities and topics in the English language lessons. Example: "I get bored by the activities in English lessons."

Apathy (4 items): refers to learners' feeling of hopelessness about their success in learning English at school. Example: "I feel hopeless about ever mastering English at school."

Confusion (5 items): the items measure the extent to which learners feel confused about language learning in class. Example: "Sometimes I feel confused because I don't understand what is happening in the English lessons."

Shame (5 items): items tap into learners' feelings of shame about their achievement and actions during English lessons. Example: "I feel ashamed if I can't answer a question during our English lessons."

The questionnaire was piloted (Albert et al., 2019; Csizér & Öveges, 2019; Csizér & Öveges, 2020; Csizér et al., 2021) in the previous academic year on the same population of Hungarian secondary school English language learners.

Data Collection Procedures

Before collecting the survey data, we contacted the principals of the participating secondary schools. We provided them with the details of our research project and asked for their consent. Next, we talked to the EFL teachers in the respective schools and informed them of the data collection procedures and asked for their consent as well. Finally, with the help of the language teachers, we asked parents to sign a consent form to allow the students to take part in our research. Once these forms were collected, we began to administer the questionnaire in paper and pencil format. However, in the second wave of the data collection, due to the lockdown of schools, data was collected online using the online version of the original

questionnaire. In order to match learners' responses in the first and second wave, we asked the learners to record their names on the instrument. Because the paper and pencil version of the questionnaire was collected by the researchers, the teachers or other participants had no access to the questionnaires; likewise, the data from the online survey was only accessible by the researchers to ensure confidentiality of the responses. The data of both waves were recorded on a spreadsheet using SPSS version 19.0 and with the help of the participants' names, the data from the two data collection times were matched.

Data Analysis Procedures

To answer our research questions, we conducted descriptive analyses and reliability analyses of the scales representing the 12 constructs. T tests were run to see whether any differences appeared in the data collected from the same group at two different points in time, with the second data collection taking place during the lockdown of schools due to the pandemic. Finally, the multiple regression analysis served exploratory purposes. That is, we were interested in identifying the variables which directly and indirectly influence autonomous learning behavior. With this in mind, we set autonomous learning behavior as the dependent variable and motivation, the various emotions, and self-efficacy as the predictors. Before running the analysis, we checked the conditions for multiple linear regression for each wave of the data.

Results and Discussion

The first research question focused on the characteristics of secondary school students' motivation, autonomy, self-efficacy and emotions during a school year where there was a sudden shift from offline to online education due to a global pandemic. In order to answer our question, we conducted descriptive and reliability analyses, the results of which we summarize in Table 1. Based on the reliability coefficients, we can say that the 12 scales

used to measure the various ID constructs had an acceptable level of reliability, with Cronbach's alpha values ranging from .64 to .93. The mean values of autonomous language learning behavior, language learning motivation and self-efficacy beliefs show that our participants reported moderate levels of independence in their learning (M_{Time1} = 3.12, SD = 0.66; M_{Time2} = 3.22, SD = 0.70), higher levels of motivated learning behavior (M_{Time1} = 3.88, SD = 0.70; M_{Time2} = 3.79, SD = 0.80), and fairly high levels of a sense of self-efficacy (M_{Time1} = 3.95, SD = 0.87; M_{Time2} = 4.00, SD = 0.83). This means that although participants do not necessarily take the initiative to manage their own language learning, they are reasonably motivated and, for the most part, feel able to cope with the language learning tasks that arise in the school context.

In terms of the emotions investigated, the general tendency that positive emotions had higher means than negative emotions is considered positive. This suggests that our participants have positive emotional experiences when learning English as a foreign language in school, with hope, in particular displaying the highest average (M_{Time1} = 4.30, SD = 0.56; M_{Time2} = 4.23, SD = 0.62). Among the negative emotions, shame was reported to be the most salient (M_{Time1} = 2.72, SD = 0.97; M_{Time2} = 2.81, SD = 1.05). These results are partially consistent with our pilot studies (Albert et al., 2019, Csizér, et al. 2021), where hope was consistently found to be the most frequently endorsed positive emotion. However, in those studies, confusion, boredom, and anxiety had higher averages than shame. This finding might suggest that while high school language learners tend to be generally hopeful regarding their English learning, the negative emotions they experience could be much more changeable and context dependent. We hypothesize that this might be linked to the greater specificity and narrower action-repertoire associated with negative emotions, in general (Fredrickson, 2003).

We also examined whether the reported levels of autonomy, motivation, self-efficacy, and the emotions measured changed significantly from the first data collection point to the second, where due to the pandemic,

EFL classes were moved to online platforms. The results of the paired samples t tests showed that there were statistically significant differences in learners' autonomous learning behavior (M_{Time1} = 3.12, SD = 0.66; M_{Time2} = 3.22, SD = 0.70; t = -2.624; p = 0.007), motivated learning behavior (M_{Time1} = 3.88, SD = 0.70; M_{Time2} = 3.79, SD = 0.80; t = 2.319; p = 0.021) and hope (M_{Time1} = 4.30, SD = 0.56; M_{Time2} = 4.23, SD = 0.62; t = 2.091; p = 0.038). While autonomous learning behavior increased, motivation and hope that they would eventually become successful language learners decreased. It can be hypothesized that students reacted to the new situation by displaying higher levels of autonomy, but the changed circumstances also decreased their motivation and made them less hopeful about their future achievements. While these findings seem to be consistent with changes that we might expect in a situation where students are suddenly given more technical autonomy (Benson, 1996, 2007) and are forced to make use of it, without baseline data for the same students showing how these variables might fluctuate over the course of a regular schoolyear, our conclusions are purely speculative.

To answer our second research question about the relationships among learners' motivation, autonomy, self-efficacy and emotions in the two waves of data collected, we conducted regression analyses. First of all, to see whether multicollinearity would pose an obstacle for our analysis, we looked at the correlations between the independent variables. From the data collected at time1, the only high correlation (above .7) was found between curiosity and enjoyment (r_{Time1} =.73 p < .01). We calculated the Variance Inflation Factor (VIF) to see whether this high correlation would indeed cause a problem when running the linear regression. Since the VIF values for enjoyment and curiosity were 2.47 and 1.38, and fell below the recommended value of 5, multicollinearity was not considered as an issue in sample$_{Time1}$. For the second wave, there were several strong correlations detected between the predictor variables: between anxiety and confusion (r_{Time2} = 0.7, p < .01), between anxiety and shame (r_{Time2} = 0.71, p < .01), between curiosity and boredom (r_{Time2} = -0.75, p < .01) and curiosity and enjoyment (r_{Time2} = 0.79, p < .01).

Table 1

Descriptive Statistics and Reliability of the Scales

Scales	Number of items	Cronbach's alpha	Mean	St. Dev.
Motivated learning behavior	5	0.79	3.88	0.70
Motivated learning behavior	5	0.84	3.79	0.80
Autonomous learning behavior	9	0.79	3.12	0.66
Autonomous learning behavior	9	0.82	3.22	0.70
Self-efficacy 1	6	0.93	3.95	0.87
Self-efficacy 2	6	0.93	4.00	0.83
Enjoyment 1	5	0.70	3.87	0.65
Enjoyment 2	5	0.77	3.80	0.73
Hope 1	6	0.74	4.30	0.56
Hope 2	6	0.75	4.23	0.62
Pride 1	5	0.86	3.70	0.96
Pride 2	5	0.86	3.62	0.95
Curiosity 1	6	0.78	3.34	0.72
Curiosity 2	6	0.83	3.35	0.79
Anxiety 1	5	0.64	2.64	0.78
Anxiety 2	5	0.70	2.67	0.82
Boredom 1	5	0.76	2.15	0.75
Boredom 2	5	0.80	2.14	0.76
Apathy 1	4	0.73	2.10	0.87
Apathy 2	4	0.75	2.11	0.88
Confusion 1	5	0.66	2.57	0.72
Confusion 2	5	0.72	2.53	0.77
Shame 1	5	0.80	2.72	0.97
Shame 2	5	0.84	2.81	1.05

We also looked at the VIF values. They were between 1.26 and 3.67, which suggest moderate multicollinearity but still fall below the recommended value of 5. Thus, for our sample$_{Time2}$, multicollinearity was not considered to be a problem.

To further analyze compliance with the assumptions of the multiple regression analysis based on the residuals, the standard residuals were calculated for the data collected at time1 and time2. For time1, standard residuals ranged between -2.89 and 2.73, which were between the recommended -3.0 and 3.0 range. For the second wave, the standard residual fell between -2.44 and 2.24, which was also acceptable. Finally, we ran the multiple regression analysis for both cohorts with autonomous learning behavior as the dependent variable. For the sake of clarity, we will only report on the connections that had an associated Beta value greater than 0.2 (see Tables 2 and 3).

