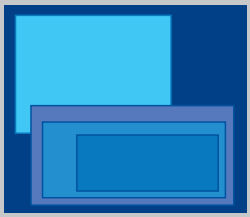


VIAL

VIGO INTERNATIONAL JOURNAL of APPLIED LINGUISTICS

Number 22-2025

Universidade Vigo

Servizo
de Publicacións

VIAL

Vigo International Journal
of Applied Linguistics

Universidade Vigo

VIAL. Vigo International Journal of Applied Linguistics.

VIAL is an international and peer-reviewed journal which is indexed and abstracted in CINDOC (Centro de Información y Documentación Científica), EBSCO, ERIH PLUS, JCR, Web of Knowledge (Arts and Humanities Citation Index, Social Sciences Citation Index), Linguistics Abstracts, Linguistics and Language Behavior Abstracts, Latindex, MLA Directory of Periodicals, PsychoInfo, Scimago Journal Rank Scopus, Ulrich's Database

Webpage: www.webs.uvigo.es/vialjournal/

Printed in Spain - Impreso en España
I.S.S.N.: 1697-0381 - Electronic ISSN: 2660-504X - Depósito Legal: VG-935-2003
Imprime e maqueta: Tórculo Comunicación Gráfica, S.A. Deseño da portada: Elena Gómez.

VIAL

Vigo International Journal of Applied Linguistics

Number 22 - 2025

Contents:

Assessing the impact of the metacognitive pedagogical cycle on adult EFL learners' proficiency and awareness: Evidence from Chilean classrooms

Marco Cancino
Department of English
Facultad de Educación y Ciencias Sociales
Universidad Andres Bello, Chile
marco.cancino@unab.cl

Natalia Oyarzún
Department of English
Facultad de Educación y Ciencias Sociales
Universidad Andres Bello, Chile
n.oyarzunopazo@uandresbello.edu

Abstract

The benefits of metacognitive instruction for the improvement of students' L2 listening comprehension and metacognitive awareness have been documented in the literature. However, it is necessary to evaluate the impact of a process-based approach such as the metacognitive pedagogical cycle (MPC) on less-skilled English as a foreign language (EFL) learners in contexts with reduced out-of-classroom interaction. Therefore, the present quasi-experimental study sought to assess the impact of the MPC on the listening comprehension and listening metacognitive awareness of 75 low-proficiency adult EFL learners. Participants were divided into an experimental group, which received MPC instruction, and a control group that received regular listening instruction over a twelve-week period. Pre- and post-listening proficiency and metacognitive awareness test results revealed that the experimental group significantly increased their proficiency and metacognitive awareness scores in a more consistent and robust manner than the control group. Additionally, learners who were exposed to MPC instruction outperformed the control group in almost all aspects of metacognitive awareness. Results indicated that less skilled learners could benefit from a metacognitive approach to listening instruction in the selected EFL setting. Pedagogical implications are discussed.

Keywords: metacognitive pedagogical cycle; listening comprehension; process-based approach; metacognitive awareness; EFL learning.

Resumen

Los beneficios de la instrucción metacognitiva en la mejora de la comprensión auditiva de la L2 y la conciencia metacognitiva de los estudiantes se han documentado en la literatura. Sin embargo, es necesario evaluar el impacto del ciclo pedagógico metacognitivo (MPC) en estudiantes menos cualificados en contextos de inglés como lengua extranjera (EFL) en donde existe una reducida interacción fuera del aula. El presente estudio cuasiexperimental buscó evaluar el impacto del MPC en la comprensión auditiva y la conciencia metacognitiva auditiva de 75 estudiantes adultos de EFL con bajo nivel de competencia. Los participantes se dividieron en un grupo experimental que recibió instrucción MPC y un grupo de control que recibió instrucción auditiva tradicional durante un período de doce semanas. Los resultados revelaron que el grupo experimental aumentó significativamente sus puntajes de competencia auditiva y conciencia metacognitiva de una manera más consistente y sólida que el grupo de control. Además, los alumnos que estuvieron expuestos a la instrucción MPC superaron al grupo de control en casi todos los aspectos de la conciencia metacognitiva. Los resultados también indicaron que los estudiantes menos hábiles podrían beneficiarse de un enfoque metacognitivo para la instrucción auditiva en el entorno educacional seleccionado. Se analizan las implicaciones pedagógicas.

Palabras clave: ciclo pedagógico metacognitivo; comprensión auditiva;enfoque basado en procesos; conciencia metacognitiva; aprendizaje de inglés como lengua extranjera.

1. Introduction

The importance of comprehension skills for successful communication has been underscored in the literature (Kobayashi, 2018; Matsumoto, 2011). According to Celce-Murcia and Olshtain (2000), the most frequently used communication skill in people's day-to-day interactions is listening. It is considered a "critical life competency" since it is the first communication skill that humans develop (Worthington & Fitch-Hauser, 2018, p. 4). In second language (L2) learning settings, listening comprehension is regarded as one of the most effective activities to learn a language (Nation & Newton, 2020). Despite its important communicative role in the learning process, listening has been the most neglected and undervalued skill in language teaching (Goh & Vandergrift, 2022). Additionally, many L2 learners often regard it as a difficult language skill to master (Hasan, 2000). These outcomes are to a certain extent influenced by the teaching approach to which learners are exposed. Some teachers tend to equate teaching L2 listening comprehension with testing it (Graham, Santos, & Francis-Brophy, 2014), which emphasizes a product-oriented focus (Nazari, 2018).

Listening comprehension instruction typically involves listening to recordings and providing answers to comprehension questions that seek to elicit how much (or how little) has been understood. This approach increases anxiety, which in turn hinders the development of listening comprehension strategies (Goh & Vandergrift, 2022). Thus, it is necessary to assess the impact of more strategic approaches that can enhance listening comprehension development. Vandergrift (2004, 2007) originally proposed a metacognitive approach to listening comprehension that included pedagogical stages through a process-based approach. This approach allows learners to monitor their own cognition and thus control their learning behavior by means of a metacognitive pedagogical cycle (MPC; Goh & Vandergrift, 2022). The MPC is supported by evidence suggesting that highly skilled listeners use a repertoire of metacognitive strategies to regulate their listening process (Goh & Vandergrift, 2022; Vandergrift, 2003). However, this approach needs to be assessed with a focus on low proficiency learners in English as a Foreign Language (EFL) contexts where listening tasks are less prominent in the classroom and are more product-based in nature. Therefore, the present study sought to assess the impact of a process-based metacognitive pedagogical cycle on the listening comprehension and listening metacognitive awareness of low-proficiency adult EFL learners. The study contributes to research describing how adult EFL learners may benefit from a systematic approach that implicitly teaches L2 listening strategies.[1] The research questions are as follows:

1. Does the metacognitive pedagogical cycle (MPC) increase the listening proficiency of adult EFL learners?
2. Does the MPC increase the metacognitive awareness of adult EFL learners?

2. Literature Review

2.1. L2 listening instruction

Listening is an active mental ability that helps individuals understand the world around them and is one of the necessary components to ensure successful communication (Rost, 2016). Moreover, listening is considered to be a complex process due to its psychological – i.e., cognitive – and social – i.e., interactive – nature (Bueno, Madrid, & McLaren, 2006). Vandergrift and Tafaghodtari (2010) recognized the complexity of the skill and the role it has in assisting learners to understand spoken L2 input and facilitating the development of the other language skills. Three types of second language listening instruction approaches are identified by Goh and Vandergrift (2022): Text-oriented, communication-oriented and learner-oriented. Text-oriented instruction drew from grammar translation methodologies (Shintani & Wallace, 2014). More specifically, this approach to listening instruction

prompted learners to recognize and understand the different components of listening input such as sounds and phonological features of key words and phrases (Goh & Vandergrift, 2022). The activities carried out in the approach involved teachers reading aloud written texts as learners completed dictation exercises or cloze-type tasks (Shintani & Wallace, 2014). Learners were given comprehension tasks only after they had listened to a passage, so they typically did not know the purpose of the activity beforehand, which hindered listening development (Brown, 2006). The search for more successful second language learning experiences prompted the emergence of the communicative language teaching (CLT) methodology in the 1970's (Goh & Vandergrift, 2022), which signified a reformulation of the current approaches to syllabus design and methodology at the time (Richards, 2006). Regarding listening instruction, authentic listening materials and pre-listening activities were included to activate prior topic knowledge in the listening task (Shintani & Wallace, 2014). Although both speaking and listening were included in CLT activities, Goh and Vandergrift (2022) argued that listening was often "the sleeping partner in the business of oral communication" and that "little attention was given to learner efforts at listening outside the classroom" (p. 10). This gains particular relevance in EFL settings, where learners do not usually engage in out-of-classroom interaction with L2 speakers. Thus, the focus was shifted to learner-oriented instruction, and the skills and strategies employed by successful listeners (Shintani & Wallace, 2014). Learners were instructed on "how to listen" by means of scaffolded listening practice and teacher modeling (Goh & Vandergrift, 2022). The approach is validated by early research on listening strategies suggesting that successful listeners used a more varied repertoire of listening strategies than unsuccessful learners (Goh, 2000; Vandergrift, 2003). Particularly, successful listeners frequently use metacognitive strategies, which in turn was found to have a positive impact on learners L2 listening development (Goh, 2008; Vandergrift, Goh, Mareschal, & Tafaghodtari, 2006; Vandgergrift & Tafaghodtari, 2010). Although learner-oriented instruction has become an important feature in language curriculum and more recognition has been given to its status and development (Goh & Vandergrift, 2022), the "listen - answer - check" pattern is still followed by many listening instructors (Siegel, 2013).

Goh and Vandergrift (2022) highlight three aspects that help understand the process of becoming a competent listener in a broad range of contexts. First, the cognitive model of language processing identifies two perspectives on listening: bottom-up and top-down processing (Brown, 2006; Lynch, 2006). Brown (2006) and Goh and Vandergrift (2022) refer to bottom-up listening comprehension processing as the segmentation of the sound stream, word meanings, and discourse markers into meaningful units to interpret the message. As for top-down listening comprehension processing, learners use prior knowledge and experiences to activate a conceptual

framework for understanding the message by forming hypotheses and modifying them to match new incoming information (Goh & Vandergrift, 2022). Learners will need to deploy and efficiently coordinate both bottom-up and top-down processing skills to achieve successful L2 listening (Brown, 2006). Proficient listeners are able to control and regulate top-down and bottom-up processing through their use of metacognitive knowledge, which in turn enhances effective text comprehension (Goh & Vandergrift, 2022). Second, listeners rely on a set of knowledge sources (namely, prior knowledge, linguistic knowledge, pragmatic knowledge, and discourse knowledge) in order to understand and make sense of what they hear (Vandergrift, 2015). Prior knowledge is stored in long-term memory (Vandergrift, 2015) and allows listeners to "match what they hear (the linguistic input) with what they know about how things work in the world (their prior knowledge)" (Goh & Vandergrift, 2022, p. 27). This connection can be nurtured in learners by means of pre-listening activities that help them predict and monitor comprehension (Vandergrift & Goh, 2022). Linguistic knowledge – i.e., vocabulary knowledge, phonological knowledge (phonemes, stress, intonation, and speech modifications such as assimilation and elision), and syntactic knowledge (grammar) – is considered fundamental for listening comprehension (Vandergrift & Goh, 2022). In fact, vocabulary has been found to be a strong predictor of effective L2 listening (Vandergrift, 2015). Pragmatic knowledge in listening allows the learner to interpret the speaker's intended message by going beyond the literal meaning of a word (Vandergrift, 2015). Second language listeners need to draw inferences when certain speech acts (e.g., requests) are presented with figurative language and culture-bound content that is closely related to sociocultural and sociolinguistic knowledge (Goh & Vandergrift, 2022). Finally, discourse knowledge refers to comprehension at the level of text organization, which can be combined with prior knowledge (transferred from the learner's L1) so that learners can anticipate the type of information that might be later accessed in a listening task (Vandergrift, 2015) and the type of questions they might be asked. Goh and Vandergrift (2022) state that these knowledge sources "are stored in the listener's long-term memory in the form of schemata (e.g., complex mental structures that group all knowledge concerning a concept)" (p. 19), and that the information retrieved from them plays a role in how cognitive processing is achieved and how successful it is. Third, the environmental features of listening are also relevant to how the process is carried out. Interactive listening refers to "the ability to interact with speakers of the target language in social situations, such as conversations" (Goh & Vandergrift, 2022, p. 29). This type of listening has transactional, interactional, or social goals that are met through successful interactions (Goh & Vandergrift, 2022). One-way listening – i.e., listening to the radio – is only transactional in nature. Its main goal is message or content-oriented (Gu, 2018), and does not require learners to interact with a speaker (Goh & Vandergrift, 2022). Listeners who take part in

either one-way or interactive listening make use of prior (i.e., top-down) knowledge to decode meaning, and linguistic (i.e., bottom-up) knowledge to distinguish sounds that are familiar. This "parallel processing" allows learners to successfully perceive, interpret, and respond to the information being heard (Lynch & Mendelsohn, 2020).

2.2. Metacognition and listening instruction

Metacognition can be defined as "a set of processes an individual uses in monitoring ongoing cognition so as to effectively control his or her own behavior" (Rhodes, 2019, p. 1). This definition positions the individual as the agent who carries out a behavior and, simultaneously, reflects on that behavior. As Goh and Vandergrift (2002, p. 88) state, metacognition allows us to be "agents of our own thinking". This critical and reflective approach to one's thinking "may result in making specific changes in how one learns" (Anderson, 2012, p. 170). Goh and Vandergrift (2022) view metacognition as comprising metacognitive knowledge, metacognitive experience, and strategy use. Metacognitive knowledge allows learners to "select, evaluate, revise, and abandon cognitive tasks, goals, and strategies considering their relationships with one another and with one's own abilities and interests with respect to that enterprise" (Flavell, 1979, p. 908). Flavell (1979) identified three types of knowledge about cognition: person, task, and strategy. Person knowledge refers to how an individual learns and how certain personal factors affect their learning, such as their motivations, strengths, and weaknesses. Task knowledge consists of learners' knowledge of features of different types of real-life listening tasks, such as discourse, grammatical, and phonological features of the words and phrases in the spoken text. Finally, strategy knowledge refers to knowledge of the strategies that can be deployed to complete a specific listening goal or to improve general listening ability (Goh & Vandergrift, 2022). Thus, language learners demonstrate various degrees of metacognitive knowledge about themselves as L2 listeners and the listening process (Graham, 2006; Goh, 1997). Metacognitive experience emerged from Flavell's (1979, p. 906) idea that thinking and learning are accompanied by other "conscious cognitive and affective experiences". These experiences reflect the individual's response to the task before it is solved or the outcome of monitoring the resources they require (e.g., feelings, estimates, and ideas) to solve the task (Akama, 2006). Goh and Vandergrift (2002, p. 90) state that metacognitive experience is useful to learners if it "leads to some productive application of strategies or further understanding about the task, themselves, and/or the world around them". In relation to strategy knowledge, increased interest in nurturing learner autonomy by means of strategy awareness has been documented over the past decades. Goh and Vandergrift (2022, p. 91) define strategy use as "the deployment of specific procedures or actions to make learning easier, faster, more enjoyable, more self-regulated, more effective, or more transferable to new situations".

Therefore, strategy knowledge can be supported by increasing awareness of the ways in which a strategy can be used, and the appropriate time for it to be deployed. Moreover, using strategies prompts cognitive processing, which can lead to deeper learning and improved performance, especially among learners who are struggling with listening comprehension tasks (Wenden, 1998).

2.3. Metacognitive processes and the metacognitive pedagogical cycle (MPC)

Teachers in many EFL contexts rely mostly on approaches whose main focus is placed on assessing learners' ability to answer listening comprehension questions rather than on the process that those learners undergo to achieve that comprehension (Goh 2010; Graham, Santos, & Vanderplank, 2011). Listening instruction research has sought to change its focus from product-based listening (listening to learn) to process-based listening (learning to listen), which has signified a shift toward metacognitive strategies (Lynch & Mendelsohn, 2020). Metacognitive instruction prompts teachers to act as guides, instructing learners on how to listen by enhancing their strategy knowledge and strategy use, and how to work on their listening without teacher support beyond the classroom (Kobayashi, 2018). Hence, a metacognitive process-based approach can assist learners in shifting their focus from listening performance to listening skill development (Bozorgian, 2014) and, honing their ability to use appropriate strategies (Goh, 2008; Maftoon & Alamdari, 2016).

Goh and Vandergrift (2022) state that the goals of a metacognitive approach to listening are to develop learners who understand the challenges of L2 listening, think about and increase their learning development individually and in collaboration, make plans to self-direct and manage their own listening progress, make a proper use of listening strategies, and display self-efficacy and motivation toward listening. Several metacognitive processes underlie these goals. Learners typically have a purpose for listening (e.g., listening to a song, having a conversation), and this purpose shapes the way they listen and helps them listen more effectively (Brown, 2006). Thus, listeners "prepare themselves for what they will hear and what they are expected to do, instead of barreling into the activity without thinking" (Goh & Vandergrift, 2022, p. 116). Then, learners must monitor their comprehension by detecting gaps in their knowledge or identifying knowledge that is deficient. Learners can thus avoid misconceptions that hinder true understanding (Dunlosky & Rawson, 2012; Roelle, Nowitzki, & Berthold, 2017). When learners encounter comprehension difficulties, they can engage in remediation planning in a subsequent step (Glogger et al., 2012). Learners must also solve comprehension problems while performing a listening task in order to reconstruct a text's main points and relevant details (Goh & Vandergrift,

2022). Learners finally evaluate whether the adjustments that were made helped them understand the listening text, which allows them to identify weaknesses in the listening process that can be avoided in future tasks. These metacognitive processes do not necessarily follow a linear pattern, as they can interact in various ways to construct meaning and comprehension (Goh & Vandergrift, 2022). In order to address these aspects by means of a pedagogical approach, Vandergrift (2004, 2007) and Goh and Vandergrift (2022) presented a metacognitive pedagogical cycle (MPC) to guide L2 learners through their metacognitive processes. The MPC involves conscious attention to one's thoughts and knowledge construction (Roelle et al., 2017) by means of five stages: Planning/predicting, first verification, second verification, final verification, reflection and goal setting. Theoretically, the metacognitive pedagogical cycle draws from knowledge about comprehension instruction derived from cognitive psychology, implicit learning, and self-regulation (Vandergrift & Tafaghodtari, 2010). It contains top-down and bottom-up dimensions of listening that can help learners increase their awareness of one-way listening processes, and nurture the creation of metacognitive knowledge, which is a key component of self-regulation and listening comprehension (Goh & Vandergrift, 2022; Vandergrift, 2007). As Roussel et al. (2017, p. 42) state, during the listening comprehension tasks self-regulated language learners are able to "plan, monitor, solve comprehension problems, and evaluate their comprehension in real time". The repeated use of the cycle with a variety of texts allows listeners with different ability levels to make progress and refine their comprehension in their own manner and at their own pace (Cross, 2010). The methodology allows them to access knowledge about listening processes by generating and verifying hypotheses and by applying prior knowledge to offset gaps in understanding (Bozorgian, 2014). Furthermore, the approach has pedagogical value, as its implementation is straightforward and can be carried out with learners of different ages, levels, and educational contexts (Cross 2010; Vandergrift, 2004).

2.4. Empirical research on the metacognitive pedagogical cycle (MPC)

Research conducted in diverse contexts has evaluated the metacognitive approach with a focus on listening outcomes and metacognitive awareness. Vandergrift and Tafaghodtari (2010) investigated the effects of the MPC on 106 French as a second language learners over a semester. They asked students in the experimental group to listen to texts through MPC-guided instruction, while learners in a control group listened to the same texts with regular instruction. Metacognitive development in L2 listening was assessed by means of the Metacognitive Awareness Listening Questionnaire (MALQ), an instrument that includes five factors based on metacognitive processes, namely, planning and evaluation, problem solving, directed attention, mental translation, and person knowledge. Results revealed

that the experimental group significantly outperformed the control group on a comprehension test and evidenced growth in problem solving and mental translation. In addition, listeners who were less skilled made more listening gains in the experimental group than highly skilled listeners. Similar results were reported by Bozorgian and Alamdari (2017), Rahimirad and Shams (2014), and Wang (2015) in EFL contexts. Positive responses to process-based MPC instruction when compared to regular product-based instruction were also reported by Mahdavi and Miri (2017) for listening comprehension and metacognitive awareness in high-beginner EFL learners. In a small-scale study, Cross (2010) measured the impact of the MPC on 20 Japanese advanced adult EFL learners. Pre- and post-test scores suggested that the MPC approach had a stronger impact on less-skilled listeners and that these gains were less prominent at higher skill levels.

In more recent studies, research on metacognitive instruction and the MPC has yielded further evidence of their effectiveness. Becker (2021) investigated the instructional benefits of metacognitive strategies in improving listening comprehension and increasing the automaticity of listening processes. Sixty-nine French as a foreign language learners were assigned to an experimental (metacognitive instruction) or a control group condition. Results revealed a significant improvement in the experimental group that was influenced by initial listening proficiency and metacognitive awareness. That is, low proficiency learners made more listening gains in both conditions, and learners in the experimental group who displayed low starting metacognitive awareness were able to make more listening gains. The author concludes that the baseline metacognitive awareness reported by learners can impact listening comprehension gains over time, which adds support to the idea that nurturing metacognition is beneficial for listening. Taghizade et al. (2022) evaluated the impact of a metacognitive intervention on the listening comprehension and metacognitive awareness of upper-intermediate introvert EFL learners. Their results indicated that the MPC-based intervention significantly enhanced introvert EFL learners' listening comprehension and metacognitive awareness, while learners reported that they were more capable of recognizing their weaknesses and strengths, their anxiety decreased, and their motivation, attention to task, and self-confidence increased. In line with Becker (2021) and Vandergrift and Tafaghodtari (2010), a stronger impact of the metacognitive approach on the listening gains of less skilled learners was reported by Chero (2023), who divided 20 young learners enrolled in an A1 general English course into two groups: skilled listeners and less-skilled listeners. The author found that the intervention significantly increased the scores of less-skilled listeners, while the MALQ questionnaire data revealed that both skilled and less-skilled listeners made significant gains in two MALQ factors (person knowledge and strategy use). Although the number of participants was somewhat reduced for quantitative analysis,

the findings highlight the effectiveness of a metacognitive instruction approach in improving the listening performance of low proficiency learners. Overall, these studies have reported consistent evidence suggesting that metacognitive instruction increases L2 listening comprehension and metacognitive awareness in the EFL classroom. When learners' metacognitive awareness is increased through instruction, their listening comprehension and overall language proficiency can also be enhanced. Still, there is a lack of studies assessing the effectiveness of the MPC in less-skilled learners and in EFL contexts where listening tasks are less frequent in the classroom and more product-based in nature.

3. Methodology

3.1. Research design

To address the research questions, the present study employed a quasi-experimental research design to measure the effect of listening instruction that is based on the metacognitive pedagogical cycle (MPC) on learners' listening comprehension performance and metacognitive awareness. Participants were placed in an experimental group that received MPC instruction and a control group that received regular listening instruction for a period of twelve weeks. Pre- and post-TOEIC Bridge listening tests, together with pre- and post-Metacognitive Awareness Listening Questionnaires (MALQs) were administered to participants in both groups to assess listening comprehension and metacognitive awareness gains, respectively.

3.2. Participants and context

The participants in the study were 75 adult EFL students (20 females, 55 males) enrolled in the third semester of an elementary EFL course in 2022 at a professional institute in Santiago, Chile. These participants belonged to four intact classes that were divided into an experimental (n = 38) and a control (n = 37) group. Participants in both groups received six hours of EFL instruction per week. The teacher delivering the classes in the experimental group was trained in the MPC and received support to plan the listening lessons. The teacher was asked to complete the treatment in a 12-week period, with 90 minutes of MPC instruction per week. The institute where the study took place is constituted as a private non-for-profit foundation that offers both technical and professional careers. Its EFL department serves more than 70.000 students in diverse technical and professional programs. Teachers follow a communicative approach to help EFL learners communicate in the L2 orally and in written form. EFL lessons are typically delivered in classrooms that do not exceed 20 students, and students are able to work with customizable, high-quality, and authentic learning materials from well-known publishers. EFL students sit the TOEIC Bridge

once they finish their EFL courses (elementary and intermediate) and are expected to reach CEFR (Common European Framework Reference for Languages) level B1 after four semesters.

3.3. Instruments

3.3.1. TOEIC Bridge listening section

The listening section of the TOEIC Bridge was administered before and after the 12-week treatment period. This instrument offers reliable global benchmarks for English language proficiency levels among individuals aged 15 and above, whose native language is not English, particularly at beginner and pre-intermediate stages. The listening comprehension segment comprises 50 questions categorized into three question formats: photographs, brief conversations, and question-answer sets. Test-takers are required to listen to spoken content, including statements, inquiries, or short interactions, and choose the accurate response from multiple-choice options. The participants in the two groups were similar in terms of baseline listening proficiency (CEFR level = A1). A *t*-test did not find significant differences (p = .548) between the baseline TOEIC Bridge listening means for the experimental group (M = 19.61) and the control group (M = 18.76).