Table 2

Results of Multiple Regression Analyses with Autonomous Learning Behavior as the Dependent Variable during Time 1

Autonomous learning behaviours	*Beta*	*t*	*p*
Motivated learning behavior	.58	11.50	< .001
Curiosity	.36	6.48	< .001
R^2		.58	
F change		5.73	
Sig. F change		.018	
Motivated learning behavior	Beta	t	p
Enjoyment	.26	3.43	.001
Hope	.23	2.56	.011
R^2		.28	
F change		4.66	
Sig. F change		.03	
Curiosity	Beta	t	p
Enjoyment	.56	9.54	< .001
Boredom	-.36	-6.54	< .001
R^2		.65	
F change		9.97	
Sig. F change		.002	

Table 3

Results of Multiple Regression Analyses with Autonomous Learning Behavior as the Dependent Variable during Time 2

Autonomous learning behaviour	*Beta*	*t*	*p*
Motivated learning behaviour	.60	12.35	< .001
Curiosity	.27	4.74	< .001
Shame	.23	4.91	< .001
Self-efficacy	.22	4.37	< .001
R^2		.67	
F change		9.59	
Sig. F change		.002	
Motivated learning behaviour	*Beta*	*t*	*p*
Hope	.48	5.80	< .001
Curiosity	.28	4.09	< .001
Apathy	.22	3.28	.001
Pride	.34	3.98	< .001
Self-efficacy	-.25	-3.08	.002
R^2		.46	
F change		9.48	
Sig. F change		.002	
Curiosity	*Beta*	*t*	*p*
Enjoyment	.54	10.20	< .001
Boredom	-.39	-7.52	< .001
R^2		.72	
F change		7.97	
Sig. F change		.005	
Shame	*Beta*	*t*	*p*
Confusion	.40	5.70	< .001
Pride	-.21	-2.77	.006
Curiosity	.28	3.59	< .001
R^2		.31	
F change		4.68	
Sig. F change		.032	
Self-efficacy	*Beta*	*t*	*p*
Pride	.35	5.32	< .001
Confusion	-.34	-6.70	< .001
Hope	.37	5.53	< .001
R^2		.64	
F change		5.92	
Sig. F change		.016	

There are some similarities, but also some differences, between the two path models derived from the results of the regression analysis of the data collected at time 1 and 2. (see Figures 1 and 2). Results based on the data collected from our participants at the two points in time (for details see Tables 2 and 3) clearly show that autonomous learning behavior is influenced by motivated learning behavior and the feeling of curiosity. The link between motivation and autonomy is not surprising, as their positive relationship has been hypothesized both on theoretical grounds (Dickinson, 1995) and demonstrated in several empirical studies (Csizér & Kormos, 2014; Kormos & Csizér, 2014; Spratt et al., 2002). Since the emotion of curiosity appears to be the driving force behind the exploration of the environment (Voss & Keller, 2013), which is probably an indispensable precondition for any autonomous action, the positive contribution of curiosity to autonomy is not surprising. Psychometric measures of curiosity and autonomy also showed a positive relationship in a questionnaire study (Kashdan et al., 2009). Furthermore, hope appears in both samples to be a key predictor of motivated learning behavior. According to Pekrun (2006), hope is an activating positive emotion, so its potential to motivate learners to take action should not come as a surprise. Finally, curiosity was strongly influenced by enjoyment and negatively predicted by boredom in both cohorts. While the relationship between enjoyment and curiosity may not be so obvious as both are activating positive emotions (Pekrun, 2006; Pekrun et al., 2007; Pekrun et al., 2009), the negative relationship between boredom and curiosity seems quite obvious, as they probably represent two ends of a continuum.

Figure 1

Path Model Based on Multiple Regression Analyses' Results with Autonomy as the Dependent Variable during Time 1

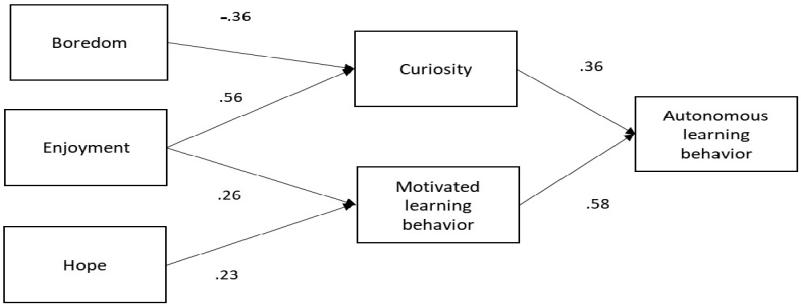
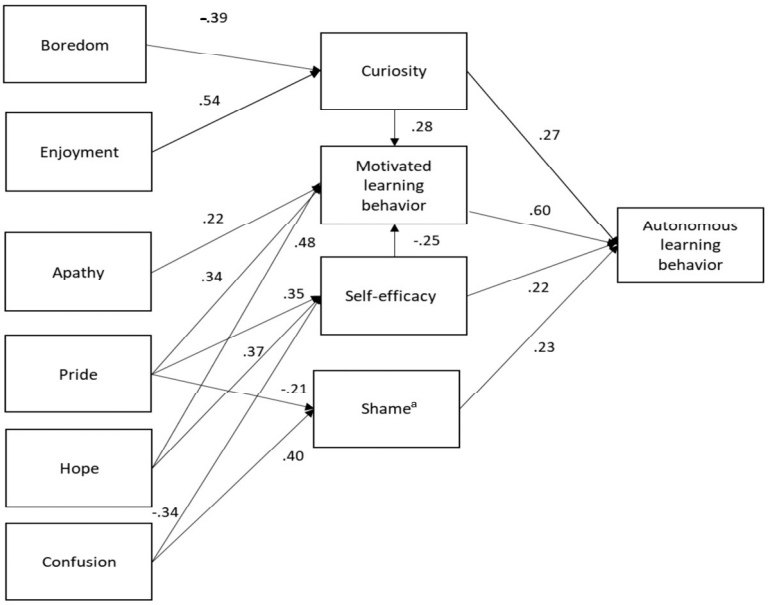


Figure 2

Path Model Based on Multiple Regression Analyses' Results with Autonomy as the Dependent Variable during Time 2



Note. Numbers refer to the Beta values in the regression analysis

[a] The Beta value of curiosity's influence on shame is .28

As for the differences in the results of the regression analyses conducted at the two different time points, the data collected during the time of the pandemic and the closing down of schools towards the end of the schoolyear seem to indicate the importance of a larger variety of emotions than earlier (see Figure 2). Besides curiosity, shame appeared to directly predict learner autonomy. Although the positive contribution of shame to autonomous learning behavior might seem counterintuitive at first, there is empirical evidence to support that in some cases shame can have a motivating effect (Turner & Schallert, 2001).

In their study examining the effects of shame on students taking a highly demanding pharmacology course, Turner and Schallert (2001) argued that

> if students believe that they have the capabilities and are committed to a clear future goal for which a course grade or information is considered to play an instrumental role within a progressive plan for obtaining their future goal, then a shame reaction may be perceived as a signal that learning-related strategies and actions need to be reevaluated and modified (p. 327)

If shame is perceived as a warning signal that the current course of action leads to unsatisfactory results, it might urge the learner to change and act more autonomously in a situation that requires autonomy, such as when schools were closed in the lockdown.

Another difference between the results of the two waves of data was a significant direct contribution of self-efficacy to autonomous learning behavior as well as to motivation in time2. These results are in line with the theory explained above and what some other studies discovered about the relationship of learner autonomy and self-efficacy. Zimmerman and Kitsantas (2005) in their regression analysis found self-efficacy to be predictive of perceived responsibility for learning. Similarly, Tilfarlioglu and Ciftci (2011) in the Turkish university context, demonstrated a strong

correlation between self-efficacy and learner autonomy, and the constructs were shown to be significant predictors of academic success. In the emerging COVID-19 situation during time2, learners' self-efficacy beliefs were brought to the foreground as persistence in applying new strategies and becoming more autonomous became more important in the online language learning context.

However, there are certain relationships in the rather complex path model at time2 that are difficult to explain. One such relationship is the negative contribution of self-efficacy to motivated learning behavior. It can be assumed that learners characterized by high levels of self-efficacy could be very confident in their abilities, so in certain situations, for example when they do not see the value or usefulness of their actions, their willingness to take action or put effort in language learning may diminish. Nevertheless, finding a plausible explanation for the positive contribution of apathy to motivated learning behavior and curiosity to shame seem even less likely. It is suggested that these findings may be anomalies resulting from the extraordinary circumstances of the pandemic; thus, they probably require further research.

An interesting result of the exploratory regression analyses is that anxiety did not appear as a predictor for autonomy and motivated learning behavior. Instead, a very similar construct, confusion, seemed to exert its indirect influence through the feeling of shame and self-efficacy beliefs. This could be explained by the notion that confusion is a very frequent cognitive concomitant of anxiety (MacIntyre & Gardner, 1994). Furthermore, as Pekrun et al. (2017) suggest, while confusion indeed represents cognitive incongruence, anxiety is viewed as an emotion "triggered by the lack of resolution of incongruity" (p. 3, Supplemental Material).

Conclusion

In our longitudinal study, we attempted to shed light on changes that might be detectable in learners' motivation, autonomy, self-efficacy, and emotions over the course of a schoolyear. The statistically significant increase in autonomy and the decrease in motivation and hope might be attributed to the sudden switch to online education, which forced learners to be more self-reliant but, at the same time, decreased their motivation and their hope of success. However, these are only tentative conclusions and our results must be treated with caution due to the lack of baseline data that could have shown any changes during a regular schoolyear.

As for the factors that contributed to our learners' autonomous learning behavior, we also observed changes in the schoolyear. While at the beginning of the schoolyear it was the students' motivation and their curiosity that contributed to autonomous learning behavior, we saw a wider range of determinants towards the end of the schoolyear. Once again, this change might either be ascribed to the unexpected changes in the circumstances or seen as a natural consequence of reaching the end of the schoolyear. It would be desirable to collect data from the same students on a regular basis to address this issue.

The pedagogical implications of our findings again depend on what we attribute the changes to. While the decrease in motivation and feelings of hope that accompanies more autonomous behavior can be seen as understandable and perhaps even adaptive under the changed circumstances, for example when students suddenly need to switch to distance learning, these processes may be of concern if they occur during a typical schoolyear. Although more autonomous behavior is usually thought of as laudable, it might also reflect dissatisfaction with what the school and teachers offer, which is an interpretation that is more in line with the decline in motivation and hope. Our findings also call attention to the fact that simply attributing positive effects to positive emotions and negative effects to

negative ones is a crude oversimplification of the complexities of the learning situation, and that contextual effects also need to be considered when interpreting such findings.

Finally, the limitations of our research should also be mentioned. The fact that there was an enormous change in the learning situation itself, that is, the first phase of data collection took place in a classroom environment while the second phase was taught online, might have influenced our results to a great extent. While in some ways we feel fortunate that we had the opportunity to collect data under such extraordinary circumstances, the lack of baseline data from a regular schoolyear makes our findings rather difficult to interpret. Therefore, we would like to collect additional data in future data collection waves in order to make our basis of comparison more sound. Despite the difficulties we encountered, we are convinced that the simultaneous and possibly longitudinal investigation of ID variables is the path researchers should follow in the future, as the ID variables investigated in our study are clearly related and their relationship changes over time.

Note 1: We have no conflict of interest to disclose.

Note 2: This study was supported by the NKFIH–129149 research grant.