3.3.2. Metacognitive Awareness Listening Questionnaire (MALQ)

Vandergrift et al. (2006) created the Metacognitive Awareness Listening Questionnaire (MALQ) utilizing Flavell's (1979) framework of metacognition. The purpose of the instrument is to evaluate the metacognitive awareness of second language (L2) listeners and their self-reported utilization of listening strategies. It consists of 21 statements in a six-point Likert scale ranging from "strongly disagree" to "strongly agree". The statements are subcategorized into the five components of metacognitive strategies proposed by Vandergrift et al. (2006) and Goh and Vandergrift (2022): Planning and evaluation (1, 10, 14, 20, and 21), directed attention (items 2, 6, 12, and 16) person knowledge (3, 8, and 15), problem solving (5, 7, 9, 13, 17, and 19) and mental translation (items 4, 11, and 18). Items 3 and 8 in person knowledge and item 16 in directed attention are reverse coded to reduce response bias. All the items in mental translation are also reverse coded since they address online mental translation strategies (e.g., item 4: "I translate in my head as I listen") that should be avoided by L2 learners because they are considered "an inefficient approach to listening comprehension" (Vandergrift et al., 2006, p. 450). The MALQ instrument has been validated through extensive use in related research (Becker, 2021; Goh & Hu, 2013; Taghizade et al., 2022). Reported internal reliability estimates (Cronbach's α) for the instrument are good and range from 0.68 to 0.78 (Vandergrift et al., 2006), which are

similar to the values obtained in the present study (Cronbach's α ranging from 0.65 to 0.77). The MALQ was translated into the learners' first language (Spanish) to ensure comprehension and was then piloted with 40 EFL students in similar EFL courses to identify confusing items. Student feedback was useful to adjust the verb forms and concepts in the L1 translation in some items. For example, some students pointed out that the L1 translation for item 15 "I don't feel nervous when I listen to English" was too vague, as they did not know the type of listening activity the item was referring to. Thus, the L1 translations of several items were modified to include the idea of listening to audio in the context of learning activities carried out in an EFL classroom. Apart from these minor translation issues, no further comprehension issues with the items or the instructions were reported. The translated MALQ items (Oyarzún, 2019) can be found in appendix A.

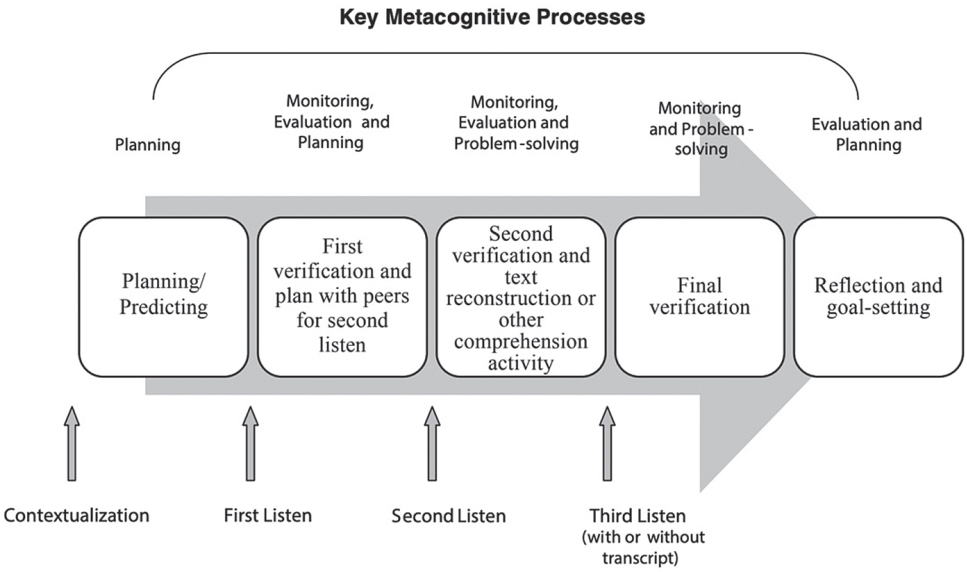
3.4. Instruction approaches

The coursebook used during the 12-week period in both the experimental and control groups was "American English File" (Latham-Koening & Oxenden, 2018). In the experimental group, five listening passages from the units being covered during the intervention weeks were selected and completed following the MPC listening guide worksheet adapted from Goh and Vandergrift (2022). Students in the control group completed the same listening activities but focused on the product-based activities included in the coursebook.

3.4.1. Experimental group

Over the 12-week duration of the study, participants regularly attended their mandatory English classes, during which one 90-minute session per week was dedicated to administering the intervention. Within these sessions, participants engaged with the five listening texts aligned with the topics covered in the English course textbook. Every treatment session was structured following the metacognitive pedagogical cycle (MPC) proposed by Vandergrift (2004, 2007) and Goh and Vandergrift (2022). Figure 1 illustrates the stages within the MPC and how they correspond to the primary metacognitive processes involved in listening.

Figure 1: Stages and processes in the metacognitive pedagogical cycle for listening instruction (Goh & Vandergrift, 2022)



The five stages were incorporated in the lessons plans for the experimental group. In the "planning/predicting" stage, listeners made decisions about their listening task and the steps they could take to make it successful (e.g., students predicted the nature of the information and the potential words they could hear). Then, during the "first verification stage," students verified their initial hypotheses and compared their understanding with that of their peers. In the "second verification stage," students focused on identifying missing information they had not deciphered earlier and resolved any discrepancies in comprehension. The "final verification stage" allowed students to further reconstruct the main points and crucial details of the text. Finally, in the "reflection and goal-setting" stage, students corrected as required, decided on the important details still needing resolution, discussed the strategies used to tackle comprehension issues, and established objectives for their next encounter with a listening task. Appendix B contains a sample lesson plan (Oyarzún, 2019) portraying the MPC stages and the activities carried out.

3.4.2. Control group

Control group participants were asked to listen to the same texts and for the same number of times as the students in the experimental group (three times), but they were not exposed to the MPC stages. They did not receive a listening guide

and did not have the opportunity to verify their predictions or collaborate with peers between listening rounds. They did not have time to review the information that was understood between listening rounds and did not receive a text script to assess their overall listening performance. Moreover, they were not asked to reflect on encountered difficulties or future listening tasks. Instead, they underwent traditional coursebook instruction involving product-based activities, following a conventional "listen-answer-check" approach. In each session, students listened to the text twice, taking notes of any information that was understood, followed by a third listening where they answered comprehension questions or completed charts with the information that they were able to decode. The teacher only intervened to verify task completion accuracy.

3.5. Data collection procedures and method of analysis

Before the study was conducted, informed consent from the participants was secured. They were informed about the characteristics of the intervention and were told that their personal details would not be disclosed and would be safely stored. Two weeks prior to the initial intervention and two weeks subsequent to the final intervention, students completed the listening section of the TOEIC Bridge as a pre-test and post-test, respectively. Thus, listening proficiency was assessed by two different versions of this test. As for the MALQ, the instrument was administered a week before the initial listening session and a week after the sessions in both groups ended. Descriptive statistics were computed for the TOEIC Bridge listening section and the MALQ data. To answer the first research question, a mixed 2x2 (time*listening condition) ANOVA was conducted with the TOEIC Bridge listening section scores. As regards the second research question, a series of mixed 2x2 (time*listening condition) ANOVAs were run with the MALQ components and the total MALQ scores. Data was examined for normality and homogeneity of variances by means of Shapiro-Wilk's tests and Levene's tests, respectively. The tests revealed that, overall, the data for all the variables did not severely depart from normality and displayed fairly equal variance in most instances.

4. Results

4.1. Descriptive statistics for the TOEIC Bridge listening section

Table 1 reports descriptive statistics for the listening section of the TOEIC Bridge. The means for participants in the experimental group prior to the intervention (M = 19.61; SD = 6.08) and following the intervention (M = 27.16; SD = 5.38) indicate a consistent improvement after the treatment. The overall means in the control group before (M = 18.76; SD = 6.11) and after the intervention (M = 20.68; SD = 6.84) suggest rather modest listening gains.

Table 1: Descriptive statistics for TOEIC Bridge listening data

Instrument	Group	Statistic	
TOEIC Bridge listening pre-test	Experimental	Mean	19.61
		SD	6.08
	Control	Mean	18.76
		SD	6.11
TOEIC Bridge listening post-test	Experimental	Mean	27.16
		SD	5.38
	Control	Mean	20.68
		SD	6.84

4.2. Descriptive statistics for the MALQ scores

Table 2 presents descriptive statistics for the overall MALQ and its components pre- and post-intervention.

Table 2: Descriptive statistics for the MALQ and its components

MALQ components	Group	N	Pre MALQ		Post MALQ	
			M	*SD*	M	*SD*
Planning and Evaluation	Experimental	38	4.1	.94	5.19	.44
	Control	37	3.97	.85	4.15	.82
Directed Attention	Experimental	38	4.4	.83	4.74	.64
	Control	37	4.03	.73	4	.66
Person Knowledge	Experimental	38	2.78	.98	3.13	.77
	Control	37	2.41	1.02	2.65	.86
Mental Translation	Experimental	38	2.8	.93	3	.90
	Control	37	2.77	1.2	2.59	1.18
Problem-solving	Experimental	38	4.68	.74	5.19	.46
	Control	37	4.54	.84	4.47	.76
Total MALQ	Experimental	38	3.75	.40	4.25	.24
	Control	37	3.54	.36	3.57	.37

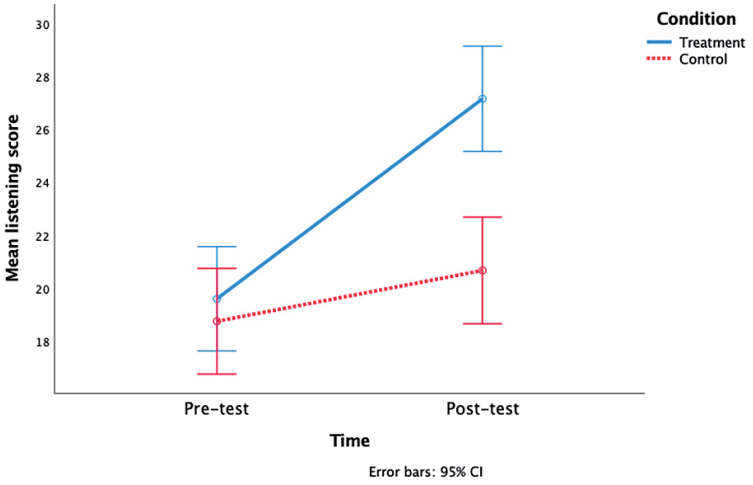
The mean scores for the overall MALQ before (M = 3.75; SD = 0.40) and after (M = 4.25; SD = 0.24) the treatment was delivered in the experimental group indicate an increase in participants' metacognitive awareness. Conversely, this trend was not observed in the control group, where MALQ mean scores before the intervention (M = 3.54; SD = 0.36) and after the intervention (M = 3.57; SD = 0.37) remained relatively consistent. Regarding the components, participants in the experimental group consistently improved their metacognitive awareness in the five components. This contrasts with the data for the control group, as the post-test MALQ means were lower than the pre-test data in directed attention, mental translation, and problem-solving.

4.3. Inferential statistics

*4.3.1. Mixed 2x2 (time*listening condition) ANOVA for TOEIC Bridge listening section data*

In order to assess the impact of the MPC on listening proficiency, a mixed 2x2 (time*listening condition) ANOVA was conducted with the TOEIC Bridge listening section data. Results revealed a significant time*listening condition interaction [F(1, 73) = 51.01, p < .001, η_p^2 = .41]. Planned comparisons indicated that participants made significant gains in both the experimental group [t(37) = -11.705, p < .001, d = 1.9, 95% CI: -8.86 to -6.25] and the control group [t(36) = -4.304, p < .001, d = 0.71, 95% CI: -2.82 to -1.02], but the effect size for the former was very robust. Figure 2 displays this trend.

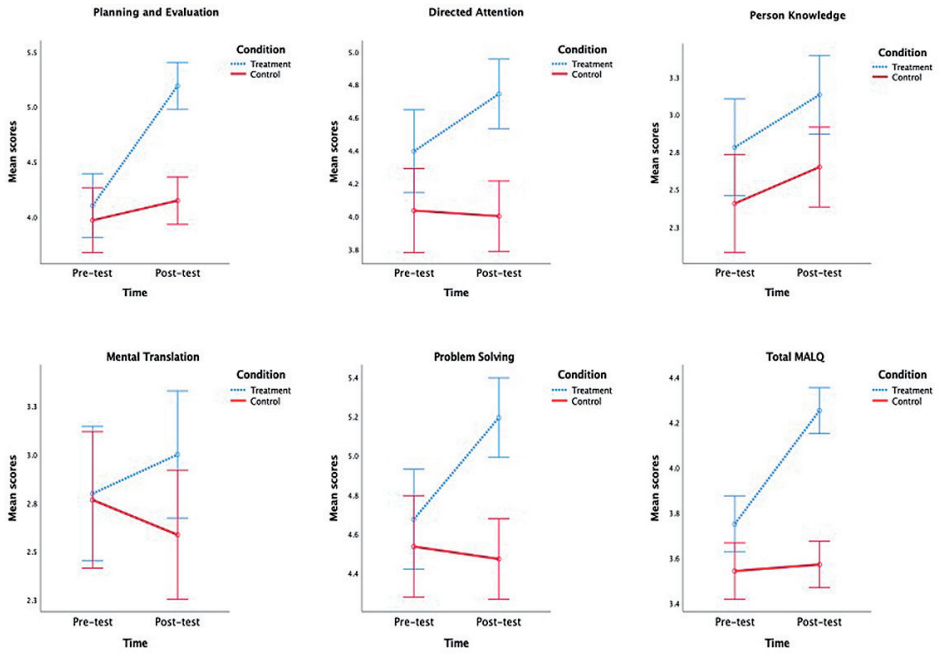
Figure 2: Time*listening condition interaction for TOEIC Bridge listening data



*4.3.2. Mixed 2x2 (time*listening condition) ANOVAs for MALQ data*

The effect of MPC instruction on metacognitive awareness was measured by means of a series of mixed 2x2 (time*listening condition) ANOVAs run with the components and the total scores for the MALQ instrument. For the total MALQ scores, a significant time*listening condition interaction [$F(1, 73) = 60.31$, $p < .001$, $\eta_p^2 = .45$] was found. Similar significant interactions were found for planning and evaluation [$F(1, 73) = 46.48$, $p < .001$, $\eta_p^2 = .39$], directed attention [$F(1, 73) = 11.715$, $p < .005$, $\eta_p^2 = .14$], mental translation [$F(1, 73) = 11.497$, $p < .005$, $\eta_p^2 = .14$], and problem solving [$F(1, 73) = 33.600$, $p < .001$, $\eta_p^2 = .32$]. The only component that did not display significant differences between participants in both conditions was person knowledge [$F(1, 73) = .564$, $p > .05$]. Still, participants in both groups made significant gains in the person knowledge component, so both types of instruction were beneficial for students in this aspect. Thus, with the exception of person knowledge, participants in the treatment group displayed significantly higher listening metacognitive awareness in the post-test than participants in the control group. Figure 3 displays the significant interaction found among the MALQ components.

Figure 3: Time*listening condition interactions for MALQ data (95% CIs)



5. Discussion

5.1. Impact of the MPC on EFL learners' listening outcomes

Results for the mixed ANOVA run with the TOEIC Bridge listening section revealed that the experimental group significantly increased their overall mean scores in a more consistent and robust manner than the control group. The size effect reported suggests that learners who are exposed to MPC instruction make significantly more listening gains than learners who receive regular listening instruction. This is in alignment with previous empirical findings addressing the effect of the MPC on listening performance (Bozorgian & Alamdari, 2017; Cross, 2010; Mahdavi & Miri, 2017; Rahimirad & Shams, 2014; Taherkhani, Aliasin, Khosravi, & Izadpanah, 2022; Vandergrift & Tafaghodtari, 2010). Employing a guided practice approach in the listening process, along with involving learners in the metacognitive processes influencing their learning, facilitated the development of a series of cognitive and metacognitive activities in the listening process, which led to substantial improvements in listening skills. The results also contribute to the evidence suggesting that less skilled learners can benefit from a metacognitive approach to listening instruction (Becker, 2021; Chero, 2023; Vandergrift & Tafaghodtari, 2010), and that listening gains can take place in EFL settings that do not present optimal conditions for learning outside the classroom. The MPC method guided participants in the experimental group to develop substantial implicit metacognitive understanding of L2 listening, which allowed them to create a more explicit view of process-based listening when facing a listening task. As Goh and Vandergrift (2022, p. 137) state, a metacognitive approach to listening "helps learners attend to implicit processes in their listening and make their knowledge of these processes more explicit", and these insights were key to substantially increase listening comprehension gains when compared to the control group. Although learners in the control group made some gains in the TOEIC Bridge post-test (not an unexpected finding, as with any type of learning approach) the product-based goals that were accomplished in this condition focused on teaching learners how to answer test-like comprehension questions (Goh, 2008), which prevented them from making more evident listening gains. Therefore, the results underscore the impact of a metacognitive approach on significantly increasing listening comprehension gains in low-proficiency EFL learners.

5.2. Impact of the MPC on EFL learners' listening metacognitive awareness

The mixed ANOVAs conducted using the data from the MALQ indicated a significant divergence in metacognitive awareness scores between the experimental

and control groups subsequent to the intervention. Participants in the experimental group made significant gains in all five categories of the MALQ. Furthermore, the experimental group significantly outperformed the control group in all the components of the MALQ, with the exception of person knowledge, where both groups increased their metacognitive awareness. Perhaps, after learners in the control group completed the MALQ pre-test, they noticed opportunities to reflect on their listening and understand the factors that influenced their learning (Vandergrift & Tafaghodtari, 2010). This may have allowed them to lower their anxiety and somewhat increase their self-concept. This increase was noticeable but not strong, as low-proficiency language learners tend to develop beliefs about their proficiency based on their performance in listening tasks (Goh & Vandergrift, 2022). It must be noted that learners in both groups displayed the lowest means in the person knowledge component, suggesting that they felt that listening was somewhat difficult for them. The significant awareness increase in the mental translation category for learners in the MPC group is consistent with Maftoon and Alamdari (2016), Robillos (2019), and Tanewong (2018), who reported learners' increased awareness of mental translation strategy avoidance. The MPC approach helped low proficiency learners to become aware of unproductive online mental translation strategies (Vandergrift et al., 2006) that should be avoided as learners develop greater automaticity in word recognition and in the processing of text and meaning. The planning and evaluation component also displayed significant differences between both groups, which is in line with Borzogian's (2014) finding that metacognitive instruction can encourage learners to plan and reflect on possible comprehension issues and envision alternatives to enhance their listening skill. Similar significant differences were reported for problem solving, whose focus on implicit learning through task performance (Vandergrift & Tafaghodtari, 2010) increased the scores for the component in the MPC group. Finally, the experimental group significantly outperformed the control group in the directed attention component, in line with research reporting that metacognitive instruction prompts learners to carefully engage in the process while performing listening tasks (Rahimirad & Shams, 2014). To sum up, the metacognitive pedagogical cycle assessed in this 12-week study yielded results that highlight the effectiveness of the approach in increasing listening comprehension performance. Moreover, learners who were exposed to the MPC approach significantly increased their metacognitive awareness in all components and outperformed the control group in all components but one (personal knowledge). Results also confirm that a metacognitive approach to listening instruction benefits less skilled learners.

6. Conclusion

The current research explored how the metacognitive pedagogical cycle (MPC) impacts adult EFL learners' listening proficiency and metacognitive awareness. Results indicated that an instruction approach such as the MPC can increase listening comprehension gains when compared to regular listening instruction, and that the metacognitive awareness of the learners exposed to the approach significantly increased after the instruction period. These findings emphasize the benefits of implementing metacognitive instruction to increase listening comprehension and metacognition in low proficiency EFL learners. Moreover, the study underscored the relevance of the MPC in developing learners' listening and metacognitive knowledge in EFL contexts that do not enable optimal conditions for learning outside the classroom and emphasize product-based listening exercises. While the study achieved its objectives, the design could have benefited from including a qualitative component to address affective factors such as anxiety, motivation, and self-efficacy, which also influence L2 learners' listening proficiency and enhance the efficacy of metacognitive teaching (Goh & Vandergrift, 2022).

There are pedagogical implications arising from the study. The results underscore a necessity for language teachers in EFL contexts to adopt more strategic approaches to teaching listening and gradually replace product-based instruction with strategic process-based instruction. Listening comprehension must be regarded as an intricate and dynamic process that needs to be understood prior to its integration with phonological components and speaking proficiency. That is, listening comprehension is a skill that is not automatically acquired as the learner engages in communicative activities (Bueno et al., 2006). Furthermore, teachers who test listening rather than teach it and equate "listening effectively" with "effective task completion" do little to help learners develop their listening competence (Graham et al., 2014, p. 53). Teachers need to be open to the exploration of practical approaches, techniques, and activities that include a focus on the process of listening and can be applied in their classrooms in order to help learners develop their listening comprehension more effectively (Graham et al., 2011). Vandergrift and Tafaghodtari (2010) warn that metacognitive instruction can become tedious if it is always carried out in the same way. Including different types of listening texts as part of MPC activities can maintain learners' engagement with the tasks as they complete metacognitive activities.

Teachers can use the MALQ instrument to help learners discover opportunities to reflect on the process of listening and thus raise their metacognitive awareness. If EFL teachers are able to gauge their students' metacognitive awareness, they can use this information to help students incorporate and revise metacognitive strategies and to

give them opportunities to reflect on their metacognitive knowledge. EFL teachers can increase metacognitive awareness through the activities and strategies they implement, which in turn will allow learners to discover and solve listening issues as they engage in listening tasks and develop a more planned approach to effectively extract information from listening input. It must be noted that learners exposed to MPC instruction were not taught listening strategies separately and explicitly in the study; strategy instruction was integrated with regular classroom listening activities (i.e., listening to texts passages from the students' textbooks) through the MPC so that listeners would pursue metacognitive strategies along with their listening tasks. This aspect adds to the practical and pedagogical benefits of including MPC instruction in the EFL classroom. Furthermore, metacognitive instruction encourages teachers to serve as guides that can effectively train learners how to listen and how to enhance their listening skills beyond the listening task itself. As teachers move away from assessing listening performance as a product and focus on listening skill development by means of appropriate metacognitive strategies, learners will become more aware of the complexities in L2 listening, will be able to strategize for self-directed management of their listening development, and will foster their confidence and motivation when undertaking listening tasks.

Acknowledgements:

This research work was funded by *Dirección General de Investigación,* Universidad Andrés Bello, Chile (DGI-UNAB) [Grant number Folio DI-09-20/JM]

7. References

Akama, K. (2006). Relations among self-efficacy, goal setting, and metacognitive experiences in problem-solving. *Psychological Reports*, *98*(3), 895-907. https://doi.org/10.2466/pr0.98.3.895-907

Anderson, N. J. (2012). Metacognition: Awareness of language learning. In S. Mercer, S. Ryan, & M. Williams (Eds.), *Psychology for Language Learning* (pp. 169-187). New York, NJ: Palgrave Macmillan. https://doi.org/10.1057/9781137032829_12

Becker, S. R. (2021). Metacognitive instruction in L2 French: An analysis of listening performance and automaticity. *Foreign Language Annals, 54*(1), 9-26. https://doi.org/10.1111/flan.12506

Bozorgian, H. (2014). The role of metacognition in the development of EFL learners' listening skill. *International Journal of Listening*, *28*(3), 149-161. https://doi.org/10.1080/10904018.2013.861303

Bozorgian, H., & Alamdari, E. F. (2017). Multimedia listening comprehension: Metacognitive instruction or metacognitive instruction through dialogic interaction. *ReCALL*, *30*(1), 131-152. https://doi.org/10.1017/s0958344016000240

Brown, S. (2006). *Teaching listening*. New York: Cambridge University Press.

Bueno, A., Madrid, D., & McLaren, N. (Eds). (2006). *TEFL in secondary education*. Granada: Editorial Universidad de Granada.

Celce-Murcia, M., & Olshtain, E. (2000). *Discourse and context in Language teaching: A guide for language teaching*. Cambridge: Cambridge University Press.

Chero, C. A. C. (2023). The impact of metacognitive instruction on EFL low-level learners' listening performance and metacognitive awareness. *International Journal of Instruction, 16*(2), 291-306. https://doi.org/10.29333/iji.2023.16217a

Cross, J. (2010). Metacognitive instruction for helping less-skilled listeners. *ELT Journal*, *65*(4), 408-416. https://doi.org/10.1093/elt/ccq073

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self-evaluations undermine students' learning and retention. *Learning and Instruction*, *22*(4), 271-280. https://doi.org/10.1016/j.learninstruc.2011.08.003

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906-911. https://doi.org/10.1037/0003-066x.34.10.906

Glogger, I., Schwonke, R., Holzäpfel, L., Nückles, M., & Renkl, A. (2012). Learning strategies assessed by journal writing: Prediction of learning outcomes by quantity, quality, and combinations of learning strategies. *Journal of Educational Psychology*, *104*(2), 452-468. https://doi.org/10.1037/a0026683

Goh, C. C. M. (1997). Metacognitive awareness and second language listeners. *ELT Journal, 51*(4), 361-369. https://doi.org/10.1093/elt/51.4.361

Goh, C. C. M. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, *28*(1), 55-75. https://doi.org/10.1016/s0346-251x(99)00060-3

Goh, C. C. M. (2008). Metacognitive instruction for second language listening development. *RELC Journal, 39*(2), 188-213. https://doi.org/10.1177/0033688208092184

Goh, C. C. M. (2010). Listening as process: Learning activities for self-appraisal and self-regulation. In N. Harwood (Ed.), *English language teaching materials: Theory and practice* (pp. 179-206). New York: Cambridge University Press.

Goh, C. C. M., & Hu, G. (2013). Exploring the relationship between metacognitive awareness and listening performance with questionnaire data. *Language Awareness, 23*(3), 255-274. https://doi.org/10.1080/09658416.2013.769558

Goh, C. C. M., & Vandergrift, L. (2022). *Teaching and learning second language listening: Metacognition in action.* New York: Routledge.

Graham, S. (2006). Listening comprehension: The learners' perspective. *System, 34*(2), 165-182. https://doi.org/10.1016/j.system.2005.11.001

Graham, S., Santos, D., & Francis-Brophy, E. (2014). Teacher beliefs about listening in a foreign language. *Teaching and Teacher Education*, *40*, 44-60. https://doi.org/10.1016/j.tate.2014.01.007

Graham, S., Santos, D., & Vanderplank, R. (2011). Exploring the relationship between listening development and strategy use. *Language Teaching Research*, *15*(4), 435-456. https://doi.org/10.1177/1362168811412026

Gu, Y. (2018). Two-way listening. In *The TESOL Encyclopedia of English Language Teaching* (pp. 1-8). Minneapolis:Wiley. https://doi.org/10.1002/9781118784235.eelt0578

Hasan, A. N. (2000). Learners' perceptions of listening comprehension problems. *Language Culture and Curriculum*, *13*(2), 137-153. https://doi.org/10.1080/07908310008666595

Kobayashi, A. (2018). Investigating the effects of metacognitive instruction in listening for EFL learners. *The Journal of Asia TEFL*, *15*(2), 310-328. https://doi.org/10.18823/asiatefl.2018.15.2.4.310

Latham-Koening & Oxenden, C. (2018). *American English File Student Book* (2nd ed.). New York: Oxford University Press.

Lynch, T. (2006). Academic listening: Marrying top and bottom. In E. Usó-Juan, & A. Martínez-Flor (Eds.), *Current trends in the development and teaching of the four language skills* (pp. 121-140). Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110197778.2.91

Lynch, T., & Mendelsohn, D. (2020). Listening. In N. Schmitt, & M. P. H. Rodgers (Eds.), *An introduction to Applied Linguistics* (3rd ed., pp. 223-239). New York, NJ: Routledge.

Maftoon, P., & Alamdari, E. F. (2016). Exploring the effect of metacognitive strategy instruction on metacognitive awareness and listening performance through a process-based approach. *International Journal of Listening*, *34*(1), 1-20. https://doi.org/10.1080/10904018.2016.1250632

Mahdavi, N., & Miri, M. (2017). Co-shaping metacognitive awareness and developing listening comprehension through process-based instruction. *International Journal of Listening*, *33*(1), 53-70. https://doi.org/10.1080/10904018.2016.1260454

Matsumoto, Y. (2011). Successful ELF communications and implications for ELT: Sequential analysis of ELF pronunciation negotiation strategies. *The Modern Language Journal*, *95*(1), 97-114. https://doi.org/10.1111/j.1540-4781.2011.01172.x

Nation, I. S. P., & Newton, J. (2020). *Teaching ESL/EFL listening and speaking* (2nd ed.). New York: Routledge. https://doi.org/10.4324/9780429203114

Nazari, M. (2018). The impact of teacher education on L2 teachers' cognitions and pedagogy of metacognitive listening instruction. *International Journal of Listening, 34*(1), 34-53. https://doi.org/10.1080/10904018.2018.1461565

Oyarzún, N. (2019). *The effects of a metacognitive, process-based approach in students listening comprehension performance and metacognitive awareness* [Unpublished master's thesis, Universidad Andrés Bello, Chile].

Rahimirad, M., & Shams, M. H. (2014). The effect of activating metacognitive strategies on the listening performance and metacognitive awareness of EFL students. *International Journal of Listening, 28*(3), 162-176. https://doi.org/10.1080/10904018.2014.902315

Rhodes, M. G. (2019). Metacognition. *Teaching of Psychology, 46*(2), 168-175. https://doi.org/10.1177/0098628319834381

Richards, J. C. (2006). *Communicative language teaching today*. New York, NJ: Cambridge University Press.

Robillos, R. (2019). Crossing metacognitive strategy instruction in an EFL classroom: Its impact to Thai learners' listening comprehension skill and metacognitive awareness. *Asian EFL Journal, 21*, 311-336.

Roelle, J., Nowitzki, C., & Berthold, K. (2017). Do cognitive and metacognitive processes set the stage for each other? *Learning and Instruction, 50*, 54-64. https://doi.org/10.1016/j.learninstruc.2016.11.009

Rost, M. (2016). *Teaching and researching listening* (3rd ed.). New York, NJ: Taylor & Francis.

Roussel, S., Gruson, B., & Galan, J.-P. (2017). What types of training improve learners' performances in second language listening comprehension? *International Journal of Listening, 33*(1), 39-52. https://doi.org/10.1080/10904018.2017.1331133

Shintani, N. & Wallace, P. M. (2014). Effects of listening support in second language classroom: A meta-analysis. *English Teaching and Learning, 38*(3), 71-101.

Siegel, J. (2013). Exploring L2 listening instruction: examinations of practice. *ELT Journal, 68*(1), 22-30. https://doi.org/10.1093/elt/cct058

Taghizade, M., Bozorgian, H., & Abadikhah, S. (2022). Introvert EFL learners' listening comprehension and metacognitive awareness: A mixed-method study. *Language Teaching Research*, 136216882211431. https://doi.org/10.1177/13621688221143105

Taherkhani, B., Aliasin, S. H., Khosravi, R., & Izadpanah, S. (2022). The interface between metacognitive strategy training and locus of control in developing EFL learners' listening comprehension skill. *Frontiers in Education*, 7, 847564. https://doi.org/10.3389/feduc.2022.847564

Tanewong, S. (2018). Metacognitive pedagogical sequence for less-proficient Thai EFL listeners: A comparative investigation. *RELC Journal, 50*(1), 86-103. https://doi.org/10.1177/0033688218754942

Vandergrift, L. (2003). From prediction through reflection: Guiding students through the Process of L2 Listening. *The Canadian Modern Language Review, 59*(3), 425-440. https://doi.org/10.3138/cmlr.59.3.425

Vandergrift, L. (2004). 1. Listening to learn or learning to listen? *Annual Review of Applied Linguistics, 24*, 3-25. https://doi.org/10.1017/S0267190504000017

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching, 40*(3), 191-210. https://doi.org/10.1017/s0261444807004338

Vandergrift, L. (2015). Listening. In P. Robinson (Ed.), *The Routledge encyclopedia of second language learning* (pp. 400-404). New York: Routledge.

Vandergrift, L., & Tafaghodtari, M. H. (2010). Teaching L2 learners how to listen does make a difference: An empirical study. *Language Learning, 60*(2), 470-497. https://doi.org/10.1111/j.1467-9922.2009.00559.x

Vandergrift, L., Goh, C. C. M., Mareschal, C. J., & Tafaghodtari, M. H. (2006). The metacognitive awareness listening questionnaire: Development and validation. *Language Learning, 56*(3), 431-462. https://doi.org/10.1111/j.1467-9922.2006.00373.x

Wang, W. (2015). Learning to listen: The impact of a metacognitive approach to listening instruction. *The Asia-Pacific Education Researcher, 25*(1), 79-88. https://doi.org/10.1007/s40299-015-0235-4

Wenden, A. (1998). Metacognitive knowledge and language learning. *Applied Linguistics, 19*(4), 515-537. https://doi.org/10.1093/applin/19.4.515

Worthington, D. L., & Fitch-Hauser, M. E. (2018). *Listening: Processes, functions, and competency.* New York, NJ: Routledge. Retrieved from https://bitly.cx/IYNMG

Appendix A. Translated items in the Metacognitive Awareness Listening Questionnaire (MALQ; Vandergrift et al., 2006; Oyarzún, 2019)

1. *Antes de empezar a escuchar un audio, me hago una idea de cómo voy a escucharlo.*
(Before I start to listen, I have a plan in my head for how I am going to listen.)
2. *Me concentro más en el audio cuando tengo problemas al entenderlo.*
(I focus harder on the text when I have trouble understanding.)
3. *Considero que la comprensión auditiva en inglés es más difícil que leer, hablar o escribir en inglés.*

(I find that listening is more difficult than reading, speaking, or writing in English.)
4. *Traduzco mentalmente a medida que voy escuchando.*
(I translate in my head as I listen.)
5. *Utilizo las palabras que entiendo para adivinar el significado de las palabras que no comprendo.*
(I use the words I understand to guess the meaning of the words I don't understand.)
6. *Cuando pierdo la concentración, la recupero de inmediato.*
(When my mind wanders, I recover my concentration right away.)
7. *Mientras escucho, comparo lo que entiendo del audio con lo que sé sobre el tema.*
(As I listen, I compare what I understand with what I know about the topic.)
8. *Siento que la comprensión auditiva en inglés es un desafío para mí.*
(I feel that listening comprehension in English is a challenge for me.)
9. *Uso mi experiencia y conocimiento para ayudarme a entender.*
(I use my experience and knowledge to help me understand.)
10. *Antes de empezar a escuchar, pienso en audios similares que podría haber escuchado antes.*
(Before listening, I think of similar texts that I may have listened to.)
11. *Traduzco palabras claves mientras escucho.*
(I translate key words as I listen.)
12. *Trato de retomar la tarea cuando pierdo la concentración.*
(I try to get back on track when I lose concentration.)
13. *Mientras escucho, rápidamente ajusto mi interpretación del audio si me doy cuenta de que no está correcta.*
(As I listen, I quickly adjust my interpretation if I realize that it is not correct.)
14. *Después de escuchar, hago memoria de como escuché y pienso en qué podría hacer de manera diferente la próxima vez.*
(After listening, I think back to how I listened, and about what I might do differently next time.)
15. *No me pongo nervioso cuando escucho un audio en inglés.*
(I don't feel nervous when I listen to English.)
16. *Cuando me es difícil entender lo que escucho, me rindo y dejo de escuchar.*
(When I have difficulty understanding what I hear, I give up and stop listening.)
17. *Uso la idea general del texto para que me ayude a deducir el significado de las palabras que no entiendo.*
(I use the general idea of the text to help me guess the meaning of the words that I don't understand.)
18. *Mientras escucho, traduzco palabra por palabra.*
(I translate word by word, as I listen.)
19. *Cuando deduzco el significado de una palabra, hago memoria de todo lo que he escuchado anteriormente para ver si mi interpretación tiene sentido.*
(When I guess the meaning of a word, I think back to everything else that I have heard, to see if my guess makes sense.)
20. *Mientras escucho, periódicamente me pregunto si estoy satisfecho/a con mi nivel de comprensión.*
(As I listen, I periodically ask myself if I am satisfied with my level of comprehension.)
21. *Tengo un objetivo en mente mientras escucho.*
(I have a goal in mind as I listen.)

Appendix B. Sample listening activity lesson plan in MPC (experimental) group: CouchSurf around the world! (Oyarzún, 2019)

MPC Stage	Activities	Description	Duration
Planning/ predicting	Pre-listening activities.	• The teacher hands students a worksheet. • The teacher asks the students what they know about "couch-surfing". • The teacher asks students to look at different pictures in the worksheet and asks them to write different words, phrases or sentences they think they are related to couch surfing in the "Your predictions" chart. • The teacher asks different students about the words and phrases they thought of and writes them on the board. • The teacher tells students that they will listen to a person talking about what couch-surfing is.	Estimated time: 12-15 minutes
First verification and plan with peers for second listen	First listen	• The teacher tells students that they will listen to the audio for the first time and while they listen, they will have to check their ideas/predictions, checking off the information they predicted correctly. Students write down new information they may have understood from the text on the "First listen" column of the worksheet. • After the first listen, students work in pairs and compare their predictions and extra information understood. They discuss confusing points and disagreements to consider other logical options as well as identify points in the text that will require further attention during the second listening.	Estimated time: 10 -12 minutes
Second verification stage and text reconstruction or other comprehension activity	Second listen	• The teacher asks students to listen to the text for a second time. They try to resolve points of confusion raised during the first listen and complete the "Second listen" section of the worksheet with the new information that was understood. • After students finish taking their notes, the teacher engages all the participants in a group discussion to confirm their comprehension of the text and to enable them to share how they succeeded in understanding words and ideas.	Estimated time: 8-10 minutes
Final verification	Third listen	• Students verify points of earlier disagreement and make changes/corrections if necessary, in the 'Third listen' column of the worksheet.	Estimated time: 5-6 minutes

MPC Stage	Activities	Description	Duration
Reflection and goal setting	Post-listening activities	• The teacher hands students the transcript of the text. • The teacher asks students to individually self-assess their comprehension level by completing a chart in their worksheet writing the ideas they understood correctly, ideas they understood wrongly and ideas they did not understand at all. • The teacher checks students' answers with the whole group and asks different students about the ideas they wrote in the different columns. • The teacher leads a class discussion on the content of the text with the whole group, and the class reconstructs the text (this can be done either in English or Spanish since it is only a comprehension verification stage). • The teacher asks students to complete the "Reflections" section of the worksheet either in English or Spanish. • The teacher encourages students to engage in a class discussion about the difficulties they encountered and the strategies they used, as well as how to approach the next lesson (in either English or Spanish).	Estimated time: 15-20 minutes

[1] The present study's methodology was based on a MA thesis carried out by the co-author (Oyarzún, N. (2019). *The effects of a metacognitive, process-based approach in students' listening comprehension performance and metacognitive awareness* (Unpublished master's thesis). Universidad Andres Bello, Chile). The present study has included some aspects of the literature review that were discussed in the MA thesis, as well as the materials in the appendix. All aspects of data collection procedures, analysis, and discussion were conducted with a new sample of participants.

Foreign language writing enjoyment: Conceptualization, sources, and measurement

Xianwen Chen
School of Foreign Languages
Sun Yat-sen University, China;
Centre for Linguistics and Applied Linguistics
Guangdong University of Foreign Studies, China
chenxw36@mail.sysu.edu.cn

Hai Xu
Centre for Linguistics and Applied Linguistics
Guangdong University of Foreign Studies, China
xuhai1101@gdufs.edu.cn

Xian Zhang
Department of Linguistics
University of North Texas, USA
xian.zhang@unt.edu

Abstract

Due to the paucity of research on foreign language writing enjoyment (FLWE) among mixed-major EFL undergraduates, this article explores the multi-dimensional conceptualization, causes, and measurement of FLWE within this student group. The two-phase investigation involved over 3,000 participants. Phase 1 utilized interviews and questionnaires to gather students' perceptions of FLWE and its sources. The findings illuminate FLWE as a multidimensional construct, with the writing topic emerging as the primary source of enjoyment, followed by text type and feedback. Phase 2 focused on developing and validating a 17-item Foreign Language Wring Enjoyment Scale (FLWES), which demonstrates sound reliability and validity. Unlike existing measures that mainly targeted limited dimensions of enjoyment among junior secondary students who are exposed to a limited range of text types primarily for examination purposes, or English majors, our instrument delves deeper into FLWE experienced by EFL undergraduates across disciplines. Our research contributes to a profound understanding of FLWE and opens avenues for measuring enjoyment in various language skills and cultural contexts.

Keywords: foreign language writing enjoyment; control-value theory; FLWES; conceptualization; validation.

Resumen

Debido a la escasez de investigaciones sobre el disfrute de la escritura en idiomas extranjeros (DEIE) entre los estudiantes universitarios de EFL de diferentes especialidades, este artículo explora la conceptualización multidimensional, las causas y la medición del DEIE dentro de este grupo estudiantil. La investigación, realizada en dos fases, involucró a más de 3,000 participantes. En la primera fase, se emplearon entrevistas y cuestionarios para recopilar las percepciones de los estudiantes sobre el DEIE y sus fuentes. Los hallazgos revelaron al DEIE como un constructo multidimensional, con el tema de escritura emergiendo como la principal fuente de disfrute, seguido por el tipo de texto y la retroalimentación. La segunda fase se enfocó en desarrollar y validar una Escala de Disfrute de la Escritura en Idiomas Extranjeros (DEIE) de 17 items, que demostró una fiabilidad y validez sólidas. A diferencia de las medidas existentes, nuestro instrumento profundiza en el DEIE experimentado por estudiantes de pregrado de EFL en diferentes disciplinas, contribuyendo así a una comprensión más profunda del DEIE y abriendo caminos para medir el disfrute en diversas habilidades lingüísticas y contextos culturales.

Palabras clave: escritura para disfrutar de una lengua extranjera; teoría del valor de control; EDELE; conceptualización; validación.

1. Introduction

Over the past decade, the rise of Positive Psychology in SLA has prompted researchers to shift away from the predominant focus on negative emotions, mostly foreign language anxiety (Cheng, 2004, 2017; Dewaele et al., 2008; Horwitz, 2010; Horwitz et al., 1986), to positive emotions, particularly enjoyment (Dewaele & Li, 2020). During this time, multiple measurement tools have been developed to gauge overall foreign language enjoyment (e.g., Botes et al., 2021; Dewaele & MacIntyre, 2014, 2017; Jin & Zhang, 2018; Li et al., 2018). However, some researchers have also started to attend to language-skill-specific enjoyment, such as foreign language writing enjoyment. For example, Tahmouresi and Papi (2021), were trailblazers in discovering the positive correlation between foreign language enjoyment and writing motivation. Notably, enjoyment in this research was gauged using an adapted form of Teimouri's (2017) joy scale, with further investigation needed to validate its factor structure and construct. Li et al. (2023) also reported that writing enjoyment positively influenced L2 writing achievement.

Regarding the assessment of foreign language writing enjoyment (FLWE), Jin (2023) crafted the English Writing Enjoyment Scale with reference to Scherer's (2000) component process model of emotion. Li et al. (2023) constructed the Foreign

Language Writing Enjoyment Scale, which specifically comprises two factors: Private and Teacher. These instruments, however, have limitations when assessing writing enjoyment in various language learning settings. Firstly, both scales do not fully capture the spectrum of writing enjoyment experiences among EFL undergraduates. Jin's (2023) scale exclusively addresses the multifaceted nature of enjoyment without exploring its underlying experiences and origins. Li et al. (2023) established a dual-factor structure. Yet, it boils down to two factors: Private and Teacher-Related, which may inadequately capture the entirety of foreign language writing enjoyment among diverse-major college students. Secondly, Li et al.'s (2023) factor structure was based on Dewaele and MacIntyre's (2014) 21-item foreign language enjoyment scale, which was crafted relying on expert knowledge rather than undergoing psychometric validation. Specifically, beyond the 7 items obtained from Ryan et al.'s (1990) Interest/Enjoyment subscale, the remaining items were grounded in scholarly expertise. Additionally, both tools lack widespread applicability among EFL undergraduates from various disciplines. Jin's (2023) scale was developed using data exclusively from undergraduate English majors at two specific universities. Li et al.'s (2023) is geared towards junior secondary students. Furthermore, prior studies did not provide a conceptualization of foreign language writing enjoyment (FLWE) and report its major causes.

To address these gaps, we present a research project that investigates the conceptualization of FLWE, its major triggers, and its assessment, using data from more than 3,000 undergraduate EFL learners.

2. Literature review

2.1. Enjoyment: Definition, dimensionality, antecedents, and dichotomy

To define "enjoyment", Csikszentmihalyi (2008, 2014) distinguished between pleasure and enjoyment. Pleasure refers to a state of satisfaction arising from fulfilling one's basic physiological drives (e.g., bodily ease). Enjoyment, on the other hand, can be conceptualized as the emotion overriding a human's homeostatic needs. In essence, enjoyment is characterized by "a sense of novelty and accomplishment" (Csikszentmihalyi, 2008, p. 46). It is what leads to personal growth and long-term happiness (MacIntyre & Dewaele, 2019).

Within Pekrun's (2006) control-value theory of achievement emotions framework, enjoyment can be approached from three perspectives: (a) valence: positive vs. negative; (b) activation: high-arousal vs. low-arousal; (c) objective focus: process-oriented vs. results-driven. Accordingly, enjoyment can be described as a positive, high-arousal emotion experienced during the process of activities or tasks (Pekrun & Perry, 2014).

The control-value theory contends that two proximal antecedents (i.e., control and value appraisals) can directly induce achievement emotions. Control appraisals pertain to an individual's controllability over their actions and outcomes whereas value appraisals include intrinsic (e.g., inherent interest in English) and extrinsic value (e.g., utilitarian functions of English). It follows that enjoyment could arise when individuals have positive control and value appraisals of an ongoing activity (Pekrun, 2006).

Furthermore, distal antecedents (i.e., situation-specific emotions, goals, beliefs, cognitive abilities, and gender) and situational antecedents (i.e., task quality, motivating factors, autonomous tasks/settings, and positive feedback) indirectly influence achievement emotions through the proximal appraisals (Pekrun, 2006). Regarding foreign language enjoyment, some situational antecedents (e.g., positive classroom atmosphere, teacher support, positive feedback, autonomous and creative tasks) were also found to contribute to this emotion (e.g., Botes et al., 2021; Dewaele & MacIntyre, 2014, 2017; Li et al., 2018).

In line with the trait-state dichotomy of emotion (Pekrun, 2006), FLE can be dichotomized into trait and state types. Trait enjoyment refers to learners' habitual positive responses to general learning situations (e.g., school). Notably, it can also be experienced recursively in a specific situation (e.g., habitual enjoyment felt in English learning situations), which is known as situation-specific enjoyment (e.g., foreign language writing enjoyment). State enjoyment, however, describes enjoyment experienced momentarily in reaction to a specific situation at a specific time (Frenzel, Goetz, & Pekrun, 2009). What differentiates situation-specific from state enjoyment is temporal scope but situational context. While a distinction was made between situation-specific and state FLE in general, whether this classification method is applicable to writing-specific enjoyment remains uncertain.

2.2. Assessment of foreign language writing enjoyment

The field of SLA has witnessed increased scholarly interest in enjoyment with the advent of positive psychology (MacIntyre, 2016; MacIntyre & Mercer, 2014). This interest has resulted in the development of various instruments for assessing foreign language enjoyment (e.g., Botes et al., 2021; Dewaele & MacIntyre, 2014, 2017; Li et al., 2018). Nonetheless, some researchers have begun to pay attention to enjoyment specific to language skills, shifting from general classroom enjoyment. For example, studies have found a positive correlation between foreign language writing enjoyment, learners' motivation and L2 writing performance (Li et al., 2023; Tahmouresi & Papi, 2021). A few studies have also examined the measurement of foreign language writing enjoyment. Drawing upon Scherer's (2000) component process model of emotion, Jin (2023) devised the English Writing Enjoyment Scale. Concurrently, Li et al. (2023)

created the Foreign Language Writing Enjoyment Scale, comprising two distinct components: Social and Teacher-Related Enjoyment. Nevertheless, these measurement tools may not completely unveil the experiences and sources of writing enjoyment for EFL undergraduates with diverse majors. Specifically, Jin's (2023) scale comprises four dimensions concerning the inherent attributes of enjoyment: cognition, motivation, affect, and expression, while overlooking its fundamental sources. Conversely, Li et al.'s (2023) scale, while featuring two factors (Social and Private) borrowed from Dewaele and MacIntyre's (2014, 2016) Foreign Language Enjoyment Scale, ultimately simplifies to two dimensions: Private and Teacher-Related. Nevertheless, these two aspects alone may not offer a comprehensive portrayal of the sources of foreign language writing enjoyment encountered by EFL undergraduates with interdisciplinary majors. In addition, Li et al.'s (2023) instrument drew inspiration from Dewaele and MacIntyre's (2014) tool, which was crafted relying on expert insights, foregoing the process of psychometric validation. Specifically, 14 out of 21 items drew extensively from scholarly knowledge, excluding the 7 items derived from Ryan et al.'s (1990) Interest/Enjoyment subscale. Furthermore, both tools lack widespread utility among the multidisciplinary EFL undergraduates. In particular, Jin (2023) constructed the scale solely based on data gathered from English majors recruited from two universities. According to Putwain et al. (2018), academic emotions are subject to domain specificity, indicating that they should be tailored to specific subjects. Apparently, English majors may possess unique academic self-perceptions and values related to English, differing from non-English majors. Li et al.'s (2023) is intended for adolescents in secondary school. Compared with college students with varied text exposure and rich writing backgrounds, junior secondary learners are confined to specific text types (i.e., narration, description, and letter writing) for exam purposes. This may impede researchers from exploring the intricacies of writing enjoyment. In addition, previous studies did not present a multi-facet conceptualization of foreign language writing enjoyment (FLWE) or document its main causes. To bridge these gaps, we initiated conceptualizing FLWE from a multi-dimensional perspective, investigating its primary catalysts, and measuring it through data collected from over 3,000 college EFL students.