References

Albert, Á., Piniel, K., & Lajtai Á. (2019). Középiskolás diákok angolórákkal kapcsolatos érzelmei egy kérdőíves felmérés tükrében [Investigating high school students' emotions in conncetion with their English classes with the help of a questionnaire study]. *Modern Nyelvoktatás*, *25*(2), 27–41.

Albert, Á., Tankó, Gy., & Piniel, K. (2018a). A tanulók válaszai a 11. évfolyamon [Answers of grade 11 students]. In E. Öveges & K. Csizér (Eds.), *Vizsgálat a köznevelésben folyó idegennyelv-oktatás hatékonyságáról* (pp. 93–163). Oktatási Hivatal.

Albert, Á., Tankó, Gy., & Piniel, K. (2018b). A tanulók válaszai a 7. évfolyamon [Answers of grade 7 students]. In E. Öveges & K. Csizér (Eds.), *Vizsgálat a köznevelésben folyó idegennyelv-oktatás hatékonyságáról* (pp. 93–163). Oktatási Hivatal.

Al-Hoorie, A. (2018). The L2 Motivational Self System: A meta-analysis. *Studies in Second Language Learning and Teaching*, *8*, 721–754. https://doi.org/10.14746/ssllt.2018.8.4.2

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

Bandura, A. (1988). Self-efficacy conception of anxiety. *Anxiety Research: An International Journal*, *1*(2), 77–98. https://doi.org/10.1080/10615808808248222

Benson, P. (1996). Concept of autonomy in language learning. In R. Pemberton, E. S. L. Li, W. W. F. Or, & H. D. Pierson (Eds.), *Taking control: Autonomy in language learning* (pp. 27–34). Hong Kong University Press.

Benson, P. (2007). Autonomy in language teaching and learning. *Language Teaching*, *40*, 21–40. https://doi.org/10.1017/S0261444806003958

Boo, Z., Dörnyei, Z., & Ryan, S. (2015). L2 motivation research 2005–2014: Understanding a publication surge and a changing landscape. *System*, *55*, 147–157. https://doi.org/10.1016/j.system.2015.10.006

Csizér, K. (2019). The L2 Motivational Self System. In M. Lamb, K. Csizér, A. Henry, & S. Ryan (Eds.), *Palgrave Macmillan handbook of motivation for language learning* (pp. 71–93). Palgrave.

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. https://doi.org/10.1007/978-3-030-75726-7_1

Csizér, K., & Kormos, J. (2009). Modelling the role of inter-cultural contact in the motivation of learning English as a foreign language. *Applied Linguistics*, *30*, 166–185. https://doi.org/10.1093/applin/amn025

Csizér, K., & Kormos, J. (2014). The ideal L2 self, self-regulatory strategies and autonomous learning: A comparison of different groups of English learners. In K. Csizér & M. Magid (Eds.), *The impact of self-concept on language learning* (pp. 73–86). Multilingual Matters.

Csizér, K., & Öveges, E. (2019). Idegennyelv-tanulási motivációs tényezők és a nyelvi vizsgák Magyarországon: Összefüggések vizsgálata egy kérdőíves kutatás segítségével [Motivation to learn a foreign language and language exams in Hungary: Studying their relationships with the help of a questionnaire study]. *Modern Nyelvoktatás*, *25*, 86–101.

Csizér, K., & Öveges, E. (2020). Nyelvtanulási autonómia és nyelvi tervezés: egy vegyes módszerű kutatás eredményei [Language learning autonomy and language planning: Results of a mixed methds study]. *Modern Nyelvoktatás*, *36*(1–2), 44–58.

Dewaele, J.-M., & MacIntyre, P. D. (2014). The two faces of Janus? Anxiety and enjoyment in the foreign language classroom. *Studies in Second Language Learning and Teaching 4*(2), 237–274. https://doi.org/10.14746/ssllt.2014.4.2.5

Dewaele, J.-M., & MacIntyre, P. D. (2016). Foreign language enjoyment and foreign language classroom anxiety: The right and left feet of FL learning? In P. D. MacIntyre, T. Gregersen, & S. Mercer (Eds.), *Positive psychology in SLA* (pp. 215–236). Multilingual Matters.

Dickinson, L. (1995). Autonomy and motivation a literature review. *System*, *23*(2), 165–174. https://doi.org/10.1016/0346-251X(95)00005-5

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

Dörnyei, Z. (2009a). Individual differences: Interplay of learner characteristics and learning environment. *Language Learning*, *59*, 230–248. https://doi.org/10.1111/j.1467-9922.2009.00542.x

Dörnyei, Z. (2009b). The L2 Motivational Self System. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 9–42). Multilingual Matters.

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

Fredrickson, B. L. (2003). The value of positive emotions. *American Scientist*, *91*(4), 330–335. https://www.jstor.org/stable/27858244

Fredrickson, B. L. (2008). Promoting positive affect. In M. Eid & R. J. Larsen (Eds.), *The science of subjective well-being* (pp. 449–468). Guilford Press.

Gardner, R. C., Tremblay, P. F., & Masgoret, A. M. (1997). Towards a full model of second language learning: An empirical investigation. *The Modern Language Journal*, *81*(3), 344–362. https://doi.org/10.2307/329310

Government decree No. 243/2003 (XII.17). On the issuance, implementation and application of the National Core Curriculum. Retrieved on 8 February 2021 from http://www.nefmi.gov.hu/letolt/kozokt/nat_070926.pdf

Holec, H. (1981). *Autonomy in foreign language learning*. Pergamon Press.

Horwitz, E. K. (1988). The beliefs about language learning of beginning university foreign language students. *The Modern Language Journal*, *72*(3), 283–294. https://doi.org/10.2307/327506

Horwitz, E. K., Horwitz, M. B., & Cope, J. A. (1991). Foreign language classroom anxiety. In E. K. Horwitz & D. J. Young (Eds.), *Language anxiety: From theory and research to classroom implications* (pp. 27–36). Prentice Hall.

Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science*, *2*(3), 260–280. https://doi.org/10.1111/j.1745-6916.2007.00044.x

Izard, C. E. (2010). The many meanings/aspects of emotion: Definitions, functions, activation, and regulation. *Emotion Review*, *2*(4), 363–370. https://doi.org/10.1177/1754073910374661

Kashdan, T. B., Gallagher, M. W., Silvia, P. J., Winterstein, B. P., Breen, W. E., Terhar, D., & Steger, M. F. (2009). The curiosity and exploration inventory-II: Development, factor structure, and psychometrics. *Journal of Research in Personality*, *43*(6), 987–998. https://doi.org/10.1016/j.jrp.2009.04.011

Keltner, D., Oatley, K., & Jenkins, J. M. (2014). Understanding emotions (3rd ed). John Wiley and Sons.

Kormos, J., & Csizér, K. (2008). Age-related differences in the motivation of learning English as a foreign language: Attitudes, selves and motivated learning behaviour. *Language Learning*, *58*(2), 327–355. https://doi.org/10.1111/j.1467-9922.2008.00443.x

Kormos, J., & Csizér, K. (2014). The interaction of motivation, self-regulation and autonomous learner behavior in different learner groups. *TESOL Quarterly*, *48*, 275–299. https://doi.org/10.1002/tesq.129

Larsen, R., & Fredrickson, B. (1999). Measurement issues in emotion research. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: Foundations of hedonic psychology* (pp. 40–60). Russell Sage Foundation.

Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press.

Macaro, E. (2008). The shifting dimensions of language learner autonomy. In T. Lamb, & H. Reinders (Eds.), *Learner and teacher autonomy: Concepts, realities and responses* (pp. 47–62). John Benjamins.

MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning*, *44*, 283–305. https://doi.org/10.1111/j.1467-1770.1994.tb01103.x

MacIntyre, P. D., & Gregersen, T. (2012). Emotions that facilitate language learning: The positive-broadening power of the imagination. *Studies in Second Language Learning and Teaching*, *2*(2), 193–213. https://doi.org/10.14746/ssllt.2012.2.2.4

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., Noels, K. A., & Clément, R. (1997). Biases in self-ratings of second language proficiency: The role of language anxiety. *Language Learning*, *47*(2), 265–287. https://doi.org/10.1111/0023-8333.81997008

MacIntyre, P. D., & Vincze, L. (2017). Positive and negative emotions underlie motivation for L2 learning. *Studies in Second Language Learning and Teaching*, *7*(1), 61–88. http://dx.doi.org/10.14746/ssllt.2017.7.1.4

Markus, H., & Nurius, P. (1986). Possible selves. *American Psychologist*, *41*(9), 954–969. https://doi.org/10.1037/0003-066X.41.9.954

Öveges, E., & Csizér, K. (2018). Vizsgálat a köznevelésben folyó idegennyelv-oktatás kereteiről és hatékonyságáról [An investigation of the foreign language teaching in the compulsory education in Hungary]. Oktatási Hivatal. https://www.oktatas.hu/pub_bin/dload/sajtoszoba/nyelvoktatas_kutatasi_jelentes_2018.pdf

Oxford, R. L. (2015). Emotion as the amplifier and the primary motive: Some theories of emotion with relevance to language learning. *Studies in Second Language Learning and Teaching*, *5*(3), 371–393. https://doi.org/10.14746/ssllt.2015.5.3.2

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, *66*(4), 543–578. https://doi:10.3102/00346543066004543

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, *18*(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9

Pekrun, R., Elliot, A. J., & Maier, M. A. (2009). Achievement goals and achievement emotions: Testing a model of their joint relations with academic performance. *Journal of Educational Psychology*, *101*(1), 115–135. https://doi.org/10.1037/a0013383

Pekrun, R., Frenzel, A. C., Goetz, T., & Perry, R. P. (2007). The control-value theory of achievement emotions: An integrative approach to emotions in education. In P. A. Schutz & R. Pekrun (Eds.), *Emotion in education* (pp. 13–36). Academic Press.

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology*, *36*(1), 36–48. https://doi.org/10.1016/j.cedpsych.2010.10.002

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Positive emotions in education. In E. Frydenberg (Ed.), *Beyond coping: Meeting goals, visions, and challenges* (pp. 149–173). Oxford University Press.