Specifically, this study aimed to address the following research questions:

1. Do Chinese EFL undergraduate learners experience enjoyment specific to English writing? If yes, how do learners conceptualize foreign language writing enjoyment (FLWE) in EFL settings?
2. What are the sources of FLWE?
3. What is the factor structure of FLWES?
4. Is the Foreign Language Writing Enjoyment Scale (FLWES) reliable and valid to measure FL writing enjoyment?

3. Method

A two-phase study was conducted. As illustrated in Table 1, Phase 1 applied a qualitative and descriptive method to identify the occurrence and sources of FLWE and to conceptualize the construct. Phase 2 connected the qualitative results from Phase 1 with quantitative methods to develop and validate the FLWES, following Devellis and Thorpe's (2021) multiple stages of item pool generation (scrutinized and refined through expert review, pilot testing, and item analysis), scale development, and scale validation.

Table 1: Research roadmap

Study	Steps	Participants	Instrument	Data analysis	Tool
Phase 1 Occurrence, conceptualization, and sources of FLWE	Qualitative stage	n_1 = 30	Interview	Qualitative analysis	Nvivo 12 Plus
		N_1 = 470	Open questionnaire		
Phase 2 Development and validation of the FLWES	Quantitative stages Initial item pool	Qualitative findings concerning sources of FLWE in Phase 1 lay the groundwork for item pool generation in Phase 2.			
	Expert review	N_2 = 6	60-item FLWES	Face validity	Manual
	Pilot testing	N_3 = 128	49-item FLWES	Face validity	Manual
	Scale development	N_4 = 1015	item FLWES	EFA	SPSS 25.0

Study	Steps	Participants	Instrument	Data analysis	Tool
	Scale validation	N_5 = 1670	Item FLWES	CFA Construct, convergent, discriminant, and criterion validity	Amos 23.0 SPSS 25.0
		n = 113	17-item FLWES	Test-retest reliability	SPSS 25.0

Note. FLWES = Foreign Language Writing Enjoyment Scale; EFA = Exploratory Factor Analysis; CFA = Confirmatory Factor Analysis.

4. Phase 1: Occurrence, conceptualization, and sources of FLWE

Phase 1 had three objectives: (a) to identify the occurrence of FLWE; (b) to provide a multidimensional conceptualization of FLWE; and (c) to provide sources for generating the item pool of the FLWES in Phase 2.

4.1. Participants

Only English majors were involved in Phase 1 because English writing is a major component of their undergraduate curriculum in China. Thus, they have extensive English writing experience, making them ideal for conceptualizing and assessing FLWE.

A total of 30 undergraduate students (17 females and 13 males; average age = 19.6, SD = 1.14, range: 18-21) from six universities of differential academic rankings participated in the interview.

Another 470 undergraduate English majors aged between 18 and 22 from 18 universities filled out an open-ended questionnaire. Among them, 26% (*n* = 123) were males and 74% (*n* = 347) were females.

4.2. Instruments

Interviews with students were conducted via Tencent Meeting, a popular online video conferencing platform. The questionnaires were uploaded onto Wenjuanxin (https://www.wjx.cn/), an online questionnaire platform. The instructions and responses were given in Chinese to guarantee full comprehension. The following questions were used in both interviews and questionnaires to solicit information concerning learners' perceptions of FLWE:

a. Have you ever enjoyed English writing?
b. If yes, when?
c. And why?
d. Can you narrate a particular episode in English writing that you found enjoyable?
e. Can you provide a detailed description of the feeling of enjoyment?

4.3. Procedure

Prior to participation, all subjects signed a consent form that outlined research objectives, procedures, the anticipated length of participation, data protection issues, and the freedom to drop out of the study without consequences. Each interview lasted 30-60 minutes. The 470 students were asked to scan a QR code generated by Wenjuanxing. Then, they filled out the surveys and submitted their answers during a class session.

4.4. Data analysis

4.4.1. The existence and conceptualization of FLWE

Participants' interview and questionnaire responses were coded via NVivo 12 Plus, a professional qualitative analysis tool.

To conceptualize FLWE, students' descriptions of the feeling of foreign language writing enjoyment were coded. Initially, "in-vivo" codes (i.e., respondents' verbatim responses) were used. Twenty initial codes emerged through iterative refinement, confirmation, and categorization. For instance, excerpts such as "I have an aspiration to recite more good sentences" and "I have an aspiration to recite more input materials relevant to the writing topic" were first categorized as "have an inspiration for self-improvement", and then grouped under an overarching category "value". Afterwards, each code was reviewed and data excerpts assigned to the same node were compared. Pekrun's (2006) three-dimensional taxonomy and antecedents of achievement emotions lay the foundation for a coding scheme classifying learners' perceptions of FLWE.

Using the scheme, the interview transcripts were then coded by following the same procedures. To ensure consistency, iterative comparisons were made between codes and categories assigned to the questionnaire responses and interview transcripts. Revisions were made if any mismatches were detected. The finalized coding scheme with the frequencies of categories is shown in Table 2.

4.4.2. Sources of FLWE

After determining the occurrence of FLWE, its sources were investigated by further analyzing participants' enjoyment excerpts. Similarly, initial coding involved abstracting in-vivo codes such as "interesting topic" and "a text type relating to my personal experiences". Herein, "text type" refers to the traditional rhetorical modes of discourse – narration, description, exposition, argumentation, and letters (Biber, 1994).

Further iterations of participants' data resulted in 60 sub-theme codes, which were fine-tuned over multiple cycles of reading. Then, the codes were collapsed into broader categories. For example, "I enjoy writing when working on an interesting topic" and "Working on a topic that relates to my real-life experience is enjoyable" were two sub-categories, which were under the broader category "Topic-Related Writing Enjoyment". These nodes and sub-nodes were then organized into a hierarchy and the coding scheme was applied to each textual response. The interview transcripts were also analyzed with the same coding scheme. The finalized coding scheme consisted of seven categories and 60 sub-categories.

To guarantee trustworthy results of Phase 1, two measures were taken. Firstly, data was collected and coded in Chinese to retain the participants' original meanings. Secondly, the data was independently rated by the researcher and a trained coder, a native Chinese speaker, and a college English teacher with a doctoral degree in applied linguistics. Inter-rater reliability was high (81%). Discrepancies were resolved, leading to a final agreement rate of 92%.

4.5. Results

4.5.1. Occurrence of FLWE

In total, 30 students recalled 74 enjoyment-eliciting instances, with an average of 2.47 scenarios per participant. Additionally, the open questionnaire indicated that 93.6% of the students (440 out of 470) reported encountering enjoyment episodes.

4.5.2. Conceptualization of FLWE

In answer to the question "Can you describe the feeling of enjoyment in relation to English writing in detail?", students provided responses (see Table 2) conforming to the three dimensions (i.e., valence, activation, and object focus) and two proximal antecedents (i.e., perceived control, and value) of achievement emotions within the control-value theory framework.

Table 2: Codes and sub-codes of FLWE and their frequencies of occurrence

Codes	Sub-codes (N)				
Valence	feel joy/ excited (191)	release one's trapped emotion (2)	heal emotional pain (2)	release negative emotions and let go of grief (2)	
Activation	smile self-consciously (80)	feel a heart palpitation (43)	one's hand shakes when getting excited about writing (9)	one's hand gets clammy when getting excited about writing (3)	
Object focus	want to continue writing (cannot stop writing) (107)	completely absorbed in writing (unaware of the passage of time) (104)			
Controllability	have confidence to produce a high-quality article (24)	have a positive attitude towards the writing task (22)	feel satisfied with the quality of one's writing (26)	feel the words just flow out of the pen (21)	
Value	go through many revisions (to improve the draft) (31)	cannot resist writing it up (21)	have an aspiration for self-improvement (e.g., recite more good sentences) (18)	desire to share one's writing with others (17)	pay attention to one's handwriting (1)

Note. N = frequency in the reports

As Table 2 indicates, FLWE is a positive feeling characterized by high levels of arousal during ongoing activities (e.g., "completely absorbed in writing", "unaware of the passage of time", and "cannot stop writing"). Specifically, (a) FLWE entails positive feelings instigated by certain events or situations (e.g., "feel joy/excited", "release one's trapped emotion", and "healing emotional pain"); (b) FLWE involves high levels of

emotional arousal (e.g., "feel a heart palpitation", and "one's hand shakes/gets clammy when getting excited about writing"); (c) FLWE occurs in the midst of activities, as reflected by a high level of engagement in writing; and (d) FLWE is instigated by control and value appraisals. Specifically, controllability over behaviors and outcomes can be exemplified by learners' positive attitude towards the writing task, their confidence to produce a high-quality article, and their contentment with writing outcomes. Concerning value appraisal, participants perceived writing as a means of expressing their thoughts. Furthermore, writing was highly valued for its positive intrinsic value, as proven by "aspired for self-improvement", "revised the article multiple times to improve its quality", and "paid attention to handwriting".

In line with Pekrun's (2006) dichotomy of state and situation-specific emotions, two types of FLWE emerged in students' questionnaires and interview responses, namely the state FLWE and the situation-specific FLWE. The former is a transient emotional state in relation to foreign language writing, which surfaces at a specific time point. The latter refers to a type of enjoyment unique to foreign language writing, which stems from repeated state experiences of enjoyment. These two types of FLWE can be exemplified by the following episodes.

State FLWE:

EXTRACT 1

> S1 (male, aged 19). While brainstorming for my memoir, I was suddenly filled with a rush of inspiration. The words flowed effortlessly from my mind to the page, providing me with an unparalleled sense of enjoyment. (Student interview)

EXTRACT 2

> S2 (female, aged 21) The teacher assigned an essay on important life decisions. Facing writer's block, I decided to craft a story. It depicts a woman's transition from recklessness to diligence in youth. Initially, people around her didn't think girls needed to strive until she witnessed a classmate's early married haggard appearance. Inspired by Principal Zhang Guimei's tale and my own experiences, writing it felt deeply resonant.

Situation-specific FLWE:

EXTRACT 3

> S3 (female, aged 20). I remember once I took part in an English writing contest and I wrote about a conversation with my grandpa, who passed away years ago.

When I was confused about life, he appeared in my dream, telling me why he joined the army and what the meaning of life was. This writing experience revived the memory of my dear grandpa. Memorizing his life and honoring him in ways that reflected on the person he was when alive and what he meant to me helped me heal from pain and sadness. (Student interview)

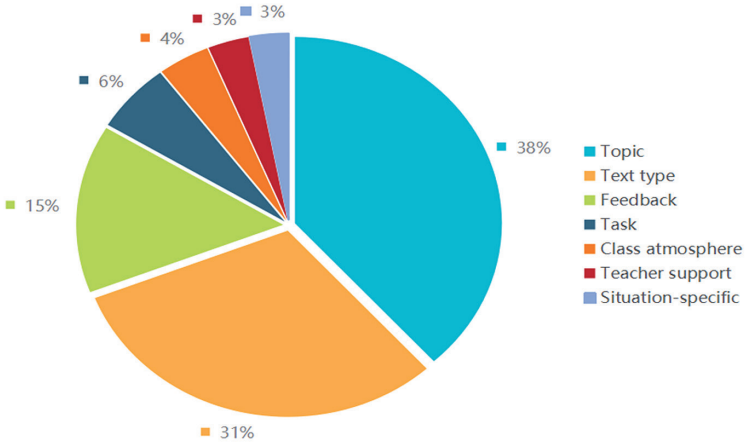
EXTRACT 4

S4 (female, aged 18). I enjoy story continuation writing tasks. One of my favorite tasks is to write a sequel to such a story: A prisoner is faced with a decision where he must open one of two doors. Behind each door is either a lady or a tiger. They may be both tigers, both ladies, or one of each. If the prisoner opens a door to find a lady, he will marry her; and if he opens a door to find a tiger, he will be eaten alive. (Student open question)

4.5.3. Sources of FLWE

Participants' enjoyment episodes yielded 1,577 instances, which fell into seven broad categories. As shown in Figure 1, most students (37.79%, 596/1,577) owned their FLWE to the topic-related factor, followed by the text type-related factor (30.44%, 480/1,577). The feedback-related factor accounted for 14.39% (227/1,577) of the sources of FLWE. The other four categories, i.e., the task-related factor (94/1,577), class atmosphere (63/1,577), teacher support (50/1,577), and situation-specific enjoyment factor (47/1,577) contributed less to FLWE, each taking up less than 10% (5.96%, 3.99%, 3.17%, and 2.98% respectively).

Figure 1: Sources of foreign language writing enjoyment



5. Phase 2: Development and validation of the FLWES

Based on the qualitative findings in Phase 1, the foreign language writing enjoyment scale (FLWES) was developed in Phase 2 by following the steps of item generation (evaluated and refined through expert evaluation, pilot testing, and item analysis), scale development, and scale validation (Devellis & Thorpe, 2021).

5.1. Participants

Phase 2 involved three different student samples. After generating an incipient item pool and having it reviewed by experts, the preliminary questionnaire was piloted on 128 undergraduate English majors recruited from a national key university. Only English majors were selected because they had substantial writing experiences because English writing was an integral part of their undergraduate education.

In the scale exploration and validation stages, both English and non-English majors were involved. Our goal was to make the findings of this study as generalizable as possible.

The phase of scale development involved another sample of 1,094 undergraduate EFL learners recruited from 21 universities across China. However, the data of 79 participants were excluded for multiple reasons such as incomplete information, self-report repetition, duplicate submissions, and repeated participation by using the same IP address that corresponded to their demographic information. As a result, 92.8% (1,015 out of 1094) of the participants provided valid responses, with a mean age of 20 (SD = 1.23), ranging from 18-24. Out of the total, 883 students were females (87%), while 132 were males (13%), with 382 (37.64%) being in Year 2, 235 (23.15%) in Year 4, 200 (19.7%) in Year 1, and 198 (19.51%) in Year 3.

The scale validation phase included another larger sample of participants (1,762) enrolled from 21 universities with different academic rankings. A total of 1,670 out of 1,762 (95%) respondents offered credible answers. In terms of gender distribution, there were 1,417 females (84.85%) and 253 males (15.15%). Regarding the grade allocation, 672 participants were freshmen (40.24%), 415 juniors (24.85%), 407 sophomores (24.37%), and 176 seniors (10.54%). The average age of the participants was 20.4 (SD = 1.24), with a range of 17-25.

For the reliability test, 120 students among the 1,670 participants were invited to participate in the retest of the FLWE Scale. In the end, 113 participants submitted reliable "test–retest" responses, with 87.61% (99) being females and 12.39% males (14). Among them, 68 (60.18%) participants were in Year 1, 25 (22.12%) in Year 4, and 20 (17.7%) in Year 2, with an average age of 19.2 (SD = 1.31).

5.2. Procedures

5.2.1. Initial item pool

Based on the qualitative data in Phase 1, a preliminary set of 60 items was formed. A factor structure was proposed for categorizing these items based on the control-value theory (Pekrun, 2006). The seven proposed factors were (a) Situation-Specific FLWE (Factor A), (b) Topic-Related FLWE (Factor B), (c) Text Type-Related FLWE (Factor C), (d) Feedback-Related FLWE (Factor D), (e) Task-Related FLWE (Factor E), (f) Teacher Support-Related FLWE (Factor F), and (g) Classroom Atmosphere-Related FLWE (Factor G).

5.2.2. Expert review

The preliminary scale was evaluated by three applied linguists and three Chinese instructors with extensive teaching experience in EFL writing. Their task involved (a) evaluating the extent to which each item accurately measured FLWE, (b) judging whether the individual item was well-articulated, and (c) determining whether each item was assigned to its appropriate category (Devellis & Thorpe, 2021). Any discrepancies were addressed through further discussion and corresponding revisions were made. Finally, 11 items were removed, resulting in the 7-factor, 49-item FLWES. The FLWES was then devised as a 5-point Likert scale, with the level of assent spanning from "strongly disagree" to "strongly agree".

5.2.3. Pilot testing

A pilot study was undertaken utilizing a sample of 128 undergraduate English majors. After completing the questionnaire, the students were asked whether they had any difficulties with the questions, and if so, to specify items that they found problematic. Based on students' feedback, 5 items were modified in terms of wording.

5.2.4. Item analysis

Extreme Group Method was employed along with Corrected Item-Total Correlation (CITC) (Dörnyei & Dewaele, 2022). The upper 27% group was compared with the lower 27% group in terms of item difficulty index. CITC evaluates the correlation between individual items and the overall test score (Devellis & Thorpe, 2021). Results indicated that all items were acceptable except an item under Factor D: I tend to enjoy writing more when I'm not being graded by my teacher. It was thus eliminated. All 48 items had acceptable CITC scores, with all the correlations between each item and its subscale above 0.40 (Field, 2013).

5.2.5. Scale development and validation

After scrutinizing the preliminary item pool, the scale was administered to a development sample, aiming to finalize the scale (Devellis & Thorpe, 2021, p.130). To confirm whether the instrument was reliable and valid, the FLWES was then distributed to another group of students designated as the validation sample.

5.2.6. Validation instruments

To assess the FLWES's criterion validity, two widely recognized measures were utilized, namely Pekrun et al.'s (2011) Achievement Emotions Questionnaire (AEQ) Enjoyment Subscale and Li et al.'s (2018) Foreign Language Enjoyment Scale (FLES). The Enjoyment subscale consists of 6 items measuring the degree of enjoyment experienced during academic tasks. The 11-item FLES, comprising 3 factors, namely FLE-Teacher, FLE-Private, and FLE-Atmosphere, gauges enjoyment in foreign language learning. Specifically, the extent to which scores on the FLWES were related to those obtained from the above two questionnaires (Devellis & Thorpe, 2021, p.42).

5.3. Data analysis

To confirm the items of the FLWES, an initial exploratory factor analysis (EFA) was performed using SPSS 25.0. Prior to the extraction of the factors, Kaiser-Meyer-Olkin (KMO) was employed along with Bartlett's Test of Sphericity to determine the suitability of the collected data for factor analysis. The Principal Component Analysis (PCA) extraction method and the Varimax rotation method with Kaiser Normalization were then applied to simplify the scale's structure.

To fine-tune the scale, the scale developed from the early EFA was administered to a different sample of students. Specifically, several types of validity tests were conducted through Confirmatory Factor Analysis (CFA) using AMOS 23.0 software, including construct validity, discriminant validity, and convergent validity. Criterion validity was evaluated by computing Pearson product-moment correlations between scores obtained on the total FLWES and the two well-established instruments mentioned earlier. Furthermore, two kinds of reliability tests were performed, namely internal consistency and test–retest. The first was examined via Cronbach's Alpha, and the second, the Pearson correlation coefficient between the scores obtained on the first and second occasions (Devellis & Thorpe, 2021).

5.4. Results

5.4.1. Scale development

The EFA findings indicated that the scale had a KMO index of 0.879, exceeding the threshold value of 0.70 (Kaiser, 1974). Additionally, Bartlett's test ($p < 0.001$) was significant at the .001 level, implying that the data were suitable for factor analysis. Following Field's guidelines (2013, p. 692), the following items were excluded: (a) items that demonstrated a low factor loading (below 0.4); (b) items that exhibited significant cross–loadings exceeding 0.4; and (c) items that were incongruous to other related items under the same factor.

Finally, 22 items were selected from the preliminary set of items, resulting in a 7-factor FLWE Scale (see Table 3). After Varimax rotation, the proposed factor - Feedback-Related Writing Enjoyment (Factor D), however, consisted of only three items with substantial loadings, with two of these items cross-loading on another factor. Thus, this factor was removed. Specifically, the remaining 6 factors were: (a) Situation-Specific FLWE (Factor A), (b) Writing Topic-Related (Factor B), (c) Text Type-Related (Factor C), (d) Task-Related (Factor E), (e) Teacher Support-Related (Factor F), and (f) Classroom Atmosphere-Related (Factor G). These factors accounted for 64.48% of the variance.

Table 3: EFA results (n = 1,015)

Factor A Situation–Specific		Factor B Topic		Factor C Text Type		Factor E Task		Factor F Teacher Support		Factor G Classroom Atmosphere	
A1	0.665	B1	0.710	C1	0.640	E1	0.732	F1	0.737	G1	0.829
A2	0.744	B2	0.758	C2	0.828	E2	0.684	F2	0.878	G2	0.819
A3	0.651	B3	0.704	C3	0.812	E3	0.678	F3	0.876	G3	0.581
A4	0.696			C4	0.478			F4	0.597		
A5	0.717										

Note. Loadings less than 0.40 were excluded from the analysis.

Table 4: CFA results (*n* = 1,670)

	X^2	*df*	CFI	TLI	SRMR	RMSEA [90% C.I.]
Benchmark			> 0.90	> 0.90	< 0.08	< 0.08
22-item (6 factors)	1392.234*	194	0.930	0.917	0.055	0.061 [0.058, 0.064]
17-item (5 factors)	488.431*	109	0.939	0.924	0.058	0.058 [0.053, 0.063]

Note. *significant at 0.01 level; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; SRMR=Standardized Root Mean Square; RMSEA = Root Mean Square Error of Approximation

5.4.2. Construct validity

Confirmatory factor analysis was carried out to ascertain if the six-factor structure identified in the exploratory factor analysis could be replicated in a new set of participants (Brown, 2015). AMOS 23.0 was used to examine a structural equation model. Multiple fit indices were utilized, including goodness of fit (χ^2/df), the Comparative Fit Index (CFI), Root-Mean-Square Error of Approximation (RMSEA), Tucker-Lewis index (TLI), and standardized root mean square residual (SRMSR). Table 4 displays the fit indices for the 22-item suggested model. Except for the normed chi-square (χ^2/df), which was 7.176, all other fitness indicators met the desired threshold values (Kline, 2016).

To enhance the suitability of the measurement model, adjustments were made by scrutinizing the standardized factor loadings and modification indices. Two items (A3 and A5) with factor loadings below 0.5 were then deleted, resulting in a 20-item scale. Furthermore, upon a thorough analysis of the modification indices, no extremely high MI values were detected (higher than 40).

5.4.3. Convergent validity

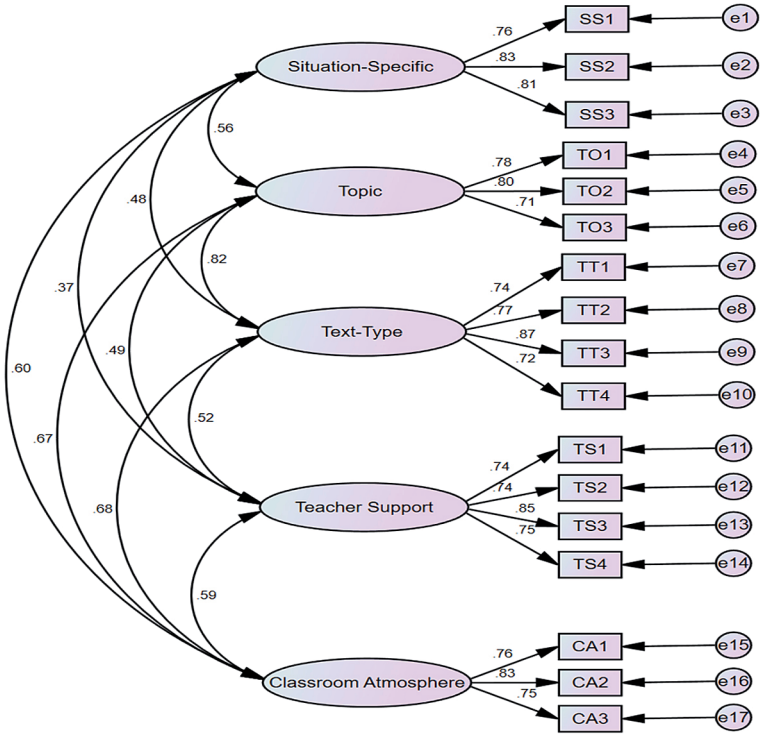
To establish convergent validity, item-total (ITC) values were scrutinized, together with composite reliability (CR), and variance extracted by constructs (AVE). To achieve acceptable convergence, the index values of ITC, CR, and AVE should reach 0.3, 0.7, and 0.5 respectively (Devellis & Thorpe, 2021). As demonstrated in Table 5, all 20 items had loadings higher than 0.5. AVE values of all constructs were greater than 0.5 except for the factor of Task-Related Writing Enjoyment (Factor E). And its CR was 0.638 (less than 0.7). These results suggested that the model fit for this subscale was less than optimal. Thus, the three items loading on Factor E were eliminated to obtain a 17-item scale of five factors. Additional CFAs indicated strong support for the model (see Table 5 and Figure 2).

Table 5: Assessment of each subscale's convergent validity and model fit indicators (*n* = 1,670)

Factor	Name	Item	Name	Convergent validity				Model fit indicators				
				λ	P	AVE	CR	x^2/df	P	CFI	RMSEA	SRMR
Factor A	Situation-Specific	A2	SS1	0.762	0.000	0.639	0.841	0.001/0	0.000	1.000	0.000	0.000
		A4	SS2	0.828	0.000							
		A5	SS3	0.806	0.000							
Factor B	Topic	B1	TO1	0.783	0.000	0.575	0.802	0.000/0	0.000	1.000	0.000	0.000
		B2	TO2	0.796	0.000							
		B3	TO3	0.706	0.000							
Factor C	Text Type	C1	TT1	0.744	0.000	0.606	0.859	27.139/2	0.000	0.992	0.087	0.019
		C2	TT2	0.772	0.000							
		C3	TT3	0.870	0.000							
		C4	TT4	0.719	0.000							
Factor F	Teacher Support	F1	TS1	0.741	0.000	0.594	0.853	52.234/2	0.000	0.983	0.095	0.024
		F2	TS2	0.742	0.000							
		F3	TS3	0.846	0.000							
		F4	TS4	0.745	0.000							
Factor G	Classroom Atmosphere	G1	CA1	0.756	0.000	0.607	0.821	0.000/0	0.000	1.000	0.000	0.000
		G2	CA2	0.830	0.000							
		G3	CA3	0.746	0.000							

Note. SS = Situation-Specific; TO = Topic; TT = Text Type; TS = Teacher Support; CA = Classroom Atmosphere

Figure 2: The five-factor foreign language writing enjoyment model



Note. SS1 = Situation-Specific (item 1); TO1 = Topic (item 1); TT1 = Text-Type (item 1); TS1 = Teacher Support (item 1); CA1 = Classroom Atmosphere (item 1); e1 = errors/residuals of its corresponding item (SS1 in this case)

Normality tests were also performed for all items on the scale. Results showed that they had Skewness and Kurtosis values ranging from -0.795 to -0.244 and 0.110-1.844 respectively, all of which fell within the typical normality ranges of -3 to 3 and -10 to 10, each meeting their respective criteria (Kline, 2016).

5.4.4. Discriminant validity

Discriminant validity refers to how much the measurements of distinct constructs are separate from each other (Bagozzi, et al., 1991, p. 425). To ensure discriminant validity, for each subscale, the square root of AVE must exceed its corresponding r^2 (correlation coefficient) (Devellis & Thorpe, 2021). All constructs had AVE values higher than their matching r^2, suggesting that there was a strong level of discriminant validity for each subscale (see Table 6 and Figure 2).

5.4.5. Criterion validity

To assess the scale's criterion validity, its association with other established scales measuring similar constructs was investigated (Bachman & Palmer, 2010). Results showed that the FLWES yielded significant correlations with Pekrun et al.'s (2011) Achievement Emotions Questionnaire (AEQ) Enjoyment Subscale (r = 0.64, r^2 = 0.4, p < 0.001) and with Li et al.'s (2018) Foreign Language Enjoyment Scale (FLES) (r = 0.84, r^2 = 0.7, p < 0.001).

Table 6: Subscale discriminant validity (n = 1670).

Subscale		r2					AVE
	A	B	C	F	G		
Factor A	Situation-Specific	--					0.639
Factor B	Topic	0.314**	--				0.575
Factor C	Text Type	0.230**	0.672**	--			0.606
Factor F	Teacher Support	0.137**	0.240**	0.270**	--		0.594
Factor G	Classroom Atmosphere	0.360**	0.449**	0.462**	0.348**	--	0.607

Note. ** significant at .01 level

5.4.6. Reliability tests

The FLWES underwent two types of reliability tests, namely the internal consistency (measured by Cronbach's Alpha) and test-retest at a two-week interval. The global FLWES and its five subscales demonstrated high reliability, as evidenced by Cronbach's Alphas of 0.913, 0.841, 0.799, 0.854, 0.848, and 0.816 respectively.

Results of test-retest reliability also indicated a high level of stability for both the overall scale and its five subscales, with reliability values of 0.927, 0.911, 0.852, 0.905, 0.883, and 0.804 respectively (Devellis & Thorpe, 2021).

6. Discussion

The novelty of this study lies in conceptualizing foreign language writing enjoyment (FLWE) from a multi-dimensional perspective and investigating its causes and assessment. In Phase 1, FLWE was found to be a multi-faceted construct. Writing topic emerged as the primary source of FLWE, with text type and feedback coming next. Based on the qualitative findings in Phase 1, Phase 2 developed the FLWES.

Results showed that the FLWES was a reliable and valid instrument gauging FLWE. This section explains how the findings help advance emotion research in foreign language writing.

6.1. The occurrence and conceptualization of FLWE

The first research question sought to explore the occurrence and conceptualization of FLWE among EFL undergraduates from various majors. The interview and questionnaire transcripts suggested that enjoyment is a frequent phenomenon in foreign language writing, echoing findings from prior research (Jin, 2023; Li et al., 2023). The results also indicated that FLWE is a positive emotion distinguished by heightened arousal while engaging in foreign language writing, driven by control and value appraisals. It can be categorized into situation-specific and state FLWE. Specifically, FLWE can be conceptualized from a multidimensional perspective by incorporating (a) the three dimensions, (b) antecedents, and (c) the state/situation-specific dichotomy of achievement emotions (Pekrun, 2006). This demonstrates the applicability of control value theory in defining writing-specific foreign language enjoyment. Previous research, however, did not provide a conceptualization of FLWE, let alone employ a multidimensional approach. Not only does this endeavour deepen our understanding of the nature of FLWE, but it also offers insights into defining enjoyment related to other language skills, such as listening, reading, and speaking.

6.2. Sources of foreign language writing enjoyment

The second research question concerns the sources of FLWE. Our findings not only confirm the positive role of classroom atmosphere, teacher support, peers' active involvement, and tasks fostering autonomy and innovation in triggering enjoyment (e.g., Botes et al., 2021; Dewaele & MacIntyre, 2014, 2017; Li et al., 2018), but also expand the sources of FLWE by including some factors that have been overlooked in extant literature such as situation-specific enjoyment, writing topic, text type and feedback. The identification of situation-specific writing enjoyment aligns with Pekrun and Perry's (2014) assertion that situation-specific emotions, as a type of distal antecedents, affect achievement emotion through control and value appraisals. According to the students' report, they consistently enjoyed writing in a foreign language and found enjoyment in the process of learning and using English through writing. In addition, they regularly experienced enjoyment in improving English proficiency through writing. These findings suggest that the students appraised the value of English writing both intrinsically and extrinsically. Participants also found enjoyment in writing about topics they were skilled in, text types they were familiar with, or those relating to personal experiences. These topics and text types were reported to have helped mitigate the negative impact of task difficulty, a situational antecedent, on

their writing enjoyment. According to Pekrun and Perry (2014), learners' perceived control can be influenced by task difficulty, thus impacting emotions. Additionally, learners' writing enjoyment could be increased when they found writing topics interesting and valuable for discussion. This echoes Pekrun and Perry's (2014) claim that tasks accommodating individual needs can facilitate positive emotions associated with activities. Furthermore, text types that allow individuals to express their ideas or emotions can evoke writing enjoyment. This aligns with Tsai et al.'s (2008) proposition that autonomous tasks can enhance learners' perceived control and intrinsic value of the activity, thus generating positive emotions in writing. Feedback is also considered crucial for eliciting achievement emotions in learning environments characterized by frequent assessments (Pekrun & Perry, 2014). In terms of teacher support, this research confirms that teachers' instructional methods have a significant impact on learners' enjoyment (e.g., Botes et al., 2021; Dewaele & MacIntyre, 2014, 2017; Li et al., 2018). A step further was taken by specifying teacher support into four types (i.e., teaching writing strategies, offering sample articles, providing language cues, and supplementing input materials relevant to the writing topic). Different from prior research, teachers' passion for teaching could be found to improve students' writing enjoyment. Just as Frenzel et al. (2009), and Dewaele and Li (2021) suggest, teachers' enthusiasm for teaching can enhance students' value appraisals, thus promoting associated emotions.

6.3. The measurement of FLWE

In response to the third research question regarding the factor structure of the FLWES, we established a 17-item FLWES consisting of 5 dimensions: (a) Situation-Specific FLWE, (b) Topic-Related FLWE, (c) Text Type-Related, (d) Teacher Support-Related, and (e) Classroom Atmosphere-Related.

This factor structure was evaluated by a series of reliability and validation tests that address the fourth research question regarding the reliability and validity of the FLWES as a measure of writing enjoyment. The FLWES was constructed based on the qualitative findings from Phase 1. This scale is proven to be a valid and reliable instrument through factor analyses and multiple assessments of validity and reliability. It is designed to provide researchers worldwide with a reliable means of gauging foreign language writing enjoyment experienced by undergraduate EFL learners. Importantly, the FLWES can be modified to suit various language learning contexts.

Unlike earlier scales (e.g., Jin, 2023; Li et al., 2023), which focused on writing enjoyment among two specific universities or secondary school students, our study involved a diverse group of EFL undergraduates representing various disciplines. These participants had extensive exposure to English writing, allowing for a more

comprehensive understanding of foreign language writing enjoyment. Notably, they reported engaging with a wide range of text types, including narration, description, exposition, argumentation, and letters. In contrast, secondary school students, primarily for examination purposes, have limited access to narration, description, and letters. Furthermore, the factor structure of our scale provides a holistic perspective of the origins of foreign language writing enjoyment among undergraduate EFL students. In contrast, prior scales focused solely on the essence of enjoyment or limited their examination to the social and private aspects of enjoyment experienced by secondary school students.

The current FLWES, to our knowledge, represents the first instrument particularly developed to measure FLWE among a wide population of undergraduate EFL learners. This endeavour may open a new fruitful area for developing more scales assessing enjoyment in relation to other language skills (i.e., reading, listening, and speaking), across diverse language contexts. Most importantly, our initial attempt may afford new insights into the skill-specificity of other achievement emotions.

7. Theoretical and pedagogical implications

The results carry both theoretical and pedagogical significance. Theoretically, the findings advance our knowledge of the multidimensional nature and measurement of foreign language writing enjoyment (FLWE), which can open up some interesting avenues for further investigations into enjoyment with other language skills or even other skill-specific achievement emotions. Additionally, the results also suggest that control-value theory is applicable in conceptualizing and assessing foreign language writing enjoyment.

Pedagogically, findings from this study imply a need for appropriate writing tasks and a positive classroom atmosphere in EFL writing so as to evoke learners' enjoyment, as suggested by existing research. Given the pivotal role of writing topics and text types in inducing FLWE as revealed by this study, foreign language teachers could assign writing topics that learners find engaging or deserving of discussion. Additionally, teachers can offer writing topics that align with students' strengths, utilize text types familiar to them, or relate to their personal experiences. Educators can also enhance students' writing enjoyment by offering text types that permit students to express their feelings or ideas without any constraints. Regarding feedback as a source of writing enjoyment, the types of feedback such as teacher, peer (Tigchelaar & Polio, 2017), and negotiated (Nassaji, 2017) are found to be related to the outcome of writing, which can boost or impede enjoyment. Therefore, we encourage teachers to incorporate more writing activities that offer negotiated feedback. Moreover, some learners were found to experience situation-specific enjoyment. Pekrun and Perry

(2014) posited that situation-specific emotions, categorized as distal antecedents, affect achievement emotions via proximal antecedents. Such a finding uncovers the necessity for developing or strengthening learners' situation-specific FLWE in the long run. Additionally, providing various kinds of support (i.e., sample articles, language cues, and input materials) can help boost writing enjoyment. Furthermore, showing enthusiasm for teaching writing can help to boost students' enjoyment. In essence, using the FLWES can assist in pinpointing the specific factor(s) of FLWE to improve writing performance. Obtaining a thorough analysis of learners' FLWE through this approach can lead to better decisions about instructional strategies to boost FLWE.

The findings in this study are subject to limitations. Although FLWES has been shown to be a valid and reliable measure of FLWE, caution should be exercised that validation is an iterative and continuous process (Devellis & Thorpe, 2021, p. 113). As such, the scale warrants further validation in distinct foreign language learning environments. Additionally, the development of the FLWES opens up opportunities for mixed-method longitudinal investigations into how learners' FLWE evolves.

8. Conclusion

This study conceptualized and devised a measure of FLWE (see Appendix 1) from a multi-dimensional perspective. It also explores the various sources of FLWE. The results corroborated the applicability of the control value theory in conceptualizing skill-specific achievement emotion, specifically FLWE. Consequently, this research enhances our understanding of FLWE, shedding light on its nature, origins, and assessment.

Findings from this study inspire further inquiry into the skill-specific nature of enjoyment or other achievement emotions. Additionally, this research offers a credible instrument that can be customized to suit diverse foreign language learning contexts.

Acknowledgements

This study was supported by the Humanities and Social Sciences Foundation of Ministry of Education of China (Grant No. 23YJC740090), and the Project of Guangdong Provincial Philosophy and Social Science (Project No. GD23WZXY02-05). We are grateful to Professor Rosa Alonso Alonso and anonymous reviewers for their insightful comments and suggestions. Correspondence should be addressed to Dr. Hai Xu.

9. References

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

Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991). Assessing construct validity in organizational research. *Administrative Science Quarterly*, *36*(3), 421-458. http://dx.doi.org/10.2307/2393203

Biber, D. (1994). An analytical framework for register studies. In D. Biber, & E. Finegan (Eds.), *Sociolinguistic perspectives on register* (pp.31–56). Oxford: Oxford University Press.

Botes, E., Dewaele, J. M. & Greiff, S. (2021). The development of a short-form foreign language enjoyment scale. *The Modern Language Journal*, *105*(4), 858-876. https://doi.org/10.1111/modl.12741

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York: Guilford Press.

Cheng, Y. (2004). A measure of second language writing anxiety: Scale development and preliminary validation. *Journal of Second Language Writing*, *13*(4), 313-335. https://doi.org/10.1016/j.jslw.2004.07.001

Cheng, Y. (2017). Development and preliminary validation of four brief measures of L2 language-skill-specific anxiety. *System*, *68*, 15-25. https://doi.org/10.1016/j.system.2017.06.009

Csikszentmihalyi, M. (2008). *Flow: The psychology of optimal experience*. New York: Harper Perennial.

Csikszentmihalyi, M. (2014). *Flow and the foundations of positive psychology*. New York: Springer.

Devellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications*. London: SAGE.

Dewaele, J. M., & Li, C. (2020) Emotions in second language acquisition: A critical review and research agenda. *Foreign Language World*, *196*(1), 34-49.

Dewaele, J. M., & Li, C. (2021). Teacher enthusiasm and students' social-behavioral learning engagement: The mediating role of student enjoyment and boredom in Chinese EFL classes. *Language Teaching Research*, *25*(6), 922-945. https://doi.org/10.1177/136216882110145

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 the language learner. In T. Gregersen, P. D, MacIntyre, & S. Mercer (Eds.), *Positive psychology in SLA* (pp. 215-236). Clevedon: Multilingual Matters.

Dewaele, J. M., Petrides, K. V., & A. Furnham. (2008). Effects of trait emotional intelligence and socio-biographical variables on communicative anxiety and foreign language anxiety among adult multi-linguals: A review and empirical investigation. *Language Learning*, *58*(4), 911-960. https://doi.org/10.1111/j.1467-9922.2008.00482.x

Dewaele, J. M., Witney, J., Saito, K., & Dewaele, L. (2017). Foreign language enjoyment and anxiety in the FL classroom: the effect of teacher and learner variables. *Language Teaching Research*, *22*(6), 676-697. https://doi.org/10.1177/1362168817692161

Dörnyei, Z., & Dewaele, J. M. (2022). *Questionnaires in second language research: Construction, administration, and processing* (3rd ed.). London: Routledge.

Field, A. (2013). *Discovering statistics using IBM SPSS*. London: SAGE.

Frenzel, A. C., Goetz, T., & Pekrun, R. (2009). Emotionen [Emotions]. In E. Wild, & J. Moller (Eds.), *Padagogische psychologie* (pp. 205-231). Berlin: Springer.

Horwitz, E. K., (2010). Foreign and second language anxiety. *Language Teaching*, *43*, 154–167. https://doi.org/10.1017/S02614480999036X

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

Jin, Y. (2023). The development and validation of the English writing enjoyment scale. *Perceptual and Motor Skills*, *130*(1), 555-575. https://doi.org/10.1177/00315125221137649

Jin, Y., & Zhang, L. J. (2018). The dimensions of foreign language classroom enjoyment and their effect on foreign language achievement. *International Journal of Bilingual Education and Bilingualism*, *24*(7), 1-15. https://doi.org/10.1080/13670050.2018.1526253

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*(1), 31-36. https://doi.org/10.1007/BF02291575

Kline, R. B. 2016. *Principles and practice of structural equation modeling*. New York: Guilford Press.

Li, C., Jiang, G., & Dewaele, J. M. (2018). Understanding Chinese high school students' foreign language enjoyment: Validation of the Chinese version of the foreign language enjoyment scale. *System*, *76*, 183-196. https://doi.org/10.1016/j.system.2018.06.004

Li, C., Wei, L., & Lu, X. (2023). Contributions of foreign language writing emotions to writing achievement. *System*, *116*, 103074. https://doi.org/10.1016/j.system.2023.103074

MacIntyre, P. D. (2016). So far so good: An overview of positive psychology and its contributions to SLA. In D. Gałajda (Ed.), *Positive psychology perspectives on foreign language learning and teaching* (pp. 3-20). London: Springer.

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., Gregersen, T., & Mercer, S. (2019). Setting an agenda for positive psychology in SLA: theory, practice, and research. *The Modern Language Journal*, *103*(1), 262-274. https://doi.org/10.1111/modl.12544

MacIntyre, P. D., Gregersen, T., & Mercer, S. (Eds.). (2016). *Positive psychology in SLA*. Bristol: Multilingual Matters.

Nassaji, H. (2017). Negotiated oral feedback in response to written errors. In H. Nassaji, & E. Kartchava (Eds.), *Corrective feedback in second language teaching and learning: Research, theory, applications, implications* (pp. 114-128). London: Routledge. https://doi.org/10.4324/9781315621432-9

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., 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., & Perry, R.P. (2014). Control-value theory of achievement emotions. In R. Pekrun, & L. Garcia (Eds.), *International handbook of emotions in education* (pp. 117-139). London: Routledge.

Putwain, D. W., Pekrun, R., Nicholson, L. J., Symes, W., Becker, S., & Marsh, H. W. (2018). Control–value appraisals, enjoyment, and boredom in mathematics: A longitudinal latent interaction analysis. *American Educational Research Journal, 55*(6), 1339-1368. https://doi.org/10.3102/0002831218786689

Ryan, R. M., Connell, J. P., & Plant, R. W. (1990). Emotions in nondirected text learning. *Learning and Individual Differences*, *2*(1), 1-17.

Scherer, K. R. (2000). Emotions as episodes of subsystem synchronization driven by nonlinear appraisal processes. In M. D. Lewis, & I. Granic (Eds.), *Emotion, development, and self-organization: Dynamic systems approaches to emotional development* (pp. 70-99). Cambridge: Cambridge University Press.

Tahmouresi, S., & Papi, M. (2021). Future selves, enjoyment and anxiety as predictors of L2 writing achievement. *Journal of Second Language Writing*, *53*, 100837. https://doi.org/1.1016/j.jslw.2021.100837

Teimouri, Y. (2017). L2 selves, emotions, and motivated behaviors. *Studies in Second Language Acquisition*, *39*(4), 681-709. https://doi.org/10.1017/ S0272263116000243.

Tigchelaar, M., & Polio, C. (2017). Language-Focused Peer Corrective Feedback in Second Language Learning. In H. Nassaji, & E. Kartchava (Eds.), *Corrective feedback in second language teaching and learning: Research, theory, applications, implications* (pp. 97-113). London: Routledge. https://doi.org/10.4324/9781315621432-8

Tsai, Y. M., Kunter, M., Lüdtke, O., Trautwein, U., & Ryan, R. M. (2008). What makes lessons interesting? The role of situational and individual factors in three school subjects. *Journal of Educational Psychology*, *100*(2), 460-472. https://doi.org/10.1037/0022-0663.100.2.460

Appendix 1: The Foreign Language Writing Enjoyment Scale (FLWES)

Factor (subscales) and items	Factor loading
Factor 1 Situation-Specific Writing Enjoyment	
(1) I habitually enjoy improving my English level through writing.	0.762
(2) Writing a foreign language (i.e., English) is enjoyable.	0.828
(3) I regularly find enjoyment in the process of learning and using English through writing.	0.806
Factor 2 Topic-Related Writing Enjoyment	
(4) I enjoy writing about an interesting topic.	0.783
(5) I tend to enjoy writing about topics I am good at because I have something to say.	0.796
(6) Working on a topic worthy of discussing is enjoyable because I can express my thoughts through words.	0.706
Factor 3 Text type-Related Writing Enjoyment	
(7) I enjoy writing a text type that allows me to express my opinions.	0.744
(8) I enjoy writing a text related to my personal experiences.	0.772
(9) Writing in a text type that enables me to express my emotions is enjoyable.	0.870
(10) Writing a text type that is familiar to me is enjoyable.	0.719
Factor 4 Teacher Support-Related Writing Enjoyment	
(11) I tend to have enjoyment in English writing if my teacher analyzes writing strategies and skills before I start to write.	0.741
(12) I tend to experience enjoyment in English writing if my teacher provides us with a sample article before I start to write.	0.742
(13) I tend to feel enjoyment in English writing if my teacher offers us language cues.	0.846
(14) I tend to enjoy English writing if my teacher offers us some input (e.g., a video clip or a reading passage) related to the writing topic before I start to write.	0.745
Factor 5 Classroom Atmosphere-Related Writing Enjoyment	
(15) I enjoy learning from peers who have a good command of English writing.	0.756
(16) My peers' positive attitudes towards English writing can positively influence my writing attitude, thus enhancing writing enjoyment.	0.830
(17) The teacher's own enthusiasm for English writing can enhance my writing enjoyment.	0.746

What shapes communicative adequacy in second language speaking performance? The contributions of complexity, accuracy, fluency, and pronunciation

Zhuo Chen
Department of General Courses
Guangzhou Panyu Polytechnic, China
zhuochen@m.scnu.edu.cn

Abstract

The necessity of considering communicative adequacy (CA) in assessing second language (L2) performance has been increasingly recognized, while its nature has yet to be fully explored. The present study examines the relationship between CA and the dimensions of complexity, accuracy, fluency, and pronunciation (CAFP) in L2 speaking assessment. Specifically, the speaking performance of 158 Chinese learners of English was subjectively rated in terms of CA and was also subjectively rated and objectively measured in CAFP. The relationship between the subjective ratings of CA and CAFP and the relationship between the subjective ratings of CA and CAFP and the objective measures of CAFP were analyzed. Results show that the subjective ratings of all CAFP dimensions were significantly correlated with and predicted CA, with pronunciation and fluency ratings making relatively greater contributions to CA than complexity and accuracy ratings, while only the objective measures of verbal complexity, speed fluency, and pronunciation significantly correlated with CA, together accounting for 45% of CA's variance. Furthermore, the subjective ratings of CAFP showed limited correlations with their objective measures. Discussions were made concerning the validity of the construct of CA, the relative contributions of CAFP to CA, and the important role of pronunciation in L2 speaking assessment.

Keywords: communicative adequacy; complexity; accuracy; fluency; pronunciation.

Resumen

La necesidad de considerar la adecuación comunicativa (CA) en la evaluación del desempeño en el segundo idioma (L2) ha sido reconocida cada vez más, aunque su naturaleza aún no ha sido completamente explorada. El presente estudio examina la

relación entre la CA y las dimensiones de complejidad, precisión, fluidez y pronunciación (CAFP) en la evaluación del habla en la L2. Específicamente, el desempeño oral de 158 estudiantes chinos de inglés fue evaluado subjetivamente en términos de CA, y también fue evaluado subjetiva y objetivamente en CAFP. Se analizó la relación entre las evaluaciones subjetivas de CA y CAFP, así como la relación entre las evaluaciones subjetivas de CA y CAFP y las medidas objetivas de CAFP. Los resultados muestran que las evaluaciones subjetivas de todas las dimensiones de CAFP están significativamente correlacionadas y predicen la CA, siendo las evaluaciones de pronunciación y fluidez las que contribuyen relativamente más a la CA que la complejidad y precisión, mientras que solo las medidas objetivas de complejidad verbal, fluidez y pronunciación se correlacionaron significativamente con la CA, representando conjuntamente el 45% de la varianza de la CA. Además, las evaluaciones subjetivas de CAFP mostraron correlaciones limitadas con sus medidas objetivas. Se realizaron discusiones sobre la validez del constructo de CA, las contribuciones relativas de CAFP a la CA y el papel importante de la pronunciación en la evaluación del habla en la L2.