Pekrun, R., Vogl, E., Muis, K. R., & Sinatra, G. M. (2017). Measuring emotions during epistemic activities: The epistemically-related emotion scales. *Cognition and Emotion*, *31*(6), 1268–1276. https://doi.org/10.1080/02699931.2016.1204989

Piniel, K., & Albert, Á. (2018). Advanced learners' foreign language related emotions across the four skills. *Studies in Second Language Learning and Teaching*, *8*(1), 127–147. https://doi.org/10.14746/ssllt.2018.8.1.6

Piniel, K., & Csizér, K. (2013). L2 motivation, anxiety and self-efficacy: The interrelationship of individual variables in the secondary school context. *Studies in Second Language Learning and Teaching*, *3*(4), 523–550. https://doi.org/10.14746/ssllt.2013.3.4.5

Plutchik, R. (1980). *Emotion: A psychoevolutionary synthesis*. HarperCollins College Division.

Reeve, J. (2009). *Understanding motivation and emotion*. John Wiley & Sons.

Resnik, P., & Dewaele, J. (2021). Learner emotions, autonomy and trait emotional intelligence in 'in-person' versus emergency remote English foreign language teaching in Europe. *Applied Linguistics Review*. https://doi.org/10.1515/applirev-2020-0096

Ruelens, E. (2019). Measuring language learner autonomy in higher education: The Self-Efficacy Questionnaire of Language Learning Strategies. *Language Learning in Higher Education*, *9*(2), 371–393. https://doi:10.1515/cercles-2019-0020

Ryan, S. (2019). Motivation as an individual difference. In M. Lamb, K. Csizér, A. Henry, & S. Ryan (Eds.), *Handbook of second language motivation* (pp. 163–182). Palgrave Macmillan.

Shao, K., Pekrun, R., & Nicholson, L. J. (2019). Emotions in classroom language learning: What can we learn from achievement emotion research? *System*, *86*, 102–121. https://doi.org/10.1016/j.system.2019.102121

Skehan, P. (1989). *Individual differences in second language learning*. Edward Arnold.

Spratt, M., Humphreys, G., & Chan, V. (2002). Autonomy and motivation: Which comes first? *Language Teaching Research*, *6*(3), 245–266. https://doi.org/10.1191/1362168802lr106oa

Tilfarlioglu, F. Y., & Ciftci, F. S. (2011). Supporting self-efficacy and learner autonomy in Relation to academic success in EFL classrooms (A case study). *Theory & Practice in Language Studies*, *1*(10), 1284–1294. http://doi.org/10.4304/tpls.1.10.1284-1294

Tremblay, P. F., & Gardner, R. C. (1995). Expanding the motivation construct in language learning. *The Modern Language Journal*, *79*(4), 505–518. https://doi.org/10.1111/j.1540-4781.1995.tb05451.x

Turner, J. E., & Schallert, D. L. (2001). Expectancy–value relationships of shame reactions and shame resiliency. *Journal of Educational Psychology*, *93*(2), 320. https://doi.org/10.1037/0022-0663.93.2.320

VanPatten, B., & Benati, A. G. (2015). *Key terms in second language acquisition*. Bloomsbury Publishing.

Voss, H. G., & Keller, H. (2013). *Curiosity and exploration: Theories and results*. Elsevier.

Wong, M. S. L. (2005). Language learning strategies and language self-efficacy: Investigating the relationship in Malaysia. *RELC Journal*, *36*(3), 245–269. https://doi.org/10.1177/0033688205060050

Zimmerman, B. J. (1994). Dimensions of academic self-regulation: A conceptual framework for education. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulation of learning and performance: Issues and educational implications* (pp. 3–21). Lawrence Erlbaum Associates, Inc.

Zimmerman, B. J., & Kitsantas, A. (2005). Homework practices and academic achievement: The mediating role of self-efficacy and perceived responsibility beliefs. *Contemporary Educational Psychology*, *30*, 397–417. https://doi.org/10.1016/j.cedpsych.2005.05.003

10.

Locked Down: Autonomous Language Learning by University Students during a Global Pandemic

Jamil Toptsi
tzamil.toptsi@btk.elte.hu

Abstract

This article aims to explore the perceived impacts of lockdown conditions on the autonomous language learning of six Hungarian students. Circumstances occurring due to the 2020-2021 COVID pandemic, such as the introduction of lockdown measures throughout the world, have led to changes affecting various aspects of peoples' lives, including the language learning endeavours of university students. These new circumstances resulting from the lockdown present potential challenges to language learners. This study aims to explore the perceived impacts that the nation-wide lockdown in Hungary had on the autonomous language learning endeavours of six university students through a longitudinal qualitative research design consisting of three rounds of semi-structured interviews during a two-month lockdown period. The participants recollected the ways in which the lockdown affected their mental states, learning environment, and language learning experiences. Other findings were uncovered in regard to the perceived role that language learning played in the lives of the participants during the pandemic. The study provides a deeper understanding into the perceived impacts of the COVID-19 pandemic on language learning and gives insight into the interactions between the various internal and external factors that make up the autonomous language learning experiences of the participants as well as their development over time.

Keywords: autonomous language learning, COVID-19, motivation, affect

Locked Down: Autonomous Language Learning by University Students during a Global Pandemic

The present study aims to gain an understanding into the perceived impacts of the 2020-2021 COVID-19 pandemic on the autonomous language learning experiences of a group of Hungarian students who are engaged in language learning in addition to their university studies. Through three rounds of interviews over the course of a two-month period, participants reported on the ways they believed the lockdown had affected their language learning experiences and on the role that language learning played during the lockdown. While the effects of lockdown conditions on language learning have been explored to some extent in recently published articles (Bailey & Lee, 2020; Maican & Cocoradă, 2021; Wargadinata et al., 2020), attention has yet to be paid to the autonomous language learning context which is the focus of the present study.

The COVID-19 global pandemic has produced deep, negative impacts on public health, the economy, and education in countries around the world (Chakraborty & Maity, 2020). In addition to these obvious impacts, more subtle impacts can be seen in the mental health of individuals forced into lockdown to slow the spread of the virus (Banks & Xu, 2020; Rossi et al., 2020; Wang et al., 2020). Certain activities that individuals might engage in, such as language learning (Egbert, 2020), have had to undergo changes to meet the new circumstances presented by the pandemic. Understanding these changes and their effects on learners can provide applied linguistics researchers with a better understanding of how language learning activities are realized during such lockdown periods, which can potentially provide insight into other instances of autonomous language learning.

The first section of this paper consists of a review of the literature concerning autonomous language learning and the various factors involved in the autonomous language learning process, as well as a look at recent studies examining the effects of the 2020-2021 COVID-19 pandemic on mental health and language pedagogy. This is followed by a description of

the research methods used, as well as a presentation and discussion of the findings. In addition to providing a summary of the findings, the conclusion discusses some of the potential implications that the study presents in regard to research methodology and pedagogy, as well as directions for future research into the autonomous language learning process.

Literature Review

The following section reviews the relevant literature on motivation, self-regulation and affect in autonomous language learning. To understand the research context of the current study, the latest research investigating the mental and psychological impacts of COVID lockdowns in a number of international contexts is examined in the second half of the literature review together with the effects that these lockdowns have had on students and language learners.

Motivation, Self-Regulation, and Affective Factors in Autonomous Language Learning

Learner autonomy has gained increased attention in applied linguistics research in recent decades and has been studied in relation to other factors involved in the learning process. As pointed out by Benson (2006), there are a variety of definitions for the concept of learner autonomy, the most widely cited being the ability to "take charge of one's own learning" as defined by Holec (1981, p. 3). Autonomous language learning can be conceptualized both within the context of the classroom (Mizuki, 2003; Thomson, 1998; Trebbi, 2003) as well as out of the classroom contexts, which include modes of learning including computer-assisted language learning (Schweinhorst, 2007), distance learning (White, 2003), and self-instruction (Fernández-Toro & Jones, 2001). Within these contexts, other individual differences factors such as motivation, self-regulation, and affect are often studied in relation to learner autonomy.

Motivation has been identified as crucial in autonomous language learning (Gao & Lamb, 2011). Motivation has been conceptualized as two dimensional, involving 1.) the choice to engage in an activity and 2.) the persistence and effort towards that activity (Dörnyei et al., 2006, p. 9). One theoretical framework for language learning motivation which has gained prominence in recent years is Dörnyei's L2 Motivational Self System. This framework is made up of three main components, two representing future selves (the ideal and ought-to L2 selves, drawn from Marcus & Nurius, 1986) and the L2 learning experience, reflecting the enjoyment of the language learning process. This framework has shown to apply in a number of contexts in which English was being learned as a foreign language (Taguchi et al., 2009) as well as in some studies with non-English L3 learning (Wang & Liu, 2017).

Self-regulation, or "planning, guiding, and monitoring one's own attention and behavior" (Berk & Winsler, 1995, p. 171), has been included as an important consideration in autonomous language learning research. A study by Kormos and Csizér (2014) investigated how learner autonomy–operationalized in their study in terms of the autonomous use of digital and traditional learning resources based on Benson (2001)–is influenced both by motivational and self-regulatory components. Kormos and Csizér's study drew on previous research into self-regulation by Tseng et al. (2006), which identified five self-regulatory components in vocabulary learning (see Table 1). In addition to Kormos and Csizér (2014), these components have informed a number of other studies. Ding and Shen (2019) examined the role that these self-regulatory components played in autonomous language learning in Massive Open Online Courses (MOOCs), while Gao and Shen (2021) investigated which self-regulatory strategies were used the most during autonomous language learning with mobile devices. These studies revealed the importance of self-regulation in distinct autonomous learning contexts.