Palabras clave: adecuación comunicativa; complejidad; precisión; fluidez; pronunciación.

1. Introduction

Communicative adequacy (CA) represents how successful a second language (L2) performance achieves its communicative goals (Pallotti, 2009). It has been increasingly used as a way to assess L2 speaking performance alongside other traditional measures such as accuracy and fluency (Koizumi & In'nami, 2024; Kuiken & Vedder, 2022b, 2022c; Révész et al., 2016). This is not only because adequacy is a major goal for L2 learners (Kuiken & Vedder, 2022a) and a basic requirement for effective oral communication (De Jong, 2023) but also for its complementary role to the complexity, accuracy, and fluency (CAF) triad (Pallotti, 2021). However, little is known about the nature of CA, i.e., what linguistic features and to what extent different linguistic features contribute to the CA of L2 performance, and how reliable it could be judged. One way to address these issues is to relate CA with other well-established constructs, as in Révész et al. (2016). Indeed, researchers have long been relating different assessments, subjective or objective, global or specific, of the same language performance to address various issues, for example, to explore the nature of certain assessment methods (Hulstijn et al., 2012; Plakans et al., 2019), the structure of L2 proficiency (Bosker et al., 2013; Suzuki & Kormos, 2020), and the raters' rating process (Kormos & Dénes, 2004; Kuiken & Vedder, 2014a).

The first appropriate benchmark for CA is CAF (Housen et al., 2012; Pallotti, 2021), since it has been most widely used and researched in language testing as well

as theoretically and empirically justified as a valid and reliable construct to gauge L2 performance and proficiency (Larsen-Freeman, 2006; Skehan, 2003). Nonetheless, it is arguable that CAF, which focuses more on the aspects of grammar, would not be sufficient to test CA without the dimension of pronunciation which captures the phonetic features. To begin with, CAF has been challenged for its comprehensiveness (Skehan, 2009). Moreover, pronunciation is a salient and basic dimension of L2 speaking performance (De Jong & Van Ginkel, 1992; De Jong et al., 2012a) and an integral part of major speaking assessments, and thus warrants to be weighed alongside CAF. Besides, pronunciation has been found to influence the comprehensibility (i.e., ease of understanding) of L2 oral production (Suzuki & Kormos, 2020) and impact the success of communication or CA (Derwing & Munro, 2009).

As such, in order to explore the validity of the construct of CA and the relative importance of CAFP in assessing CA, the present study investigates the relationship between three types of assessment of the same speaking performances, namely, raters' subjective ratings and several objective measures of the CAFP of these performances, as well as raters' subjective ratings of their CA. It is hoped that the results can offer insights into the theoretical construction and the pedagogical practice of L2 speaking assessment.

2. Literature review

2.1. Communicative adequacy

Pallotti (2009) was one of the first to explicitly address the importance of adequacy in relation to CAF. Based on a scrutiny of the major problems in defining and operationalizing the constructs of CAF, he proposed that *adequacy*, which is defined as "the degree to which a learners' performance is more or less successful in achieving the task's goals efficiently" (p. 596), should be used as an independent construct, complementing and interpreting measures of CAF. Since Pallotti's seminal work, the argument that interpretations of the CAF of L2 performance may not be satisfactory without considering adequacy has been increasingly recognized (De Jong et al., 2012b; Kuiken & Vedder, 2022a). Several empirical studies have been carried out to explore the nature of the construct of CA and its relationship with CAF.

Hulstijn et al. (2012) analyzed the relationship between L2 Dutch learners' speaking proficiency rated in terms of communicative adequacy, i.e., "the adequacy with which participants were able to perform communicative speaking tasks" (p. 205) and several other aspects of their linguistic competences. They extrapolated the CA scores into the B1 or B2 speaking proficiency levels of the Common European Framework of Reference for Languages and found that different linguistic competences had varied

differentiating power to the two proficiency levels, thus implicitly demonstrating the usefulness of CA in assessing L2 speaking performance. Drawing from the same data pool, De Jong et al. (2012a) directly related L2 learners' functional adequacy (a measure of communicative success) to their linguistic skills. Analyses revealed that, except for the speed of articulation measures (i.e., response latency and response duration), all the other skills were significantly related to the learners' functional adequacy scores, together accounting for 76% of its variance. This result illustrates the componential nature of functional adequacy. De Jong et al. (2012b) further demonstrated that different degrees of task complexity (complex vs. simple tasks) influence functional adequacy measures (rated on six levels), fluency measures (including breakdown fluency, speed fluency, and repair fluency), and lexical diversity measures (Guiraud's index) in different ways. This result effectively supports their argument that measuring CAF "does not amount to the same thing as measuring overall speaking performance" (p. 135). They suggested that CAF combined with CA is likely to predict the overall success of a speaking performance.

In the area of L2 writing assessment, using the data from the Communicative Adequacy and Linguistic Complexity (CALC) study, Kuiken et al. (2010), and Kuiken and Vedder (2014b) looked specifically into the construct of CA and addressed its relationship with linguistic complexity. Kuiken et al. (2010) found significant high correlations between the subjective ratings of CA and the subjective ratings of linguistic complexity. As for the objective measures of complexity, significant correlations were found between CA and lexical diversity and accuracy, but not with syntactic complexity. Kuiken and Vedder (2014b) further revealed that significant correlations existed between overall ratings of CA and linguistic complexity in L2 as well as L1 writing assessment. In addition, raters reported attaching more importance to communicative adequacy than to linguistic complexity. The two studies have expanded our understanding of the nature of CA in L2 writing assessment. More importantly, they justified the research effort of making distinctions between CA and CAF and demonstrated the usefulness of the construct of CA in L2 assessment.

It is worth noting that there is a discrepancy in terminology in the studies presented in this subsection so far. Specifically, half of these studies used the term *communicative adequacy*, while the other half used *functional adequacy*. Kuiken et al. (2010), and Kuiken and Vedder (2014b) used these two different terms for the same construct, which was explicitly defined as "how well participants manage to fulfill the communicative requirements set by the speaking task" (De Jong et al., 2012b, p. 123) based on Pallotti (2009). To maintain consistency, the present study employs the term *communicative adequacy*, following Pallotti (2009, p. 599), who referred to adequacy as "the appropriateness to communicative goals and situations", considering

that *communicative adequacy* displays a focus on whether an L2 speaking performance succeeds in achieving the intended communicative goal. Additionally, this is to distinguish it from the task-related construct of *functional adequacy* that Kuiken and Vedder (2017) developed for L2 writing production.

Taken together, previous studies on CA featured a general agreement on Pallotti's (2009) basic argument of CA and its relative role to CAF, but a coherent and clear definition of CA is still lacking, showing a limited understanding of its nature. Besides, the objective measures of the three dimensions of CAF can only shed light on our understanding of the nature of the construct of CA, and these measures alone are not enough for a comprehensive understanding. Moreover, one or two linguistic dimensions of complexity or fluency would not be sufficient to present a holistic construct as CA and there is still no empirical evidence to show that measuring the three dimensions of CAF jointly equals measuring CA. Therefore, exploring the relationship of CA with CAF is "the most tempting endeavor" (Kuiken et al., 2010, p. 95).

Révész et al. (2016) conducted a study to this aim. They investigated the extent to which CAF predicted CA in L2 oral tasks. Results showed that filled pause frequency was the strongest predictor of CA, explaining 15% of CA's variability, while all other indices of CAF dimensions had significant but small contributions (from 1% to 7%). A model with all the objective measures as predictors accounted for 41% of CA's variance, a power much weaker than the 76% in De Jong et al. (2012a). Their main explanation was that they did not take pronunciation quality as a variable, while in De Jong et al. (2012a), a pronunciation measure was included and showed the strongest contribution (34%) among all the nine linguistic features. Specifically, Révész et al. (2016, p. 846) stated that "A possible explanation why our model was able to explain less variance may lie in the fact that we did not consider pronunciation quality, while this factor had a strong impact on adequacy in De Jong et al. (2012a)". According to their statement, "pronunciation quality" is similar to the "pronunciation quality" analyzed in De Jong et al. (2012a, p. 11), which includes the correctness of speech sounds, word stress, and intonation. Moreover, according to them, it would have been possible to investigate pronunciation in their dataset as De Jong et al. (2012a) did. Whether what Révész et al. (2016) explained was the case should be examined empirically, as it has significant implications for L2 teaching and testing practices regarding the role of pronunciation.

2.2. Complexity, accuracy, fluency, and pronunciation

Unlike CA, the CAF triad has been explored extensively in SLA research (Pallotti, 2021). Most researchers agree that the multifaceted and multidimensional nature of L2 language performance and proficiency is most frequently and adequately captured

by the constructs of CAF (e.g., Housen & Kuiken, 2009; Skehan, 2003). In this study, a fourth dimension - pronunciation is added to CAF in response to the special nature of oral performance and the argument that pronunciation is an integral part of L2 speaking proficiency (Isaacs & Thomson, 2013; Suzuki & Kormos, 2020), resulting in the quartet of CAFP. This addition was also made for the following reasons.

Firstly, although the construct of pronunciation is operationalized differently in different studies, they all focus on the phonological and the acoustic aspects of oral language production, and it has been shown to play an important role in judging overall spoken language proficiency. For example, Higgs and Clifford (1982) found that pronunciation was important across different proficiency levels of the FSI (Foreign Service Institute) speaking test scale. In De Jong and Van Ginkel (1992), pronunciation contributed most to a global proficiency score at lower proficiency levels, and at higher levels it made equal contributions as fluency and accuracy. Iwashita et al. (2008) also revealed that pronunciation in terms of L1 target-like syllables was showing greater relative impact on overall proficiency scores.

Secondly, studies have shown that pronunciation figured prominently in the comprehensibility of L2 speeches, which is "central to interlocutors' communicative success" (Isaacs & Trofimovich, 2012, p. 475). Therefore, it is particularly pertinent to our analysis of the contributing factors of CA. Isaacs and Trofimovich (2012) examined 19 segmental, suprasegmental, fluency, lexical, grammatical, and discourse-level variables to explore which linguistic features influence raters' judgments of L2 comprehensibility at different proficiency levels. The pronunciation measure was the only one showing significant discriminating power between all levels of learners. Saito and Shintani (2016) also found that raters' judgment of L2 speech comprehensibility was primarily influenced by pronunciation and fluency measures and secondarily by lexical and grammatical ones. Similarly, in Suzuki & Kormos (2020), raters' judgment of comprehensibility was best predicted by the objective measures of fluency, grammatical accuracy, and pronunciation. Moreover, in the post-rating interview, pronunciation was the only feature mentioned by all the raters to have influenced their judgment of comprehensibility.

In addition, pronunciation has been found to be independent of CAF. De Jong and Van Ginkel (1992) revealed that accuracy and comprehensibility were two perspectives of a single dimension of message conveyance, while pronunciation and fluency constitute two separate dimensions of speech production. Therefore, they suggested that accuracy, pronunciation, and fluency be treated and measured separately. In Pinget et al. (2014), raters rated the foreign accentedness of the speeches on a nine-point scale ranging from 1 'no accent' to 9 'very strong accent' based on their judgments on the pronunciation of sounds, word stress and intonation patterns.

Their ratings show how much the speakers' pronunciation deviated from the norms of Standard Dutch (p. 359). They also found that raters' ratings of fluency and perceived foreign accent were predicted by different objective acoustic measures and they were only weakly correlated. This finding supported their argument that fluency and accent were separate constructs and should be assessed independently. Note that the raters' rating of accent was based on their judgment of the pronunciation of sounds, word stress, and intonation, and the objective measures of accent consisted of phonemic error rate and pitch range. Therefore, their result for the variable of accent was also indicative of the interdependent relationship between fluency and pronunciation.

3. Research questions

As such, the present study was designed to address the following research questions (RQs).

RQ1: How do raters' subjective ratings of the CAFP of the test takers' speaking performance relate to their subjective ratings of the CA of the same performance?

RQ2: How do the objective measures of the CAFP of the test takers' speaking performance relate to raters' subjective ratings of the CA of the same performance?

RQ3: How do the objective measures of the CAFP of the test takers' speaking performance relate to raters' subjective ratings of the CAFP of the same performance?

4. Methodology

4.1. Participants

One hundred and fifty-eight year-one postgraduate students in an engineering university in China took part in the study on a voluntary basis. They all passed the national entrance examination for postgraduate students in China, which included an English test. According to the examination syllabus, their English proficiency was intermediate to advanced. Their ages ranged from 22 to 28 (M = 22.59, SD = 1.80), with 53 females and 105 males.

4.2. Speaking task

The subjects were instructed to make a 2-3 minutes' speech on the topic of "A positive change in life". They had two minutes to prepare for the speech. Their speeches were rated on the spot for communicative adequacy and they were also recorded for further rating, coding, and measurement.

4.3. Transcription

The 158 recorded speeches were 0.73-3.57 minutes long (M = 1.89, *SD* = 0.56) and were transcribed by an experienced researcher in applied linguistics. Twenty randomly selected transcripts were checked for accuracy by another experienced researcher in applied linguistics to ensure reliability. The check reveals that the first researcher's transcription was accurate and reliable.

4.4. Assessment

4.4.1. Raters' subjective ratings of CA and CAFP

Eight experienced raters, who were also EFL teachers in the subject university, were employed to ensure the ecological validity of the study (Kormos & Dénes, 2004). However, the teacher raters might have been biased in what to listen for, as they were experienced EFL teachers and were familiar with the common mistakes in students' oral performance. To reduce the possibility of any bias, they were clearly informed of the criteria for rating and were given trials to ensure that they fully understood the criteria before they rated the target speeches.

The raters rated the speeches in two ways, i.e., holistically in terms of CA and specifically in CAFP. The holistic rating of CA was done on the spot during the test by the eight raters. They were provided with the definition by Pallotti (2009) (presented in the Literature Review). Following the practice in De Jong et al. (2012a), Hulstijn et al. (2012), Kuiken and Vender (2014b), and others, they were asked to rate the CA of the participants' performance on the same 6-level scales (i.e., *unsuccessful*, *weak*, *mediocre*, *sufficient*, *quite successful*, and *very successful*), but with no detailed descriptors, to ensure that the raters made an intuitive judgment of this insufficiently researched construct (Suzuki & Kormos, 2020). Another consideration is that, as De Jong (2018) pointed out, for the studies relating measures of CAF to overall proficiency, one serious problem is that "what is apparent in the descriptors is likely to emerge as a significant predictor" (p. 243). For example, if the rubrics of oral proficiency contain descriptions of fluency, then it would be natural that a moderate or strong correlation would be found between fluency measures and proficiency ratings (e.g., in Iwashita et al., 2008). Therefore, we purposefully chose not to provide detailed rubrics for the rating of CA. This was also employed to check whether as a construct proposed by SLA researchers, CA is sensible and valid for other practitioners. Nonetheless, the inter-rater reliability coefficient in terms of Cronbach's Alpha for CA was 0.933, showing that the raters were consistent in their judgment. Their average was calculated as the final score for each participant.

The rating of CAFP was done by the same raters based on the recordings of the participants' speeches three weeks after the test. The raters were not informed before the second rating task to ensure that they were not biased or influenced by their first holistic rating practice. Unlike the rating of CA, the rating of CAFP was guided by rubrics adopted from the IELTS speaking band descriptors to ensure that the raters treated CAFP separately, not treating any of them in a broad sense (see Bosker et al., 2013, for discussions of fluency broadly defined being interpreted as overall speaking proficiency). Besides, we used different rating scales for the two ratings, i.e., a 6-level scale for CA and a 9-point scale for CAFP, to avoid artifactual test homogeneity (De Jong & Van Ginkel, 1992) and to further reduce possible interference from the first rating practice. The Cronbach's Alpha for the raters' rating of CAFP was 0.969, 0.946, 0.973, and 0.972, respectively, also showing a high degree of consistency. Similarly, their scores of CAFP were averaged to obtain a final score for each subject.

It should also be noted that unlike Bosker et al. (2013), who used different groups of raters for overall and specific ratings, we deem it important to use the same raters for holistic and specific ratings, as the possible consistency or inconsistency between the ratings of CA and CAFP by the same raters may better reveal the nature of CA, while using different groups of raters is itself a potential source of discrepancy.

After the second rating task, the raters were also asked to rank the relative importance of CAFP in CA.

4.4.2. Objective measures of CAFP

The operationalization of CAFP followed three major principles. First, we employed the widely accepted definitions and measures in the literature to ensure comparability with similar studies. Second, we selected the measures that are generally found to be effective to ensure their validity. Most importantly, instead of trying to achieve full coverage, we used limited yet complementary and distinct measures for each construct to avoid metrics redundancy (Norris & Ortega, 2009) and to reduce the chance of confounding different measures and leading to intercollinearity, since it would make it impossible to estimate the variance uniquely predicted by each predictor variable (De Jong, 2018).

Complexity. Complexity has been variously defined, for example, as "the extent to which learners produce elaborated language" (Ellis & Barkhuizen, 2005, p. 139) or as "the size, elaborateness, richness and diversity of the learner's linguistic L2 system" (Housen & Kuiken, 2009, p. 465). According to Pallotti (2009), complexity poses most problems among the CAF triad for its polysemous nature. For instance, Ellis (2003) listed as many as eight measures of complexity.

This study employed a generally used measure – the amount of subordination, specifically, the ratio of clauses to the Analysis of Speech Units (AS-units) in the subjects' speeches. It was complemented by a measure of specific linguistic features, i.e., the number of different grammatical verb forms in their speeches, including tense, modality, and voice. AS-unit was used instead of t-unit to better suit the features of oral texts (Foster et al., 2000).

Accuracy. Accuracy is generally constructed as "the ability to produce error-free speech" (Housen & Kuiken, 2009, p. 461). The two most often identified measures in the literature are general accuracy and specific measures of accuracy. The former refers to identifying all types of errors and calculating the percentage of error-free clauses or errors per 100 words (e.g., Skehan & Foster, 1999) or the incidence of errors per t-unit (e.g., Bygate et al., 2013), while the latter refers to identifying particular types of error, for example, correct use of vocabulary (Skehan & Foster, 1997) and correct use of plurals (Ortega, 1999).

Following Ellis and Barkhuizen's (2005) proposal, the general measure of accuracy used in this study was "errors per 100 words" and the specific measure was "correct verb forms", which has been long and widely used in the literature (Wigglesworth, 1997; Yuan & Ellis, 2003) (See Table 1).

Fluency. Fluency has also been defined in a variety of ways, for example, as "the capacity to produce speech at normal rate and without interruption" (Skehan, 2009, p. 510), or "the production of language in real time without undue pausing or hesitation" (Ellis & Barkhuizen, 2005, p. 139). Its measurements are mostly of two kinds: in terms of temporal variable, being related to the speed of speaking/writing, and by means of hesitation phenomena, being related to pauses, repetitions, and other dysfluency features (Lennon, 1990).

The present study also operationalized fluency in these two ways. One is the speed of speaking, measured by the number of syllables per minute of speech (Yuan & Ellis, 2003). The other is dysfluency or repair fluency, indexed by the number of reformulations, repetitions, false starts, and replacements per AS-unit, the definitions of which followed Skehan and Foster (1999).

Pronunciation. Two major ways of measuring pronunciation can be identified from the literature. First, the subjects are instructed to read given linguistic materials which include some target words or sounds. The reading is then analyzed with only the chosen words or sounds assessed (e.g., in De Jong et al., 2012a). In the second approach, the subjects' elicited speech is analyzed for certain segmental or suprasegmental features, for instance, how target-like the pronunciation of meaningful words and syllables was

(Iwashita et al., 2008, p. 33), or by calculating the segmental, word stress, or super-segmental error rate (e.g., in Suzuki & Kormos, 2020). Our pronunciation measure mainly referred to Saito and Shintani (2016) and considered the number of erroneously pronounced words, that is, words that have segmental errors (substitution, omission, or insertion of individual consonant and vowel sounds) and word stress errors (misplaced or missing primary stress). The pronunciation was compared to standard British English and American English pronunciation, as that is what the subjects had been learning at school.

All measures and calculations of CAFP are summarized in Table 1.

Table 1: Summary of objective CAFP measures and their calculations

Dimensions	Aspects	Calculations
Complexity	Syntactic	Ratio of clauses to AS-units
	Verbal	Number of different grammatical verb forms in the speech
Accuracy	Specific	Percentage of all verbs used correctly to the total number of verbs
	General	Errors per 100 words
Fluency	Speed	Number of syllables per minute of speech
	Repair	Number of reformulations, repetitions, false starts, and replacements per AS-unit
Pronunciation	–	Number of erroneously pronounced words per 100 words

The first researcher analyzed the transcripts and made the computations. Twenty randomly selected transcripts were coded by the second researcher. The inter-rater correlation coefficient was 0.732, indicating a moderate and acceptable level of reliability (Koo & Li, 2016).

4.5. Data analysis

Three sessions of analyses were carried out using r software (R Core Team, 2023) to address the research questions, namely, descriptive, correlation, and regression analyses. The statistical analysis was complemented by the raters' view of the relative importance of CAFP in CA.

5. Results

5.1. Descriptive statistics

The descriptive statistics of subjectively rated and objectively measured CAFP and subjectively rated CA are summarized in Table 2. Skewness and kurtosis values verify that the scores were normally distributed. Mean scores of subjectively rated CAFP suggest that the subjects had relatively high scores in accuracy, followed in turn by fluency, pronunciation, and complexity.

Table 2: Descriptive statistics of CAFP and CA measures (*N* = 158)

Types	Measures	Mean	SD
Subjective	Complexity	5.521	0.727
	Accuracy	5.730	0.669
	Fluency	5.619	0.879
	Pronunciation	5.623	0.697
	CAFP average	5.624	0.626
	CA	3.960	0.501
Objective	Syntactic complexity	1.646	0.254
	Verbal complexity	4.171	1.365
	Specific accuracy	0.807	0.093
	General accuracy	5.427	2.485
	Speed fluency	118.704	28.230
	Repair fluency	5.672	4.358
	Pronunciation	1.075	1.171

5.2. Results of Research Question 1

The first research question explores how raters' subjective ratings of the specific dimensions of CAFP relate to their subjective ratings of CA. It was answered through a series of correlation and regression analyses. The results of Pearson correlation analysis between the raters' ratings of CAFP and between the ratings of CAFP and CA are presented in Table 3.

Table 3: Correlation between subjective ratings of CAFP and between subjective ratings of CAFP and CA (N = 158)

	Complexity	Accuracy	Fluency	Pronunciation	CAFP average
Complexity					
Accuracy	0.533**				
Fluency	0.649**	0.580**			
Pronunciation	0.688**	0.595**	0.614**		
CAFP average	0.852**	0.791**	0.865**	0.853**	
CA	0.764**	0.745**	0.793**	0.857**	0.938**

Note. **p < .01.

The correlation coefficients between the ratings of CAFP range from 0.533 to 0.688 (p < .01), representing an interrelated yet interdependent relationship among these four dimensions. As for the subjective ratings of CA and CAFP, although the two types of assessments were conducted separately with an interval of three weeks, they were strongly correlated (r = 0.745 - 0.857, p < .01), especially for the average score of CAFP and the rating of CA (r = 0.938, p < .01), indicating that the raters were consistent in rating and their judgment of L2 oral performance. Among CAFP, pronunciation and fluency had relatively higher correlations with CA than complexity and accuracy.

To investigate the extent to which CAFP accounts for the variance of the ratings of CA, a set of stepwise regression analyses were conducted. First, each variable of CAFP was used to predict CA separately. The values of adjusted R^2 shown in Table 4 indicate that pronunciation is the strongest predictor of CA, explaining 73% of the observed variance in CA, with fluency, complexity, and accuracy accounting for 63%, 58%, and 55%, respectively (p < .001).

Table 4: Results of regression analyses for subjective ratings of CAFP as predictors of CA

			Predictors				
Step 1	**R^2**	**Adjusted R^2**	**Estimates**	***SE***	**Statistic**	***p***	**VIF**
Pronunciation	0.734	0.732	0.62	0.3	20.75	<0.001	2.29
Fluency	0.629	0.626	0.45	0.03	16.26	<0.001	2.07
Complexity	0.583	0.581	0.53	0.04	14.78	<0.001	2.27
Accuracy	0.555	0.552	0.56	0.04	13.94	<0.001	1.76
Step 2							
P + F + C+A	0.894	0.891					
Pronunciation (P)			0.31	0.03	10.88	<0.001	
Fluency (F)			0.17	0.02	7.74	<0.001	
Complexity (C)			0.10	0.03	3.76	<0.001	
Accuracy (A)			0.18	0.03	6.82	<0.001	

Secondly, we fitted multilevel models with more than one variable as predictors. According to Plonsky and Ghanbar (2018), the assumptions of multiple regression, e.g., normality, outliers, and multi-collinearity among predictor variables, were checked. Shapiro-Wilk normality tests were carried out on all variables, and the *p* values, ranging from 0.622 to 0.838, confirmed their normal distribution. Variance Inflation Rate (VIF) values were calculated to check multi-collinearity, which showed no such problem (VIF = 1.76 - 2.29). The best fitted model, with the lowest AIC (Akaike Information Criterion) of -113.81, has pronunciation, fluency, accuracy, and complexity as predictors, accounting for 89% of the variance of CA (see Table 4).