Table 1

Self-Regulatory Components from Tseng et al. (2006)

Component	Definition
Commitment Control	The learner's capacity to preserve or increase their goal commitment
Metacognitive Control	The learner's capacity to monitor and control their concentration and avoid unnecessary procrastination
Satiation Control	The learner's capacity to eliminate boredom and increase their attraction or interest in a task
Emotion Control	The learner's capacity to manage disruptive emotional states and moods, as well as generate emotions that are conducive to their goals
Environment Control	The learner's capacity to eliminate negative environmental influences and exploit positive environmental influences

In addition to self-regulation, affect has also been cited as an important aspect of autonomous language learning (Murray & Lamb, 2018). While much of the previous applied linguistics research on affect has focused on the classroom context with special attention towards anxiety (Arnold, 1999; MacIntyre & Gregersen, 2012), more recent studies have investigated positive affective factors as well, and have expanded the focus of research outside of the classroom to autonomous language learning. Hurd's (2007) study on distance learners revealed the positive and negative emotions experienced by participants engaged in distance-learning tasks, while Lee and Lee (2021) carried out a questionnaire study which showed that motivational variables and autonomous digital learning behaviours are predictive of foreign language enjoyment. As pointed out in Murray and Lamb (2018), the cognitive and affective systems of autonomous language learners "are formed by the learning environment" (p. 260), with major changes to that environment potentially leading to changes in these systems

(Ushioda, 2015). Such perspectives, when taken with the studies mentioned above, reveal the ways in which affect not only has an impact on the autonomous learning process but in turn is affected by factors present within this process.

Due to the continuing development of digital language learning resources, autonomous language learning continues to gain importance for today's learners (Lai, 2017). Recent studies into distance learning during the COVID-19 pandemic have highlighted this importance even further. A number of Spanish studies have examined this particular context, with García-Alberti et al. (2021) and Gonzalez, et al. (2020) examining how autonomous distance learning took place with online resources during the pandemic and Palau et al. (2021) examining differences in autonomous learning among different student groups. These studies into autonomous learning during lockdown conditions have shown how such changes can impact individuals in general and students specifically.

Impacts of the COVID-19 Global Pandemic on Psychological and Mental Health

Several studies have looked at the impact that lockdown conditions have had on the mental health of those experiencing the pandemic and its associated conditions. One such study (Rossi et al., 2020) was based on a large-scale questionnaire distributed to Italians (n = 18,147) three weeks into the country's COVID lockdown in 2020. The questionnaire contained scales measuring post-traumatic stress symptoms (PTSS), depression, anxiety, insomnia, stress, and adjustment disorder (difficulty adjusting to a stressful event which occurs). Based on the scores, a substantial number of the participants reported the previously mentioned issues, with 37% reporting PTSS. Being a woman was associated with all of the previously mentioned scales, while being young was associated with all of them except for insomnia and adjustment disorder. Another study from the UK (Banks & Xu, 2020) drawing data from the UK Household Longitudinal Study (UKHLS)

showed an increase in mental health problems among participants based on their responses to the mental health portion of the 2020 survey. Sharp increases were seen in the responses to a scale measuring 12 different aspects of general mental health. Similarly to the Italian study, young women were shown to be impacted more severely in comparison to the general population (p. 11). In addition, a Chinese study (Wang et al., 2020) examined the psychological impacts and mental health status of respondents (n = 1,210) through the use of two scales measuring two constructs: the *Impact of Event Scale-Revised* and the *Depression, Anxiety and Stress Scale*, respectively. Similarly to the previous two studies, female gender and student status were associated with a greater psychological impact as well as higher measurements of depression, anxiety and stress. In addition to these three studies, a Hungarian questionnaire study by Szabó et al. (2020) also showed the impact that the pandemic has on the mental health of those in lockdown. In this study, similarly to those previously mentioned, young people and women reported increased levels of stress during the initial lockdown period which occurred in April when the data was gathered. Examining these studies provides a snapshot into the experiences of learners during the pandemic and the effects that the pandemic had on them.

However, it should also be noted that all of the previously mentioned studies were quantitative in nature and did not gather more detailed information on the unique circumstances faced by individual participants. An exception was a study by Son et al. (2020), which provided a qualitative perspective into the effects of COVID-19 on the mental health of university students in the US through a mixed-methods study. A number of themes arose in relation to the difficulties faced by the students, which included worrying about themselves or others, distraction, lack of motivation, monotony, irregular sleep patterns, reduced interaction, challenges with online classes, decreased appetite, and uncertainty, among others. Another study investigating French university students showed similar findings (Le Vigouroux et al., 2021).

The Pedagogical Impacts of the Pandemic

Students and teachers involved in language programs around the world have been deeply impacted by the 2020 COVID pandemic (Marinoni et al., 2020), and these impacts continue as this article is being written. In the past year, a number of articles have been published investigating how students and teachers involved in language teaching were affected by the move to distance learning in schools around the world. Such studies act as a window into the behaviours and emotions associated with those involved in distance language learning.

Some insight into the difficulties faced by students has been gathered from their instructors. In a study by Bailey and Lee (2020), a group of EFL teachers were provided with a questionnaire eliciting data on what they felt were the challenges as well as benefits associated with distance learning for their students as well as themselves. The study showed that, on average, instructors perceived that their students had trouble with online collaboration, attendance, self-regulated learning, and distraction, among other issues (p. 198). Although the study gathered important information from teachers, it lacks the more detailed and personal data that could be gained from an emic perspective gathered from the students themselves.

A number of recent studies have put a greater focus on the impacts that the pandemic has had on the learning experiences of students themselves. A more learner-focused study by Wargadinata et al. (2020) in the Indonesian context explored the digital resources and learning activities of students engaged in autonomous Arabic learning during the lockdown. While this study provided insight into the autonomous language learning process used by the participants, it lacks rich qualitative data from the learners that could help researchers to learn more about the perceived impacts of the lockdown on their learning experiences. A more recent study providing a perspective from language learners carried out by Maican and Cocoradă (2021) revealed the presence of positive and negative emotions among students in regard to distance learning, and participants identified

stressors as well as coping behaviours for dealing with them. The studies mentioned above provide details regarding the learning experiences of students during the lockdown and help to paint a more detailed picture of how these unique conditions impacted their learning process.

The studies discussed in this section have shown that while there has been some initial research into the impacts of the pandemic on students, there is still little research specifically focused on language learners. Research into autonomous language learning during the pandemic is even scarcer. Studies investigating this context have not gathered an in-depth perspective into student experiences in the autonomous language learning context during the COVID pandemic, highlighting the need for interview studies which can provide such perspectives. This study aims to address this gap in the research in order to better understand the impact that the pandemic had on autonomous language learners.

Methods

The preceding section showed that the effects of the COVID lockdown on the autonomous language learning process, and language learning in general, are under-researched, and that there is insight to be gained from an emic perspective into the experiences of language learners subject to lockdown conditions. In light of this research gap, this article will focus specifically on the effects of the lockdown on language learning as perceived by Hungarian university students engaged in autonomous language learning, and is centred on the following research question:

> What were the perceived impacts of lockdown conditions over a two-month period on the autonomous language learning experiences of Hungarian university students?

To answer the research question presented above, a longitudinal qualitative study was designed with three data collection points at the beginning, middle, and end of a two-month period. The context examined in

this study warranted the use of interviews to provide an exploratory approach to understanding the unique learning situations faced by the students. Longitudinal data collection was used in order to gain insight into the ways in which the perceived effects of the lockdown changed over time, as well as the motivation and affective factors of the students. As discussed in Ortega and Iberri-Shea (2005), language learning represents "a complex process that happens through and over time" (p. 26), the study of which warrants the use of longitudinal research design to uncover the ways that learners and the learning process develop over time.

Participants

The participants in the study were chosen from a population of university students and had to be engaged in language learning activities in their free time, either completely independently or alongside formal/informal instruction. The participants, who were either in the English programme at their university or teacher trainers, all took part in English-medium courses and were learning an additional non-English language. These participants were reached out to prior to the start of a government mandated COVID lockdown in Hungary and were all subject to the legal requirements of this lockdown, which included an 20:00 to 5:00 curfew, a ban on dining-in at restaurants, cafes, and bars, and a limit on gatherings of more than 10 individuals, among other measures (Magyar Közlöny [Hungarian official gazette], 2020). Six participants were found through convenience and snowball sampling who met the criteria, all of whom were Hungarian and female. Although the lack of male participants was not intentional and simply occurred as a result of the sampling used, the sample nonetheless provides insight into a group shown by previous research to be especially prone to the negative mental health impacts of the lockdown: young women (Banks & Xu, 2020; Rossi et al., 2020; Szabó et al., 2020). Participants all ranged between the ages of 20-22 and possessed a good command of the English language. In addition, participants provided their self-reported

proficiency in the target languages that they were learning. Pseudonyms were given to each participant to protect their anonymity. This information on the participants can be found below in Table 2.

Table 2

Information on the Six Participants

Pseudonym	Age	Target language	Proficiency
Bogi	20	German, Korean	B2-C1, A2-B1
Fanni	21	Italian	B2
Judit	21	Italian	B1
Kati	22	Spanish	A1
Rose	22	Japanese	N5-N4*
Viki	21	German	A2

*The Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2001) was used by all of the participants in their self-reported data, with the exception of Rose, who reported using the Japanese Language Proficiency Test (JLPT) levels (N5 being the lowest, N1 being the highest).

The Instrument

Semi-structured interviews were used to gather information about the participants' autonomous language learning during the lockdown period. This interview format was chosen in order to provide flexibility in regard to the participants' responses and to avoid limiting the accounts of the participants (Dörnyei, 2007). In addition, the interview schedules evolved in

different ways to accommodate the unique data collected from each participant, while at the same time maintaining a shared structure followed in each interview. In addition, the longitudinal design allowed for member checking to be carried out, as data collected in previous interviews was followed up on in later interviews with participants, allowing participants to clarify or follow-up on their previous responses. This helped to ensure interpretive validity in the study (Dörnyei, 2007).

Procedure

All participants were interviewed using the same protocol in the first round. The interview guide contained items designed to collect information on the biographical data of the participants in addition to their autonomous learning behaviours, previous language learning experiences, language learning goals, motivation, self-regulatory strategies (drawn from Tseng & Dörnyei's (2006) study on self-regulation), and the perceived impacts the current lockdown had on their current learning efforts. This initial, uniform interview schedule can be found in the Appendix. The next round of interviews followed up on the previous responses, examining how the issues discussed previously had changed and developed over the weeks that passed between the interviews. The last round of interviews continued this follow-up format, while also gathering summative thoughts regarding the overall language learning experience during the two-month lockdown.

Data was collected from the participants through an online video-calling platform because face-to-face interviews were not feasible at this time due to the COVID-19 lockdown. Consent was given by the participants to record the audio from the call and use the data for research purposes. Participants were informed about the aims of the study and were given the opportunity to opt out at any time. In addition to the video calls, messages were also sent in the chat box of the video-calling platform in order to clarify

certain details given by the participants, such as the names of applications and web pages used for the purposes of their autonomous learning.