Additionally, the results of the raters' view on the relative importance of CAFP in rating CA show that, on average, the eight raters took accuracy and pronunciation as equally important, followed by fluency, and they all ranked complexity as the least important.

5.3. Results of Research Question 2

The second research question addresses how objective measures of CAFP relate to raters' subjective ratings of CA. Similar correlation and regression analyses were carried out.

Table 5 shows few significant correlations among the objective measures of CAFP. Within each of the four dimensions, only general accuracy significantly correlated with specific accuracy, while the two measures of complexity had no significant correlation with each other, nor did the two measures of fluency. As for the relationship between the objective measures of CAFP and the subjective ratings of CA, speed fluency and verbal complexity moderately correlated with the subjective ratings of CA, with objective measures and subjective ratings of pronunciation showing a significant yet limited correlation.

Table 5: Correlation between objective measures of CAFP and between objective measures of CAFP and subjective ratings of CA (N = 158)

	Syntactic complexity	Verbal complexity	Specific accuracy	General accuracy	Speed fluency	Repair fluency	Pronunciation
Syntactic complexity							
Verbal complexity	0.015						
Specific accuracy	0.174*	-0.149					
General accuracy	0.145	0.060	-0.724**				
Speed fluency	-0.143	0.281**	0.080	-0.056			
Repair fluency	-0.472**	0.096	-0.160*	-0.238**	-0.262		
Pronunciation	0.283	-0.131	0.011	0.147	0.019	-0.343**	
CA	-0.010	0.478**	-0.216	0.198	0.513**	-0.148	-0.303*

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

Each objective measure was then used to predict CA. The adjusted R^2 presented in Table 6 shows that speed fluency, verbal complexity, and pronunciation accounted for 26%, 23%, and 9% of CA's variance, respectively.

Besides, models with more than one variable as predictors were also fitted to find the best fitting model. Preliminaries such as normality, outliers, and multi-collinearity among predictor variables were also checked and no such problems were revealed. As in Suzuki and Kormos (2020), and Saito and Shintani (2016), the linguistic measures that did not correlate significantly with the outcome variables were excluded from constructing the model. Subsequently, speed fluency, verbal complexity, and pronunciation together significantly predicted the raters' subjective judgment of CA, explaining 45% of its variance.

Table 6: Results of regression analyses for objective measures of CAFP as predictors of subjective ratings of CA

Predictors					
Step 1	R^2	Adjusted R^2	Estimates	*SE*	*p*
Speed fluency	0.263	0.258	0.01	0.00	<0.001
Verbal complexity	0.229	0.224	0.18	0.03	<0.001
Pronunciation	0.092	0.086	-0.13	0.03	<0.001
Specific accuracy	0.047	0.041	-1.17	0.42	0.053
General accuracy	0.039	0.033	0.04	0.02	0.056
Repair fluency	0.022	0.016	-0.02	0.01	0.063
Syntactic complexity	0.000	-0.006	-0.02	0.16	0.898
Step 2					
SF + VC + P	0.455	0.445			
Speed fluency (SF)			0.01	0.00	<0.001
Verbal complexity (VC)			0.12	0.02	<0.001
Pronunciation (P)			-0.12	0.03	<0.001

5.4. Results of Research Question 3

Table 7: Correlation between objective measures and subjective ratings of CAFP (*N* = 158)

Objective measures	Subjective ratings			
	Complexity	Accuracy	Fluency	Pronunciation
Syntactic complexity	0.011	-0.060	0.237**	-0.128
Verbal complexity	0.568**	0.182*	0.350	0.449
Specific accuracy	-0.158*	-0.126	-0.114	-0.236**
General accuracy	0.148	0.138	0.286**	0.183*
Speed fluency	0.399**	0.540**	0.504**	0.353**
Repair fluency	-0.103	-0.162*	-0.494**	0.037
Pronunciation	-0.167*	-0.013	-0.032	-0.493**

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

The third research question asks how the objective measures and subjective ratings of CAFP relate to each other. The Pearson correlation results summarized in Table 7 show that there was a moderate correlation between verbal complexity, speed fluency and repair fluency, and pronunciation with their corresponding subjective ratings.

Three regression models were accordingly constructed. Results in Table 8 show that objectively measured verbal complexity, fluency, and pronunciation accounted for 32%, 39%, and 24% of the variance of their corresponding subjective ratings.

Table 8: Results of regression analyses for objective measures of CAFP as predictors of their corresponding subjective ratings

Predictors	**Outcome variable**					
Objective	Subjective	R^2	Adjusted R^2	Estimates	*SE*	*p*
Verbal complexity	Complexity	0.323	0.318	0.30	0.04	<0.001
Pronunciation	Pronunciation	0.243	0.238	-0.29	0.04	<0.001
SF + RF	Fluency	0.394	0.387			
Speed fluency (SF)				0.01	0.00	<0.001
Repair fluency (RF)				-0.08	0.01	<0.001

6. Discussion

6.1. Discussion of Research Question 1

Results of the first research question show that subjectively rated CA strongly correlated with and was significantly predicted by all subjective ratings of CAFP. The best model fitted with all four variables accounted for 89% of the variance of CA. This adds to former research findings that objective measures of CAF were related to CA in L2 speaking (De Jong et al., 2012a; Révész et al., 2016) and that subjective ratings of linguistic complexity and accuracy were correlated with subjective ratings of CA in L2 writing (Kuiken & Vedder, 2014b).

Interestingly, unlike previous studies (Kuiken & Vender, 2014b; Révész et al., 2016), the present study employed very general descriptors of CA and still generated consistent results. Specifically, although, except for a general operational definition and basic descriptors of the 6-level scales of CA, no specifics were provided to the

raters, the inter-rater reliability score was high for CA. It is even higher than that for CAFP, the rating of which was guided by detailed rubrics. This indicates that the raters seemed to agree on what CA means in oral performance and carried out the assessment with sufficient reliability. Therefore, in line with the former studies on CA, the present study enhances the evidence that CA is a valid construct for assessing L2 performance and could be reliably rated.

Another finding is that among the four dimensions of CAFP, the best predictors of CA were pronunciation and fluency, while the importance ranking data showed that the eight raters generally perceived accuracy and pronunciation more important for CA in L2 speaking than fluency and complexity. This interesting discrepancy between what raters did and what they thought may indicate that, metacognitively, raters believed that the L2 speech has to be accurate to be communicatively adequate. However, in actual communication, when the speech is above a threshold of accuracy and does not hinder understanding, its CA is not significantly affected (Pallotti, 2021). This finding bears out De Jong and Van Ginkel's (1992) view that oral production assessment is often based first on pronunciation and fluency, and then on the appropriateness of lexical or syntactical choices.

Comparing the present result with the results of other studies that examined the relative contribution of subjectively rated specific features to ratings of overall speaking proficiency, there are discrepancies as well as similarities. Our finding differs from McNamara (1990), who found a crucial role for grammatical and lexical accuracy but a limited role for fluency in assessing overall communicative effectiveness. Nonetheless, the present study partially replicates the findings of Higgs and Clifford (1982) that at lower proficiency levels, raters put most emphasis on vocabulary and pronunciation; as the level goes up, they gave more consideration to fluency and grammar. It is also in line with De Jong and Van Ginkel (1992) that at low proficiency levels, pronunciation contributed most to overall ability ratings, and then were accuracy and comprehensibility, with fluency contributing very little, while at higher proficiency levels, all subskills made equal contributions.

The similarities and discrepancies may, to some extent, be due to the fact that different studies chose different sets of dimensions and constructs, and even for the same dimensions, a wide range of measures and ratings were used.

The finding of the most significant role of pronunciation in predicting CA implies that when evaluating L2 speaking through specific performance features, pronunciation, as an important and unique feature of oral performance, should be adequately recognized and sufficiently analyzed (De Jong & Van Ginkel, 1992). This also suggests that the CAF triad may not present a comprehensive picture of L2

speaking performance (De Jong, 2023; Pallotti, 2009). However, even CAFP may not be comprehensive enough, as in this study, the four dimensions together accounted for 89% of the variance of CA. Since language performance cannot be meaningfully interpreted solely linguistically and without referring to its quality and effectiveness, we support the view of Pallotti (2009) and other relevant studies that communicative adequacy should be employed as "a separate performance dimension" and "as a way of interpreting CAF measures" (p. 590).

6.2. Discussion of Research Question 2

Among the objective measures of CAFP, only the two accuracy measures correlated with each other, while the two complexity measures did not, nor did the two fluency measures. Since we used "complementary and distinct measures" of CAFP to reduce the redundancy of measurements, it is reasonable that the measures of the same dimension had limited or no significant correlations.

Another main finding is that only three out of the seven objective measures of CAFP significantly predicted the subjective ratings of CA, i.e., verbal complexity, speed fluency, and pronunciation. This finding is not singular. In Révész et al. (2016), for nine out of ten CAF measures used to predict CA, their R^2 was quite low, ranging from 0.01 to 0.07. Only breakdown fluency made the highest contribution of 0.15. Similarly, among the nine linguistic skills analyzed by De Jong et al. (2012a), only knowledge of vocabulary and quality of intonation measures were significant predictors of CA, while all seven other variables made no significant contributions. In Douglas (1994), no significant relationship was observed between the subjective scores and objective measures. The underlying reason for this frequently reported limited relationship between the objective measures of CAF and the subjective ratings of oral proficiency may be that there exist a wide variety of measures for each CAF dimension. While each has its rationale, none or no combinations of them are sufficient enough to represent the whole of L2 oral proficiency, not to mention that the present study employed a holistic and more function-oriented construct as CA. However, this does not diminish the value of CAF in assessing the specific aspects of L2 performance and in describing its multidimensionality.

On the other hand, the significant role of verbal complexity, speed fluency, and pronunciation in predicting CA is also informative. Ellis and Barhuizen (2005) found that, unlike accuracy and fluency measures, complexity measures do not provide "a totally consistent picture" (p. 156), pointing to the fact that different complexity measures do not correlate closely with each other. Therefore, it is not unnatural that, in our study, verbal complexity had significant predictive power to CA, while syntactic complexity did not. Similarly, in L2 writing, Kuiken et al. (2010) found significant

correlations between CA and lexical variation but not with syntactic complexity. The significant role of speed fluency and pronunciation replicates the results of several former studies. For instance, Iwashita et al. (2008) showed that token frequency, speech rate, and pronunciation significantly influenced oral language proficiency. Ginther et al. (2010) found that measures of speech rate had strong and moderate correlations with proficiency scores. In De Jong et al. (2012a), vocabulary knowledge and pronunciation were the best predictors of speaking proficiency.

The non-significant role of objective accuracy measures is also understandable. Note that the result of RQ1 shows that the contribution of subjective accuracy to CA is much lower than the other three dimensions. This may again reflect that, when rating CA, raters did not take accuracy as a primary influencing factor as long as the L2 production was accurate enough for understanding. Our result demonstrated Pallotti's (2009) illustration that an inaccurate statement like *No put green thing near bottle* is perfectly functional for achieving the intended communicative goal compared with *colorless green ideas sleep furiously*.

In all, it can be concluded from RQ2 that raters took a relatively holistic view when rating CA, which cannot be sufficiently captured by a combination of objective linguistic features. Additionally, different CAFP measures had different degrees of salience in predicting CA.

6.3. Discussion of Research Question 3

Under RQ3, the objective measures and subjective ratings of each CAFP dimension were related to each other. Similar results emerged and can help to explain the results of RQ1 and RQ2. Specifically, the objective measures of verbal complexity, fluency, and pronunciation significantly predicted their corresponding subjective ratings. The explanation of this limited significant relationship is similar to that of RQ2 and will not be repeated.

The significant relationship found between speed fluency and repair fluency with the subjective fluency ratings in this study has been reported in the literature. For example, Préfontaine et al. (2016) showed that the mean length of runs, articulation rate, and the frequency of pauses all played influential roles in raters' judgments of fluency. Besides, Kormos and Dénes (2004) also found that speech rate, mean length of runs, and pace significantly predicted subjective fluency scores. In Bosker et al. (2013), objective measures of pauses and speed were significant predictors of subjective fluency ratings. Taken together, it can be inferred that objective measures and subjective perceptions of the temporal features of oral production generally have a high degree of consistency.

7. Conclusion

7.1. General conclusion

This study explored the nature and the measurement of the construct of CA of L2 speaking performance by analyzing its relationship with the linguistic dimensions of CAFP. To achieve this, the oral English performance of 158 Chinese EFL learners was subjectively rated in terms of CAFP and CA and was objectively measured using seven indices of the CAFP dimensions. These three types of measurements were related to each other through a series of correlation and regression analyses. It was found that the subjective ratings of all CAFP dimensions are significantly correlated with and predicted CA, with pronunciation and fluency ratings making relatively greater contributions to CA than complexity and accuracy, while only the objective measures of verbal complexity, speed fluency, and pronunciation significantly correlated with CA. Furthermore, the subjective ratings of CAFP showed limited correlations with their objective measures. These findings point to the usefulness and validity of CA as a construct for assessing L2 speaking performance and they also demonstrate its complementary role to CAFP in L2 speaking assessment. Moreover, the significant role of pronunciation and fluency in predicting CA calls for more focused teaching and research effort. Furthermore, the limited correlation between the objective measures and subjective ratings highlights the importance of using both types of assessment to obtain a more comprehensive picture of L2 speaking performance.

7.2. Implications

The findings bear several implications. Theoretically, the construct of CA has been shown to have the potential to be taken as a promising measurement of L2 oral performance and it can sufficiently capture the facets of complexity, accuracy, fluency, and pronunciation of L2 speaking performance. Moreover, in assessing L2 speaking, the construct of CAFP is more valid and sufficient than CAF. Methodologically, this study adds to the empirical studies demonstrating the effectiveness of relating different types of assessment as a way to examine the assessment methods and explore the nature of L2 proficiency. Pedagogically, the prominent role of pronunciation in assessing oral performance calls for more teaching effort on pronunciation. Besides, the communicative or functional aspect of L2 speaking performance needs to be sufficiently considered in the teaching and testing practice. In addition, to get a comprehensive picture of L2 learners' speaking performance, it is recommended that both subjective ratings and objective measures be employed.

7.3. Limitations

One limitation of the study is that the result of the contributions of CAFP to CA was found with subjects who were at the intermediate to advanced levels of proficiency. According to the Relative Contributions Model (Higgs & Clifford, 1982), different proficiency levels would see different magnitudes of influence from the same linguistic factors. Therefore, it is worthwhile to examine the research questions with learners at lower proficiency levels. Moreover, the raters were only asked to rank the relative importance of CAFP for CA, but the underlying reasons or beliefs were left unexplored. Post-rating interviews would be more revealing (as in Trofimovich & Isaacs, 2012). Last, the analysis of CA in the literature almost exclusively concerns the raters' perspective. To obtain a comprehensive picture of CA, the test takers' self-assessments of CA and their similarities and differences with the raters' assessment also need to be investigated.

Acknowledgments

The publication of the work was supported by Guangzhou Panyu Polytechnic (grant number: 2023JG04). The author is grateful to the teachers and students at Information Engineering University in China, who generously participated in the study and to Dr. Fang Nan for her help with the data analysis.

8. References

Bosker, H. R., Pinget, A-F., Quené, H., Sanders, T. J. M., & De Jong, N. H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing, 30*(2), 159-175. https://doi.org/10.1177/0265532212455394

Bygate, M., Skehan, P., & Swain, M. (2013). *Researching pedagogic tasks: Second language learning, teaching, and testing*. London: Routledge. https://doi.org/10.4324/9781315838267

De Jong, J. H. A. L., & Van Ginkel, L. W. (1992). Dimensions in oral foreign language proficiency. In L. T. Verhoeven, & J. H. A. L. De Jong (Eds.), *The construct of language proficiency: Applications of psychological models to language assessment* (pp. 187-205). Amsterdam: John Benjamins Publishing. https://doi.org/10.1075/z.62.19jon

De Jong, N. H. (2018). Fluency in second language testing: Insights from different disciplines. *Language Assessment Quarterly*, *15*(3), 237-254. https://doi.org/10.1080/15434303.2018.1477780

De Jong, N. H. (2023). Assessing second language speaking proficiency. *Annual Review of Linguistics*, *9*(1), 541-560. https://doi.org/10.1146/annurev-linguistics-030521-052114

De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012a). Facets of speaking proficiency. *Studies in Second Language Acquisition*, *34*(1), 5-34. https://doi.org/10.1017/S0272263111000489

De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012b). The effect of task complexity on functional adequacy, fluency and lexical diversity in speaking performances of native and non-native speakers. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 121-142). Amsterdam: John Benjamins Publishing. https://doi.org/10.1075/lllt.32.06jon

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, *42*(4), 476-490. https://doi.org/10.1017/S026144480800551X

Douglas, D. (1994). Quantity and quality in speaking test performance. *Language Testing*, *11*(2), 125-144. https://doi.org/10.1177/026553229401100203

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford: Oxford University Press.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, *21*(3), 354-375. https://doi.org/10.1093/applin/21.3.354

Ginther, A., Dimova, S., & Yang, R. (2010). Conceptual and empirical relationships between temporal measures of fluency and oral English proficiency with implications for automated scoring. *Language Testing*, *27*(3), 379-399. https://doi.org/1177/0265532210364407

Higgs, T. V., & Clifford, R. (1982). The push toward communication. In T. V. Higgs (Ed.), *Curriculum, competence, and the foreign language teacher* (pp. 57-79). Lincolnwood, IL: National Textbook Company.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, *30*(4), 461-473. https://doi.org/10.1093/applin/amp048

Housen, A., Kuiken, F., & Vedder, I. (2012). Complexity, accuracy and fluency. In A. Housen, F. Kuiken, & I. Vedder, (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (Vol. 32) (pp. 1-20). Amsterdam: John Benjamins Publishing. https://doi.org/10.1075/lllt.32.01hou

Hulstijn, J. H., Schoonen, R., De Jong, N. H., Steinel, M. P., & Florijn, A. (2012). Linguistic competences of learners of Dutch as a second language at the B1 and B2 levels of speaking proficiency of the Common European Framework of Reference for Languages (CEFR). *Language Testing*, *29*(2), 203-221. https://doi.org/10.1177/0265532211419826

Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly*, *10*(2), 135-159. https://doi.org/10.1080/15434303.2013.769545

Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition*, *34*(3), 475-505. https://doi.org/10.1017/S0272263112000150

Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics*, *29*(1), 24-49. https://doi.org/10.1093/applin/amm017

Koizumi, R., & In'nami, Y. (2024). Predicting functional adequacy from complexity, accuracy, and fluency of second-language picture-prompted speaking, *System*, *120*, 103208. https://doi.org/10.1016/j.system.2023.103208

Koo, T. K., & Li, M. Y. (2016). A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, *32*(2), 145-164. https://doi.org/10.1016/j.system.2004.01.001

Kuiken, F., & Vedder, I. (2014a). Raters' decisions, rating procedures and rating scales. *Language Testing*, *31*(3), 279-284. https://doi.org/10.1177/0265532214526179

Kuiken, F., & Vedder, I. (2014b). Rating written performance: What do raters do and why? *Language Testing*, *31*(3), 329-348. https://doi.org/10.1177/0265532214526174

Kuiken, F., & Vedder, I. (2017). Functional adequacy in L2 writing: Towards a new rating scale. *Language Testing*, *34*(3), 321-336.

Kuiken, F., & Vedder, I. (2022a). Measurement of functional adequacy in different learning contexts: Rationale, key issues, and future perspectives. *TASK*, *2*(1), 8-32. https://doi.org/10.1177/0265532216663991

Kuiken, F., & Vedder, I. (2022b). Speaking: Complexity, accuracy, fluency, and functional adequacy (CAFFA). In L. Gurzynski-Weiss, Y. J. Kim (Eds.), *Instructed second language acquisition research methods* (pp. 329-352). Amsterdam: John Benjamins Publishing. https://doi.org/10.1075/rmal.3.14kui

Kuiken, F., & Vedder, I. (2022c). The assessment of functional adequacy in language performance. *TASK. Journal on Task-Based Language Teaching and Learning*, *2*(1), 1-7. https://doi.org/10.1075/task.21009.kui

Kuiken, F., Vedder, I., & Gilabert, R. (2010). Communicative adequacy and linguistic complexity in L2 writing. In I. Bartning, M. Martin, & I. Vedder (Eds.), *Communicative proficiency and linguistic development. Intersections between SLA and language testing research* (pp. 81-100). Amsterdam: European Second Language Association.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, *27*(4), 590-619. https://doi.org/10.1093/applin/aml029

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, *40*(3), 387-417. https://doi.org/10.1111/j.1467-1770.1990.tb00669.x

McNamara, T. F. (1990). Item response theory and the validation of an ESP test for health professionals. *Language Testing*, *7*(1), 52-76. https://doi.org/10.1177/026553229000700105

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

Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition*, *21*(1), 109-148. https://doi.org/10.1017/S0272263199001047

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, *30*(4), 590-601. https://doi.org/10.1093/applin/amp045

Pallotti, G. (2021). Measuring complexity, accuracy, and fluency (CAF). In P. Winke, & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 201-210). New York: Routledge. https://doi.org/10.4324/9781351034784-23

Pinget, A.-F., Bosker, H. R., Quené, H., & De Jong, N. H. (2014). Native speakers' perceptions of fluency and accent in L2 speech. *Language Testing*, *31*(3), 349-365. https://doi.org/10.1177/0265532214526177

Plakans, L., Gebril, A., & Bilki, Z. (2019). Shaping a score: Complexity, accuracy, and fluency in integrated writing performances. *Language Testing*, *36*(2), 161-179. https://doi.org/10.1177/0265532216669537

Plonsky, L., & Ghanbar, H. (2018). Multiple regression in L2 research: A methodological synthesis and guide to interpreting R^2 values. *The Modern Language Journal*, *102*(4), 713-731. https://doi.org/10.1111/modl.12509

Préfontaine, Y., Kormos, J., & Johnson, D. E. (2016). How do utterance measures predict raters' perceptions of fluency in French as a second language? *Language Testing*, *33*(1), 53-73. https://doi.org/10.1177/0265532215579530

R Core Team. (2023). *R: A Language and Environment for Statistical Computing* [Software]. R Foundation for Statistical Computing. https://www.R-project.org/

Révész, A., Ekiert, M., & Torgersen, E. N. (2016). The effects of complexity, accuracy, and fluency on communicative adequacy in oral task performance. *Applied Linguistics*, *37*(6), 828-848. https://doi.org/10.1093/applin/amu069

Saito, K., & Shintani, N. (2016). Do native speakers of North American and Singapore English differentially perceive comprehensibility in second language speech? *TESOL Quarterly*, *50*(2), 421-446. https://doi.org/10.1002/tesq.234

Skehan, P. (2003). Task-based instruction. *Language Teaching*, *36*(1), 1-14. https://doi.org/10.1017/S026144480200188X

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*(4), 510-532. https://doi.org/10.1093/applin/amp047

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, *1*(3), 185-211. https://doi.org/10.1177/136216889700100302

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, *49*(1), 93-120. https://doi.org/10.1111/1467-9922.00071

Suzuki, S., & Kormos, J. (2020). Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition*, *42*(1), 143-167. https://doi.org/10.1017/S0272263119000421

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, *15*(4), 905-916. https://doi.org/10.1017/S1366728912000168

Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing*, *14*(1), 85-106. https://doi.org/10.1177/026553229701400105

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, *24*(1), 1-27. https://doi.org/10.1093/applin/24.1.1

Examining two form-focused instruction combinations and their relationship with language-analytic ability in young learners: A study on the English possessive determiners *his/her*

María Martínez-Adrián
Department of English, German and Translation and Interpretation Studies
University of the Basque Country UPV/EHU
maria.martineza@ehu.eus

Kevin Iglesias-Diéguez
Department of English, German and Translation and Interpretation Studies
University of the Basque Country UPV/EHU
kevin.iglesias@ehu.eus

Abstract

Recent work conducted in input-limited classrooms with young learners has made a call for the incorporation of form-focused instruction (FFI) in the form of focused tasks and/or additional treatments that could help learners expand their attention to grammar. Current research has also uncovered the need for the examination of the interaction between FFI, individual differences, such as language-analytic ability, and language achievement. Thus, this paper examines the effect of two forms of FFI (dictogloss+self-correction and dictogloss+metalinguistic explanations [ME]) on learners' accuracy in the use of the possessive determiners *his/her* in the final product of the dictogloss tasks performed. It also looks into the interface between FFI and language-analytic ability as measured by the MLAT-ES for child language. To do so, 33 Basque/Spanish bilingual learners of English (aged 10-11) were split into a dictogloss+self-correction and a dictogloss+ME group. Participants were asked to complete three collaborative dictogloss tasks targeting the possessive determiners *his/her* in three consecutive weeks. The dictogloss+ME group achieved greater accuracy in the use of the possessive determiners *his/her* than the group with self-correction. Likewise, the benefits observed in the former were independent of participants' language-analytic ability, all of which suggests a blurring effect of the ME treatment received.