The interviews were conducted in English for the practical benefit of the researcher, who at the time of the data collection lacked a high level of proficiency in the mother tongue of the participants. Despite being second-language speakers of English, the data collection process went smoothly as the participants possessed high levels of English proficiency. Three rounds of data collection took place with each participant, with each interview taking place approximately one month apart. The government mandated lockdown remained in force throughout the entirety of the study, which provided the participants with the opportunity to reflect on the two months spent under these conditions. The amount of data recorded from the participants ranged from 18 to 45 minutes for each interview (M = 27.5). Throughout the data collection process, new data arose, which led to the introduction of new items in later versions of the interview protocols. This helped to narrow the focus of the interviews towards the issues that appeared most frequently in the data and were thus the most relevant to the specific context under investigation.

Data Analysis

Data analysis was carried out throughout the longitudinal data collection process, with the initial interview data informing the design of interview protocols used in subsequent rounds. After each round of interviews, the recordings were digitally transcribed and thematic analysis was used to classify and categorize the relevant excerpts from the data (Barkhuizen et al., 2014). The first step of the analysis process involved multiple thorough readings of the transcribed data and simultaneous notetaking to code the relevant excerpts which connected the language learning experiences of the participants to the COVID lockdown. After this initial coding process was completed, the codes were examined and

organized into categories, representing the emerging themes from the data. In the case of the first and second round of interviews, these themes were then used to inform the design of proceeding interview protocols and were followed up on in the subsequent rounds of interviews with the participants. After the three rounds of data collection were complete, the author made one more pass through the data from all three rounds of interviews, repeating the previous process with all of the interview transcripts to double-check the coding scheme and reassess the themes emerging from the data. The results of this analysis are included in the following section alongside the discussion of the findings.

Results and Discussion

The emerging themes uncovered through the analysis represent the perceived impacts of the COVID lockdown on the autonomous language learning experiences of the participants. An additional theme was also identified in connection to the role that language learning played in the COVID lockdown. Together, these four themes include: 1.) negative impacts on the mental states of the learner, 2.) positive impacts of the lockdown on the learner, 3.) impacts on the learner's environment, and 4.) role of language learning during the lockdown. The results for each category are presented below alongside a discussion of the findings.

Negative Perceived Impacts on the Mental States of the Learner

A number of negative perceived impacts were reported by the participants in regard to the effects that the COVID lockdown had on their mental states; the data showed that in turn, such perceived impacts also affected the participants' language learning endeavours. The negative influences of the lockdown were mainly related to motivation, depression, and stress, issues which came up repeatedly in the interview data, as well as in a previous study on university students under lockdown (Son et al., 2020).

Similarly to the items used by the Patient Health Questionnaire (Spitzer et al., 1999) to measure depression in the Son et al. study (2020), participants in the present study identified issues such as low energy levels, trouble concentrating, feeling bad about themselves, and having feelings of hopelessness or disinterest. In the Son et al. study, stress was measured in addition to depression using the Perceived Stress Scale (Cohen & Hoberman, 1983) containing items examining the extent to which individuals are upset, nervous, angry, and in control during a specific period. Experiences similar to those addressed by the items of the scale were mentioned by the participants. In regard to issues related to motivation, depression, and stress, it is important to mention that while the lockdown was identified as one of the contributing factors leading to such issues among the participants, other perturbations such as university exams that they were taking during this period also caused mental distress, sometimes even more so than the lockdown. Thus, the impacts discussed above can be seen as an interaction between the COVID lockdown and the other environmental factors in the background of participants.

In regard to the negative impacts of the lockdown on the learner's mental states, motivation was the most discussed theme by the participants, who mainly reported on the ways that the lockdown lowered their motivation to engage in the autonomous language learning process. This can be seen in the excerpt below:

> A lot of things have changed, so my motivation is like, dead. I barely do my things in the rhythm that I should be doing things, and what I noticed is that I can't really enjoy the small things anymore … I feel somehow limited in everything I do in my everyday life, and demotivated in basically everything. (Viki, interview 1)

Similar sentiments were found in the interview data of a number of other participants, with only two of the participants (Bogi and Kati) not reporting negative motivational impacts from the pandemic. As seen in the

excerpt above, motivation is linked to the maintenance (or lack thereof) of a daily routine, something also reflected in the testimony of Judit, who related her lower levels of motivation to difficulties with time management. Such connections between motivation and time management can be seen in previous research, such as in the study by Kormos and Csizér (2014) showing a relationship between intended learning effort and time management. Similar interactions between motivation and self-regulation can also be found in Ding and Shen's (2019) study on autonomous learners taking part in online courses.

Issues related to motivation also appeared alongside reports of depression and exhaustion resulting from the lockdown. When discussing potential challenges that she expected to face in regard to her language learning during the lockdown, Fanni stated that "if I'm going to feel a bit down, then I think my motivation level is also going to be lower" (Interview 1). However, in this case, such challenges to language learning motivation did not materialize, and the enjoyment that she experienced in the language learning process actually increased her motivation to learn, as she highlighted in her second interview:

> These last weeks of the semester are always very hard, so I think everyone wants to find something that they enjoy doing, so for me it was learning Italian… these last few weeks I was learning more, and I became more motivated, I would say. (Fanni, interview 2)

Here the role of the L2 learning experience in the learners' motivational system can be seen; this is in line with Dörnyei's L2 motivational self system (2009). Similar findings regarding the role of enjoyment were also noted in the case of Maican and Cocoradă's (2021) study, with the scale showing negative correlations with stressors and negative emotions which can have a negative impact on motivation.

In addition, an aura of uncertainty surrounded the lockdown and was a cause of stress for some participants. This was reflected in both Viki and

Judit's data. When asked about stress encountered during the language learning process, Viki went into detail about the source of her stress:

> Stress is not directly related to learning the language, but you know, these questions about the future… I have doubts and it comes to my mind when I'm learning the language if it's worth it at all, but it's not because of the learning process but because of the situation that we're in. (Interview 3)

Judit also reflected on the negative effects caused by the uncertainty related to the COVID lockdown:

> This coronavirus situation really affects my motivation level because of this uncertain situation, we don't know what's going to happen … I think that that affects motivation because you don't know if you should learn anything at all because who knows what's going to happen in the next two days. (Interview 3)

The interview excerpts above reflect the seemingly interconnected nature between the elements of stress, uncertainty, motivation, and depression, which appear to impact or draw upon each other in various ways. The interactions between these factors can then be seen to affect the language learning experience of many of the participants. It should also be noted, however, that many of the negative issues experienced improved over time, and some participants became used to the situation in which they found themselves. This can be seen in the second round of interviews with Fanni, who claimed that the lockdown had "kind of become natural" (Interview 2) and Kati, who had "got used to this quarantine thing" (Interview 2). This ability to adapt to difficult circumstances reflects the concept of resilience, which was previously shown to play an important role in language learning and language learning motivation (Kim & Kim, 2017).

The Perceived Positive Impacts of the Lockdown on the Learner

Despite the negative aspects of the lockdown, the participants also mentioned a number of ways in which the lockdown was beneficial to their language learning progress. All of the participants but one reported that without the lockdown, they would have likely spent less time studying the target language. When asked how she thought her language learning would have been different without the lockdown, Fanni stated the following:

> I think I would learn less because I could go meet my friends. I could go do things instead of staying at home. If I'm at home all day then I always have at least one hour to sit down and study Italian, but if I could go anywhere I would probably spend that one hour with something else (Interview 3).

Similar sentiments were expressed by the other participants except for Rose, who believed that she would have studied the same amount of Japanese regardless of whether the lockdown had been in place or not. Other unique benefits from the lockdown were reported by participants. Unlike the other participants who had previous experience with their target language, Kati began learning hers during the pandemic. In the first interview, she described how she was inspired to learn Spanish after watching a Spanish-language series on Netflix, and in the third interview she claimed that if it had not been for the lockdown conditions, she likely would not have started learning the language. Bogi also appeared to benefit from the lockdown, providing the following statement on how she believed the lockdown contributed to her motivation:

> Since I have to spend so much time at home, I think I will feel bad for not making time to learn languages, and then I will feel guilty, and then I will try to do something about that. So maybe my motivation will be a bit higher because of that. (Bogi, Interview 1).

The statement above shows interaction between different affective states occurring in response to a perturbation such as the COVID lockdown. While many participants were demotivated by the circumstances, Bogi's attitudes towards the situation and her own internal motivational disposition produced a unique response to the lockdown. Such a case shows the complex nature of language learning motivation, and the various ways in which it can interact with internal and external factors related to the learner (Ushioda, 2015). Such positive reflections on the lockdown period can also be seen as a positive reframing of a negative situation, something which was also seen in studies examining university students in the US (Son et al., 2020) and in France (Le Vigouroux et al., 2021).

Perceived Impacts on the Learner's Environment

In addition to its effects on the mental states of the participants, the lockdown also had a perceived impact on the learning environments of the participants. Many of these changes negatively influenced their language learning efforts during the lockdown, while other changes limited the type of language learning activities that participants were able to engage in during the lockdown period.

The most common issue raised by participants was in regard to disturbances in their learning environment at home due to the lockdown conditions. In her first interview, Rose discussed distractions caused by her four siblings, who, due to the lockdown, were all taking part in distance learning at home. This additional background noise led to difficulties with concentration. Similar problems were mentioned by Judit, who reported increased distraction resulting from her mother who was working from home due to the lockdown. Bogi, who was forced to leave the student dormitories due to the lockdown, encountered disturbances to her language learning efforts when she went back to her hometown. These took the form of distraction caused by the presence of family members, as well as family

responsibilities which took up a portion of her time. Such disturbances caused by family members were also reported in a study carried out in Romania by Maican and Cocoradă (2021) and represent a loss of environmental control (Tseng et al., 2006) by the learners.

The other main issue which was reported in this category was the interference by the lockdown in the language learning activities of the participants. Some participants, such as Viki and Kati, encountered difficulties with finding a private tutor or conversation partner to supplement their autonomous learning activities because of the lockdown. Viki also lamented not being able to attend language exchange events that would normally be taking place in Budapest had it not been for the lockdown. One participant, Rose, also experienced changes in learning activities which occurred due to her Japanese class shifting from offline to online. This shift led to a decrease in the amount of vocabulary covered in the course, which prompted Rose to increase her own autonomous vocabulary learning activities at home. Similar increases in the autonomous learning behaviours of university students after the implementation of distance learning in response to COVID-19 lockdowns were also noted in Gonzalez et al. (2020).

The Role of Language Learning during the Lockdown

The final theme which emerged from the data reflects the role that the language learning process played for the participants during their time under lockdown. In contrast to the previous section, which contained reflections on how the lockdown was beneficial to the language learning process, this theme included reflections on how the language learning process improved the learner's conditions during the lockdown. It should be noted here that all of the participants expressed enjoyment of the language learning process, which in some cases raised their spirits despite the debilitating circumstances surrounding the lockdown.

In the final interview, based on comments in the previous interviews regarding their enjoyment of the language learning process, participants were asked whether they would recommend language learning as an activity to a bored friend during the lockdown, and why. This item produced a number of insights into the benefits of language learning as perceived by the participants. Alongside their comments on the enjoyable nature of language learning, some participants also reflected on an aversion to "wasting their time", as seen in the following excerpts:

> "I think that when people learn new things they can feel more useful, like, they spent their time in a useful way instead of wasting it, and I think that's a great feeling that's important to me." (Rose, Interview 3)

> "They [language learners in lockdown] could keep themselves busy, and I guess they wouldn't feel like they were wasting their time if they tried to learn a new language. It's also fun." (Fanni, Interview 3).

Such notions reflect the virtually unanimous sentiment felt among the participants that their language learning endeavours contributed positively to their lockdown experience. While not all types of language practice were deemed as enjoyable (some participants, such as Judit, were not fond of learning the grammar rules of their target language), there was a general consensus that language learning was a beneficial and enjoyable activity to take part in during the idle time at home resulting from the lockdown conditions. In the previously mentioned studies by Son et al. (2020) and Le Vigouroux et al. (2021), coping mechanisms including keeping routines, positive reframing, and distracting oneself with other tasks were all used by the students interviewed. As the findings in the current study show, language learning provided similar opportunities for the participants in lockdown, providing an activity that was enjoyable as well as rewarding. Kubota (2011) also examined such perspectives of language learning by investigating the role of English learning as an enjoyable leisure activity in

Japan, highlighting the positive role that language learning can have for learners who enjoy the learning process.

Conclusion

This study investigated the perceived impacts of the 2020-2021 COVID lockdown on Hungarian university students engaged in autonomous language learning activities. The findings revealed the participants' thoughts regarding the negative impacts of the lockdown on their mental states, as well as the positive effects that it had on their language learning efforts. Participants also recollected ways in which the lockdown had changed their learning environment and elaborated on the role that language learning had played in their lives during the two-month lockdown period.

Throughout the lockdown period, lower levels of motivation as well as stress and feelings of depression were reported by the participants as a result of the lockdown which were in line with previous research into the impact of COVID on the general population (e.g., Rossi et al., 2020) as well as students specifically (Son et al., 2020). Positive aspects of the lockdown period were also noted by the participants, who reported that the lockdown gave them time to engage in language learning that they may not have taken advantage of under normal conditions. Similar positive impacts of the COVID lockdown were also reported in Gonzalez et al. (2020) in which the lockdown appeared to have a positive impact on learning outcomes for the students studied. Further research into how these positive and negative impacts coincide and relate to one another could be more closely studied in future research.

In addition to the perceived negative impacts on their mental states, reports from the learners also elaborated on the ways in which the lockdown had an effect on their learning environment, with changes in the home-lives of the participants reported alongside limitations in the types of autonomous learning activities in which they could engage. As Murray and Lamb (2018)

highlighted, autonomy is "inextricably linked to space and place" (p. 254) and cannot be understood without examining the context in which it takes place. Through qualitative data collection, information about the context can be gathered from the participants and insights can be gained into how these contextual factors impact the autonomous learning experience.

An additional theme emerged in the data which, although not directly related to the research question, provided insight into the perceived benefits of engaging in the autonomous language learning process during the lockdown. Participants saw their language learning activities as an enjoyable way to pass the time spent at home during the lockdown, and many shared their thoughts on the sense of fulfilment that the learning process brought to them during this period. The identification of language learning as a sort of coping mechanism used during difficult times, while introduced in Maican and Cocoradă's (2021) study into language learning during the COVID lockdown, has yet to be seriously investigated in applied linguistics research, warranting future research in this direction.

In terms of methodological implications, the longitudinal data collection used in this study revealed some of the ways that the participants developed and adapted to the lockdown conditions over the two-month period. Studies such as these can provide key emic insights into autonomous language learning for applied linguistics researchers trying to better understand a highly prevalent but as of yet under-researched topic. This context continues to show increasing importance considering the continuously emerging new opportunities for autonomous learning provided by digital technology, as well as calls for rethinking our current conceptualization of autonomous language learning (Sockett & Toffoli, 2012). While the qualitative data gathered in the study provided a rich account of the language learning experiences of the participants during the lockdown, the addition of quantitative data would provide further insight into the various affective and contextual processes outlined in this study and their development over time. With the help of quantitative data, specific factors

related to motivation or self-regulation could be measured over time and trajectories could be drawn to show their development, as was carried out in studies drawing from a dynamic systems theory perspective (e.g., Piniel & Csizér, 2015).

Although the current study focuses on autonomous language learning outside of the formal classroom setting, the findings can still have implications for language teachers. As studies of successful language learners have shown (Griffiths, 2008), learners who take advantage of resources outside the classroom have greater chances of success. By carrying out in-depth, detailed studies of autonomous language learning using longitudinal designs such as the one used in this study, teachers have a chance to not only gain a better understanding of the issues learners face outside of the classroom but also of how these issues manifest over time. Future research could also gather information about the specific language learning activities that students engage in through the use of technology, such as Skype (Tian & Wang, 2010) or DuoLingo (Lowen et al., 2019). By exploring autonomous language learning from different angles and understanding it more holistically, teachers will be able to direct students to appropriate resources and give helpful advice to students on how to take control of their language learning endeavours.

References

Arnold, J. (Ed.) (1999). *Affect in language learning*. Cambridge University Press.

Bailey, D. R., & Lee, A. R. (2020). Learning from experience in the midst of COVID-19: Benefits, challenges, and strategies in online teaching. *CALL-EJ*, *21*(3), 178–198. http://callej.org/journal/21-2/Bailey-Lee2020.pdf

Banks, J., & Xu, X. (2020). The mental health effects of the first two months of lockdown and social distancing during the Covid-19 pandemic in the UK. *IFS Working Papers*, 1–28. https://dx.doi.org/10.1920/wp.ifs.2020.1620

Barkhuizen, G., Benson, P., & Chik, A. (2014). *Narrative inquiry in language teaching and research.* Routledge.

Benson, P. (2001). *Teaching and researching learner autonomy in language learning*. Longman.

Benson, P. (2006). Autonomy in language teaching and learning. *Language Teaching*, *40*, 21–40. https://doi.org/10.1017/S0261444806003958

Berk, L. E., & Winsler, A. (1995). *Scaffolding children's learning: Vygotsky and early childhood education*. National Association for the Education of Young Children.

Chakraborty, I., & Maity, P. (2020). COVID-19 outbreak: Migration, effects on society, global environment and prevention [Manuscript in preparation]. *Science of the Total Environment*, *728*, 1–20. https://doi.org/10.1016/j.scitotenv.2020.138882

Cohen, S., & Hoberman, H. M. (1983). Positive events and social supports as buffers of life change stress. *Journal of Applied Social Psychology*, *13*(2), 99–125. https://doi.org/10.1111/j.1559-1816.1983.tb02325.x

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching and assessment*. Cambridge University Press.

Ding, Y., & Shen, H. (2019). Delving into learner autonomy in an EFL MOOC in China: A case study. *Computer Assisted Language Learning*, *32*, 1–23. https://doi.org/10.1080/09588221.2019.1681464

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

Dörnyei, Z. (2009). The L2 motivational self system. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 Self* (pp. 9–42). Multilingual Matters.

Dörnyei, Z., Csizér, K., & Németh, N. (2006). *Motivation, language attitudes and globalisation: A Hungarian perspective*. Multilingual Matters.

Egbert, J. (2020). The new normal? A pandemic of task engagement in language learning. *Foreign Language Annals*, *53*, 314–319. https://doi.org/10.1111/flan.12452

Fernandez-Toro, M., & Jones, F. R. (2001). *DIY techniques for language learners*. CILT.

Gao, C., & Shen, H. (2021). Mobile-technology-induced learning strategies: Chinese university EFL students learning English in an emerging context. *ReCALL*, *33*(1), 88–105. https://doi.org/10.1017/S0958344020000142

Gao, X., & Lamb, T. (2011). Exploring links between identity, motivation and autonomy. In G. Murray, X. Gao, & T. Lamb (Eds.), *Identity, motivation and autonomy in language learning* (pp. 1–10). Multilingual Matters.

García-Alberti, M., Suárez, F., Chizón, I., & Feijoo, J. C. M. (2021). Challenges and experiences of online evaluation in courses of civil engineering during the lockdown learning due to the COVID-19 pandemic. *Education Sciences*, *11*, 1–19. https://doi.org/10.3390/educsci11020059

Gonzalez, T., de la Rubia, M. A., Hincz, K. P., Comas-Lopez, M., Subirats L., Fort, S., & Sacha, G. M. (2020). Influence of COVID-19 confinement on students' performance in higher education. *PLoS ONE*, *15*(10), 1–23. https://doi.org/10.1371/journal.pone.0239490

Griffiths, C. (2008). *Lessons from good language learners.* Cambridge University Press.

Holec, H. (1981). *Autonomy in foreign language learning*. Pergamon.

Hurd, S. (2007). Distant voices: Learner's stories about the affective side of learning a language at a distance. *Innovation in Language Learning and Teaching*, *1*(2), 242–259. http://dx.doi.org/10.2167/illt020.0

Japanese Language Proficiency Test. https://www.jlpt.jp/e/about/levelsummary.html

Kim, T., & Kim, Y. (2017). The impact of resilience on L2 learners' motivated behaviour and proficiency in L2 learning. *Educational Studies*, *43*(1), 1–15. https://doi.org/10.1080/03055698.2016.1237866

Kormos, J., & Csizér, K. (2014). The interaction of motivation, self-regulatory strategies, and autonomous learning behaviour in different learner groups. *TESOL Quarterly*, *48*(2), 275–299. https://doi.org/10.1002/tesq.129

Kubota, R. (2011). Learning a foreign language as leisure and consumption: Enjoyment, desire, and the business of *eikaiwa*. *International Journal of Bilingual Education and Bilingualism*, *14*(4), 473–488. https://doi.org/10.1080/13670050.2011.573069

Lai, C. (2017). *Autonomous language learning with technology: Beyond the classroom.* Bloomsbury Academic.

Le Vigouroux, S., Goncalves, A., & Charbonnier, E. (2021). The psychological vulnerability of French university students to the COVID-19 Confinement. *Health Education & Behavior*, *48*(2), 123–131. https://doi.org/10.1177/1090198120987128

Lee, S. J., & Lee, K. (2021). The role of informal digital learning of English and L2 motivational self system in foreign language enjoyment. *British Journal of Educational Technology*, *52*(1), 358–373. https://doi.org/10.1111/bjet.12955

Loewen, S., Crowther, D., Isbell, D. R., Kim, K. M., Maloney, J., Miller, Z. F., & Rawal, H. (2019). Mobile-assisted language learning: A Duolingo case study. *ReCALL*, *31*(3), 293–311. https://doi.org/10.1017/S0958344019000065

MacIntrye, P. D., & Gregersen, T. (2012). Affect: The role of language anxiety and other emotions in language learning. In S. Mercer, S. Ryan, & M. Williams (Eds.), *Psychology for language learning* (pp. 103–118). Palgrave Macmillan.

Magyar Közlöny, A Kormány 478/2020. (XI. 3.) *Korm. rendelete a veszélyhelyzet kihirdetéséről.*

Maican, M. A., & Cocoradă, E. (2021). Online foreign language learning in higher education and its correlates during the COVID-19 pandemic. *Sustainability*, *13*(2), 1–21. https://doi.org/10.3390/su13020781

Marinoni, G., Van't Land, H., & Jensen, T. (2020). *The impact of COVID-19 on higher education around the world.* International Association of Universities. https://www.iau-aiu.net/IMG/pdf/iau_covid19_and_he_survey_report_final_may_2020.pdf

Markus, H., & Nurius, P. (1986). Possible selves. *American Psychologist*, *41*(9), 954. https://doi.org/10.1037/0003-066X.41.9.954

Mizuki, P. (2003). Metacognitive strategies, reflection, and autonomy in the classroom. In A. Barfield & M. Nix (Eds.), *Learner and teacher autonomy in Japan 1: Autonomy you ask!* (pp. 143–153). Learner Development Special Interest Group of the Japan Association of Language Teachers.

Murray, G., & Lamb, T. (2018). Space, place, autonomy and the road not yet taken. In G. Murray & T. Lamb (Eds.), *Space, place and autonomy in language learning*. (pp. 249–262). Routledge.

Ortega, L., & Iberri-Shea, G. (2005). Longitudinal research in second language acquisition: Recent trends and future directions. *Annual Review of Applied Linguistics*, *25*, 26–45. https://doi.org/10.1017/S0267190505000024

Palau, R., Fuentes, M., Mogas, J., & Cebrián, G. (2021). Analysis of the implementation of teaching and learning processes at Catalan schools during the COVID-19 lockdown. *Technology, Pedagogy and Education*, *30*(1), 183–199. https://doi.org/10.1080/1475939X.2020.1863855

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

Rossi, R., Socci, V., Talevi, D., Mensi, S., Niolu, C., Pacitti, F., Di Marco, A., Rossi, A., Siracusano, A., & Di Lorenzo, G. (2020). COVID-19 Pandemic and lockdown measures impact on mental health among the general population in Italy. *Frontiers in Psychiatry*, *11*, 1–6. https://doi.org/10.3389/fpsyt.2020.00790

Schwienhorst, K. (2007). *Learner autonomy and CALL environments*. Routledge.

Sockett, G., & Toffoli, D. (2012). Beyond learner autonomy: A dynamic systems view of the informal learning of English in virtual online communities. *Procedia–Social and Behavioral Sciences*, *34*, 212–215. https://doi.org/10.1016/j.sbspro.2012.02.042

Son, C., Hegde, S., Smith, A., Wang, X., & Sasangohar, F. (2020). Effects of COVID-19 on college students' mental health in the United States: Interview survey study. *Journal of Medical Internet Research*, *22*(9), 1–14. https://doi.org/10.2196/21279

Spitzer, R. L., Kroken, K., & Williams, J. B. (1999). Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. Primary care evaluation of mental disorders. Patient health questionnaire. *JAMA*, *282*(18), 1727–1744. https://doi.org/10.1001/jama.282.18.1737

Szabó, A., Ábel, K., & Boros, S. (2020). Attitudes toward COVID-19 and stress levels in Hungary: Effects of age, perceived health status, and gender. *Psychological trauma: Theory, research, practice, and policy*, *12*(6), 572–575. https://dx.doi.org/10.1037/tra0000665

Taguchi, T., Magid, M., & Papi, M. (2009). The L2 motivational self system amongst Chinese and Japanese learners of English: A comparative study. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 66–97). Multilingual Matters.

Thomson, C. K. (1998). Junior teaching internship: Promoting cooperative interaction and learner autonomy in foreign language classrooms. *Foreign Language Annals*, *31*(4), 569–583. https://doi.org/10.1111/j.1944-9720.1998.tb00600.x

Tian, J., & Wang, Y. (2010). Taking language learning outside the classroom: Learners' perspectives of eTandem learning via Skype. *Innovation in Language Learning and Technology*, *4*(3), 181–197. https://doi.org/10.1080/17501229.2010.513443

Trebbi, T. (2003). Curriculum development and learner autonomy in the foreign language classroom: Constraints and possibilities. In D. Little, J. Ridley, & E. Ushioda (Eds.), *Learner autonomy in the foreign language classroom: Teacher, learner, curriculum and assessment* (pp. 166–184). Authentik.

Tseng, W. T., Dörnyei, Z., & Schmitt, N. (2006). A new approach to assessing strategic learning: The case of self-regulation in vocabulary acquisition. *Applied Linguistics*, *27*, 78–102. https://doi.org/10.1093/applin/ami046

Ushioda, E. (2015). Context and complex dynamic systems theory. In Z. Dörnyei, P. D. MacIntyre, & A. Henry (Eds.), *Motivational dynamics in language learning* (pp. 139–163). Multilingual Matters.

Wang, C., Pan, R., Wan, X., Tan, Y., Xu, L., Ho, C. S., & Ho, R. S. (2020). Immediate psychological responses and associated factors during the initial stage of the 2019 coronavirus disease (COVID-19) epidemic among the general population in China. *International Journal of Environment Research and Public Health*, *17*(5), 1–25. https://doi.org/10.3390/ijerph17051729

Wang, T., & Liu, Y. (2017). Dynamic L3 selves: A longitudinal study of five university L3 learners' motivational trajectories in China. *The Language Learning Journal*, *48*(2), 201–212. https://doi.org/10.1080/09571736.2017.1375549

Wargadinata, W., Maimunah, I., Febriani, S. R., & Humaira, L. (2020). Mediated Arabic language learning for Arabic students of higher education in COVID-19 situation. *Izdihar: Journal of Arabic Language Teaching, Linguistics, and Literature*, *3*(1), 59–78. https://doi.org/10.22219/jiz.v3i1.11862

White, C. (2003). *Language learning in distance education.* Cambridge University Press.

Appendix

The interview guide for the first round of interviews

Obtaining consent

Today I am here with [pseudonym]. I would like to interview you in regard to your plans to practice a foreign language at home during the current COVID-19 lockdown. The data collected from this interview are going to be anonymous and will be used for research purposes only. You may opt out of the research study at any point. The interview is expected to take between 30-40 minutes. Do you consent to taking part in this interview?

Biographical information and background

How old are you?
What do you study?
Where are you staying during the lockdown?
How would you describe your general mental state during the 2020 COVID pandemic?

Language learning

What language do you plan to learn during the lockdown?
Why did you choose this language?
What is your previous experience with this language?
Do you enjoy learning this language?
Are you interesting in the culture of the target language?
How would you describe your current proficiency in the target language?
Are there any other languages you speak besides English, Hungarian, and this language?

Are you taking any formal or informal lessons in addition to your autonomous language learning?
What are your goals in regard to learning the target language?
How long do you think it will take to achieve your goal?
How often do you practice the target language (days per week, hours per day)?
What materials, applications, or resources do you use to study the target language?
Are there any challenges that you expect to face in your language learning process over the next month?

Affective/motivational factors, self-regulation

How would you describe your current level of motivation to learn the target language?
What are the things that have a negative impact on your motivation?
What are the things that have a positive impact on your motivation?
Do you experience issues with distraction during the language learning process?
Do you experience issues with boredom during the language learning process?
Do you experience stress during the language learning process?
What is your ideal study environment?
Are you able to achieve your ideal study environment under the current circumstances?

Closing

Is there anything else you would like to add in relation to the things we talked about today?
Thank you very much [name] for sitting down with me today. If you agree, I will contact you in about three weeks for a second interview to see how the learning process went over the course of the three weeks.

In harmony with the spirit of the first volume of the series, in this publication we sought to foreground the 'power of the local' by addressing issues of direct relevance to our immediate and broader local contexts, and concomitantly to contribute to the academic discourse on current applied linguistics concerns. The studies featured in our volume provide further insights into the diverse research topics explored by the faculty and graduate students at the Department of English Applied Linguistics as well as evidence the professionalism and dedication with which they engaged in their investigations in the course of 2021. The research reported also constitutes pertinent input for our content courses, and we hope it will inspire the research activities of our students.

Gyula Tankó and Kata Csizér, Editors

ISBN 978-963-489-423-0