Keywords: form-focused instruction (FFI); metalinguistic explanations (ME); language-analytic ability; young learners of English as a Foreign Language (EFL); dictogloss.

Resumen

Trabajos recientes llevados a cabo con jóvenes aprendices en contextos de input limitado señalan la necesidad de la incorporación de instrucción centrada en la forma (FFI) mediante tareas focalizadas y/o tratamientos adicionales que podrían ayudarles a prestar mayor atención a la gramática. Investigaciones actuales también han enfatizado la necesidad de examinar la interacción entre FFI, las diferencias individuales, como la habilidad analítica sobre la lengua, y las ganancias lingüísticas. Por tanto, este artículo analiza el efecto de dos tipos de FFI (dictoglosia+auto-corrección y dictoglosia+explicaciones metalingüísticas [ME]) sobre la precisión en el uso de los determinantes posesivos en inglés *his/her* en el producto final de la dictoglosia. También se centra en la interfaz entre FFI y la habilidad analítica sobre la lengua medida a través del MLAT-ES, diseñado para jóvenes aprendices. Para ello, se dividió a 33 aprendices de inglés bilingües en euskera y español (de 10 y 11 años de edad) en dos grupos: dictoglosia+auto-corrección y dictoglosia+ME. Los participantes completaron tres dictoglosias centradas en los determinantes posesivos *his/her* en tres semanas no consecutivas. El grupo dictoglosia+ME obtuvo un incremento mayor en su precisión en el uso de los determinantes posesivos *his/her* que el grupo dictoglosia+auto-corrección. Además, los beneficios observados fueron independientes de la habilidad analítica sobre la lengua de los participantes, lo cual sugiere que el tratamiento empleado puede beneficiar a todos los participantes por igual.

Palabras clave: instrucción centrada en la forma; explicaciones metalingüísticas; habilidad analítica sobre la lengua; jóvenes aprendices de inglés como lengua extranjera; dictoglosia.

1. Introduction

The investigation of attention to form operationalised by language-related episodes (LREs) during the performance of collaborative tasks in the case of young learners has attracted the attention of researchers in the last decade. The existing differences between adults and children in how they approach the language learning process makes research with young learners a priority as previous research findings coming from adult populations cannot be overextended to young learners (Oliver & Azkarai, 2017). In this respect, recent investigations conducted with young learners while performing collaborative speaking+writing tasks have revealed that, unlike adults, LREs are not so elaborated and meaning- and form-related aspects receive

attention in equal terms (Martínez-Adrián & Gallardo-del-Puerto, 2021). In the light of the results obtained, there is an acute need to expand learners' attention to grammar forms through focused tasks alone or in combination with additional treatments, which are techniques encompassed by form-focused instruction (FFI). Even if this line of research is receiving increasing attention in the last years in input-limited settings (Calzada, 2021; Gorman & Ellis, 2019; Serrano, 2011), it is still in its infancy, and more investigations targeting different types of FFI are specially needed.

In particular, even though previous research has attested the benefits of focused tasks such as dictogloss tasks in drawing learners' attention to formal features, these seem to depend on the specific target, this task offering greater advantages in the case of the third person singular *-s* morpheme than in the case of articles (Calzada & García Mayo, 2020). Likewise, some additional treatments to the implementation of dictogloss tasks have not yielded the expected results (Calzada, 2021). Thus, the effectiveness of the incorporation of FFI in the form of focused tasks that elicit specific linguistic features and additional treatments that could maximise the potential of these collaborative tasks by raising learners' awareness of these features either before or after their performance (Loewen & Sato, 2018) merits further investigation in young learners.

A call in recent research has also been made as regards the interaction between FFI and individual differences (IDs) (Roehr-Brackin & Tellier, 2019; Suzuki, 2022; Suzuki & Dekeyser, 2017). Among these IDs, language-analytic ability has been suggested as a predictor of success for the acquisition of grammar properties (Li, 2016; Roehr-Brackin & Tellier, 2019) and FFI to have the capacity to level out these differences (Suzuki, 2022). However, research along these lines is scarce, especially with young learners. In particular, how different forms of FFI may affect the relationship between language-analytic ability and grammar accuracy remains unknown.

This paper aims to fill these gaps by comparing the written production of two groups of Basque/Spanish bilingual young learners of English who have received different types of FFI. Partially building on Gorman and Ellis (2019), one group performed three collaborative focused tasks (i.e. a dictogloss task) and was allowed to self-correct their written production before starting the following task, whereas the other group received two sessions of metalinguistic explanations (ME) in addition to the collaborative dictogloss tasks with no self-correction.

Examining their gains in accuracy in the use of the English possessive determiners ***his/her*** can shed light on the effectiveness of different types and combinations of FFI on language production. More specifically, a collaborative task such as a dictogloss, unspeeded because of its writing component, generates reflection about language, and can force learners to draw upon any metalinguistic knowledge they might have (Ellis, 2003;

Izumi, 2003; Kuiken & Vedder, 2012).[1] In particular, ME can draw students' attention to possessive determiners and trigger its noticing prior to the start of a new task, helping them to build their metalinguistic representations. During the performance of these tasks, learners might resort to the explicit knowledge about these features developed during the ME sessions either in an overt or covert way (Gutiérrez, 2013). As a result, the availability of metalinguistic representations can lead to an increase in learners' accuracy of language forms during production (Johns, 2003; Mitchell, 2000; Thepseenu & Roehr, 2013) and an impact on language development (Ellis, 2003; Erlam, 2013; Izumi, 2003). Moreover, an analysis of how their gains vary with respect to their language-analytic ability can provide further insight into the interplay between FFI and students' IDs.

This paper is structured as follows. Section 2 will review the main findings concerning FFI interventions and highlight the relevance of language-analytic ability in language learning. In sections 3 and 4, the research questions will be posed, and the methodology followed to answer them will be presented. Section 5 will describe the results of the study, which we will discuss in section 6. This paper finishes with some concluding remarks and avenues for further research.

2. Literature review

2.1. Form-focused instruction

FFI can be defined as "any pedagogical effort which is used to draw the learners' attention to language form either implicitly or explicitly [...] within meaning-based approaches to second language (L2) instruction [and] in which a focus on language is provided in either spontaneous or predetermined ways" (Spada, 1977, p. 73). It has often been characterised in terms of its explicitness (see Spada, 2023) and, although both implicit and explicit FFI can be beneficial for L2 learning, the latter has been observed to be more advantageous in meta-analyses (i.e. secondary research aimed at identifying cumulative results arising from a body of primary studies (Roehr-Brackin, 2018, p. 91)) of the effect of different types of L2 instruction such as Norris and Ortega (2001) and Goo et al. (2015), particularly in the short run, and in the development of explicit knowledge. This type of FFI comprises techniques such as L1-L2 contrasts and explanations, among others.

Research on FFI is currently immersed in exploring different types of FFI in different contexts (Nassaji, 2016) and in whether FFI is more effective or necessary for certain forms, particularly in the case of young learners, a population not receiving attention until quite recently (see Collins & Ruivivar, 2020, 2021; Roehr-Brackin, 2024, in this respect). The underlying reason for the scarcity of research along these lines with children may be related to the belief that children learn implicitly, which renders FFI

unsuitable. However, 10- to 12-year-olds show the capacity for abstract thought and are able to focus on linguistic issues (Hakes, 1980). Similarly, recent research suggests that metalinguistic abilities may increase at ages 8–9 when children are provided with age-appropriate FFI (Roehr-Brackin & Tellier, 2019), and that explicit instruction may affect children and adults in the same way (Ferman & Karni, 2010; Lichtman, 2021).

The ample majority of studies on the effect of FFI have targeted learners from Canadian immersion programmes who, despite the large amount of exposure received, lagged behind in terms of the accurate production of certain features, such as past tense/verb aspect (Harley, 1989), grammatical gender in French (Harley, 1998), L2 English question formation (White et al, 1991), and English possessive determiners (White & Ranta, 2002), among others. These investigations proved the effectiveness of various FFI treatments on the development of explicit knowledge of these linguistic features. Nevertheless, as suggested in current research (see Hanan, 2015), the effectiveness of these treatments could have been strengthened by the large exposure to meaningful input offered in immersion contexts.

Evidence from input-limited contexts is less abundant. One of those earlier studies in these contexts was Hanan (2015) with English-speaking learners of German. The combination of explicit information about the target features selected, together with structured input activities, led to gains in learners' ability to use the accusative case in masculine definite articles in German. However, this investigation did not include an explicit-only condition, and consequently, the extent to which explicit information about the target feature contributed to learning remained unknown.

Gorman and Ellis (2019) pioneered in this case by providing young English as a foreign language (EFL) learners (9–12-year-olds) with metalinguistic information in between performing communicative tasks. They also examined direct written corrective feedback as another FFI intervention not previously tested with young learners. In particular, they compared two experimental condition groups (ME and direct written corrective feedback) to a dictogloss+self-correction group who did not receive any treatment at all but were allowed only to self-correct. The study lasted for five weeks in which the three groups completed four dictogloss tasks individually targeting the present perfect simple, a novel feature to the participants. In week 1, the three groups performed the first dictogloss (as a pre-test). In week 2, the treatment received by the dictogloss+ME group consisted of an eight-minute mini-lesson one week after the performance of the first dictogloss task, while the second treatment group were provided eight minutes to study their direct written corrective feedback. The dictogloss+self-correction group were given eight minutes to edit their previous production, not receiving any type of feedback. Participants from the experimental and comparison groups were then asked to perform the second dictogloss. This

procedure was subsequently repeated in week 3 when the three groups performed the third dictogloss. After week 4, in which learners did not perform any task, the three groups completed the fourth dictogloss (as a delayed post-test).

The examination of the results indicated that neither FFI in the form of dictogloss+ME nor dictogloss+direct written corrective feedback had an impact on the learners' accurate use of the target structure over the performance of a dictogloss task coupled with self-correction. As regards ME, the authors acknowledge the lack of practice and the need for more engaging ME for children. With respect to direct written corrective feedback, they suggest that learners may not have been developmentally ready to pay attention to the corrections made, or that they were too focused on reconstructing the content in new pieces of writing that prevented them from drawing upon the explicit knowledge they had gained during the treatment phase. Thus, these authors made a call for further research along these lines.

Other studies conducted in EFL settings have also tested individual FFI types (Calzada, 2021; Serrano, 2011). The study by Calzada (2021) with L1 Spanish 6th grade primary school learners (11–12 years old) examined the effect of pre-task instruction on the accurate use of the third person singular *-s* morpheme and the possessive determiners *his/her* in a dictogloss task. To this end, one group received two 10-minute pre-task focus-on-form mini-lessons to raise learners' awareness of the third person singular *-s* morpheme and the possessive determiners, respectively. These mini-lessons were delivered following the same sequence of activities before the dictogloss task: input enhancement, elicitation and provision of the rule, and practice. Another group just performed the two dictogloss tasks collaboratively, while a third group carried out these tasks individually.

The examination of the reconstructed texts showed that the group with pre-task focus-on-form mini-lessons was more accurate in the use of the target forms, but the mean accuracy score still remained somewhat low across groups (below 50%). The analysis of the results in terms of LRE production also confirmed greater benefits of the pre-task focus-on-form mini-lessons in the case of the third person singular *-s* morpheme than in the case of the possessive determiners. As explained by the author, this finding could be ascribed to (1) the existence of more instances of the third person singular *-s* morpheme in the input text, which could have resulted in greater opportunities for languaging, (2) to the design of the task which could have prompted the omission of possessive determiners or the substitution by 'the', as well as (3) the lesser amount of instruction and low frequency of this feature in the input.

Other studies targeting the English possessive determiners (Serrano, 2011) have explored the effect of FFI in the form of metalinguistic explanations and practice.

Two groups of primary school learners immersed in a Content and Language Integrated Learning (CLIL) programme participated in the study. One group received metalinguistic instruction for 6 weeks, while a control group did not. The three groups completed three different tasks at a pre- and a post-task stage: a metalinguistic judgement task to look into whether learners demonstrated knowledge of the target features, as well as an error correction task and a production oral task to measure their performance. The results obtained indicated that the treatment received by the experimental group did not have an impact on metalinguistic knowledge as measured by the metalinguistic judgement task, but a slightly positive impact on the performance of the controlled task (error correction) and the production oral task. The positive effect of metalinguistic instruction was suggested to be dependent on individual differences, such as learners' analytic ability.

Even if FFI has attracted the attention of researchers targeting young learners in input-limited settings in recent years, investigations that comprise a wider range of FFI techniques, including additional FFI treatments to the performance of focused tasks, are needed. Some of the investigations already conducted did not yield significant improvements on the selected targets (i.e. Gorman & Ellis, 2019), in others subtle improvements were observed, and greater benefits were attested for certain features (e.g. third person singular *-s* morpheme) more than others (possessive determiners *his/her*) (i.e. Calzada, 2021).

This paper will try to contribute along these lines by examining two forms of FFI (dictogloss tasks+ME and dictogloss tasks+self-correction) following Gorman and Ellis (2019), though with the incorporation of three modifications which could reinforce the learning potential of a dictogloss task for the target feature selected (possessive determiners *his/her*): (1) the administration of a collaborative task instead of an individual dictogloss task, which has proved effective in a variety of contexts (Basterrechea & García Mayo, 2014; Calzada, 2021), and which could be more appropriately integrated in a meaning-oriented learning context, such as CLIL educational programmes, in which participants from the current investigation are enrolled; (2) the inclusion of an engaging treatment with a noticing and an awareness phase followed by practice along Bouffard and Sarkar (2008), and Lyster (2007, 2015); and (3) the incorporation of precise instructions for learners to draw their attention to form rather than meaning during task performance.

The examination of the use of possessive determiners in the written output will allow us to assess the differential effect of two forms of FFI on the application of the explicit knowledge in productive language use. As a written task, characterised by a more self-determined pace (Kuiken & Vedder, 2012), and as a task performed in collaboration, the assumption is that this task generates reflection about language

(Erlam, 2013; Gutiérrez, 2013; Thepseenu & Roehr, 2013), and thus the possibility of drawing upon explicit knowledge about possessive determiners facilitated by explicit instruction in the form of ME, all of which could lead to greater accuracy for these language forms in the ME group.

2.2. Language-analytic ability

Language-analytic ability, being a component of language learning aptitude, has been defined as the ability to infer linguistic systematicities from the input and make generalisations, encompassing both grammatical sensitivity and inductive language learning ability (Skehan, 1998, 2002). Previous studies have mostly explored its relationship with grammar development in adults, with a meta-analysis concluding that this ID is a good predictor of grammar learning over other components of aptitude (Li, 2016). Likewise, recent studies have also confirmed that this ID is a predictor of success for the acquisition of grammar properties in children (Roehr-Brackin & Tellier, 2019; Rosa-González, 2011).

Latest research has also been immersed in how FFI interacts with different IDs (DeKeyser, 2021; Lado & Sanz, 2021; Li, Ellis, & Zhu, 2019; Pawlak, 2021; Suzuki, 2022). For language-analytic ability in particular, a compensation pattern where FFI inhibits the effects of the ID, being equally effective for all students, has been observed in adults (Suzuki, 2019; Suzuki, Yokosawa, & Aline, 2022) and adolescents (White & Ranta, 2002). As Suzuki (2022) claims, "extra support (e.g., the aid of explicit information) can level out learners' differences" (p. 299). In this regard, current investigations are concerned with the interaction between this ID and different types and combinations of FFI (Kachinske & DeKeyser, 2019; Li et al., 2019) as it remains unknown if such a compensation pattern will be observed with any FFI implementation (see Serrano, 2011 for an example where a compensation pattern between FFI in the form of ME only and language-analytic ability was not found).

Since language-analytic ability has also been found to be a good predictor in children as young as 8 years old (Roehr-Brackin & Tellier, 2019; Rosa-González, 2011), examining the interaction between different types of FFI and this ID in young learners will undoubtedly contribute to uncover the optimal combination of individual and contextual factors needed to maximise the benefits of explicit learning and teaching in this population. Considering the scarcity of studies that have undertaken this research (e.g. Li et al., 2019) and the recent calls to compare different teaching approaches in the same age group, or to compare same-age children with different levels of language learning aptitude participating in the same teaching approach (Roehr-Brackin, 2022, 2024), the present investigation will try to fill these gaps by exploring the interaction of language-analytic ability and FFI in two groups of young learners: one receiving FFI

in the form of focused tasks (i.e. dictogloss) coupled with self-correction, and another one in the form of ME in addition to the performance of the same focused tasks.

3. Research questions

Despite the increasing attention FFI is receiving in the case of young learners in limited-input settings, research is still in its infancy and more investigations targeting different types and combinations of FFI interventions are needed. Likewise, even if FFI has been reported to blur differences ascribed to language-analytic ability (Suzuki, 2022), research with young learners is still limited and, in particular, how different ways of FFI may affect the relationship between this ID and language achievement remains an open question. Thus, this study, integrated in a bigger project where different FFI conditions have been explored in intact classes as a response to a call for more authentic classroom research (Loewen & Sato, 2018; Philp, Walter, & Basturkmen, 2010; The Douglas Fir Group, 2016), will try to contribute to the aforementioned gaps by addressing the following questions:

1. Does FFI in the form of metalinguistic explanations after the performance of a focused task lead to greater accuracy in the use of the target feature selected in reconstructed texts than self-correction after the performance of a focused task?

2. How do different forms of FFI (dictogloss+metalinguistic explanations; dictogloss+self-correction) mediate the relationship between language-analytic ability and use of the target feature selected?

4. Methodology

4.1. Participants

Thirty-three Basque/Spanish bilinguals (20 girls and 13 boys) from two intact classes in 5th grade from a semi-private school voluntarily participated in this study after obtaining parental consent. All of them came from Spanish-speaking families, but thanks to their enrolment in a trilingual programme, students achieve a good command in L2 Basque, so they can be considered Basque/Spanish bilinguals dominant in L1 Spanish (Cenoz, 2009). They were between 10 and 11 years old (M=10.2, SD=0.4) and, on average, had started learning English at the age of 3 (M=3.3, SD=1.8). All of them claimed to have a perfect command of Spanish, a Romance language.

Learners were taking part in a trilingual CLIL programme where Spanish, Basque and English were the languages of instruction. CLIL lessons were characterised by the

lack of FFI interventions in their lessons. As regards content subjects, they were taking mathematics in Spanish; physical education, arts and music in Basque; and sciences, robotics and religion in English. In addition to this, they also received lessons that focused on the language itself: 3 times a week for English, and 4 times a week for Spanish and Basque each. In total, they were exposed to approximately 363 hours of English a year during class time. Outside of the school, 30% of students in each class attended private English lessons for 2 hours on average (M=2.1, SD=1.1). The characteristics of these participants reflect the regular profile of learners during middle childhood (7–11 years old) attending a semi-private school in a bilingual community in Spain, usually enrolled in a CLIL programme, and some of them attending private English lessons. Thus, the outcomes obtained in the present study could be potentially applicable to typically intact classes. At the time of testing, they had an A1-A2 level of English according to the Flyers Test (Cambridge Assessment English, 2018). The administration of this test allowed us to ensure that we had two comparable groups in terms of general language proficiency, as confirmed by a Mann-Whitney U test (z=1.087, p=0.290, d= 0.385). In terms of the target feature knowledge, they were also comparable as evidenced by the lack of statistically significant differences in a grammaticality judgement task (GJT) administered at the outset of the study (z=0.674, p=0.509, d=0.234). This task included 16 items related to the target feature in both gender-matched and gender-mismatched contexts as well as 12 distractors (see section 4.2 for an explanation of gender-matched and gender-mismatched contexts in the use of English possessive determiners; see Iglesias-Diéguez & Martínez-Adrián, in press for a more detailed analysis of the GJT).

Each of the two classes were randomly assigned to one of two conditions: the dictogloss+self-correction group (n=15; 10 girls and 5 boys), who received explicit FFI in the form of collaborative dictogloss tasks coupled with self-correction, and the dictogloss+ME group (n=18; 10 girls and 8 boys), who were involved in awareness-raising sessions in addition to the collaborative dictogloss tasks. To carry out the dictogloss, participants in each group were paired up considering their proficiency level as measured by the Flyers test, resulting in 6 pairs and 1 triad in the dictogloss+self-correction group and 9 pairs in the dictogloss+ME group. Having proficiency-matched pairs with children fosters collaborative patterns of interaction and allows them to pool their resources and remain on task (Basterrechea & Gallardo-del-Puerto, 2020; García Mayo & Imaz Agirre, 2019).

4.2. Target structure

So as to investigate the effect of FFI in the form of ME, we selected the English third person possessive determiners *his/her* as the target structure. Unlike Gorman and Ellis (2019), this feature is not novel to the participants from the present study,

but it has been selected on the grounds that it poses considerable difficulty to students with Romance L1s (White et al., 2007) due to the differences in gender assignment between Romance languages and English. Whereas the English possessive determiner depends on the gender of the possessor, Romance languages assign the gender of the possessive determiners based on the gender of the possessed entity. In particular, possessive determiners in Spanish do not bear any marking of gender, the determiners being equal regardless of the gender of the possessor or of the possessed entity (e.g. *su coche* 'his/her car'). However, number agreement occurs with the possessed entity (e.g. *su coche* 'his/her car', *sus coches* 'his/her cars'), and the gender assignment typical of Romance languages can still be seen in third person possessive pronouns (e.g. *el suyo* 'his/hers' [masc.], *la suya* 'his/hers' [fem.]). This difference in gender-agreement systems gives rise to two different contexts from the perspective of a learner of English with a Romance L1 background: (1) *gender-matched contexts*, when the gender of the possessive determiner agrees with the gender of the entity being possessed (e.g. *his grandfather*, *her aunt*); and (2) *gender-mismatched contexts*, when the gender of the possessive determiner differs from the gender of the entity being possessed (e.g. *his grandmother*, *her uncle*).

Previous studies emphasise the importance of the context as they report different accuracy scores by Basque/Spanish bilinguals in the production of English possessive determiners. While both contexts pose difficulty for learners of English, gender-mismatched contexts seem to be more challenging for speakers with a Romance L1 background (Imaz Agirre & García Mayo, 2013). Therefore, Basque/Spanish bilinguals need to overcome this crosslinguistic influence, especially in gender-mismatched contexts, to achieve target-like use of the possessive determiners in English. In this regard, note that Basque does not have grammatical gender. Therefore, the difficulties in dealing with the possessive determiners *his/her* in the case of Basque/Spanish bilinguals are ascribed to the influence of gender-assignment in Spanish, as has been documented in previous studies with L1/L2 Basque learners of English (Imaz Agirre & García Mayo, 2013).

4.3. Design

Two groups participated in four sessions as indicated in Table 1. In the first session, all participants completed part II of the MLAT-ES test (Stansfield & Reed, 2005) to measure their language-analytic ability. In session 2, both groups completed the first dictogloss task, which served as the pre-test to check their accuracy in the use of the possessive determiners. Session 3 was devoted to the round of treatment based on dictogloss 1 for the dictogloss+ME group (15 minutes) followed by the second dictogloss in both groups. To counteract any effect that the additional time with the outcome of dictogloss 1 may have in the dictogloss+ME group, pairs in the dictogloss+self-correction group were given 15 minutes to go through their previous

writing from dictogloss 1 and self-correct it. The same procedure was followed in session 4 with the round of treatment for the dictogloss+ME group based on dictogloss 2, the self-correction time for the dictogloss+self-correction group, and the completion of the third dictogloss in both groups, which was used as the post-test.

Table 1: Design of the study

	Dictogloss+self-correction group	**Dictogloss+ME group**
Session 1	MLAT-ES (Part II) (30')	MLAT-ES (Part II) (30')
Session 2	Dictogloss 1 [PRE-TEST] (10')	Dictogloss 1 [PRE-TEST] (10')
Session 3	Self-correction of dict. 1 (15') Dictogloss 2 (10')	Round 1 of ME treatment (15') Dictogloss 2 (10')
Session 4	Self-correction of dict. 2 (15') Dictogloss 3 [POST-TEST] (10')	Round 2 of ME treatment (15') Dictogloss 3 [POST-TEST] (10')

The design of this study builds on Gorman and Ellis (2019), although there are three important differences. First, the dictogloss task was carried out in pairs to foster students' focus on form as previous studies show that carrying out a written task collaboratively can be beneficial for language development (Fernández Dobao, 2012; Kim, 2008; Storch, 1999) and results in a positive experience for students (Fernández Dobao, 2020; Fernández Dobao & Blum, 2013). Secondly, as suggested by Gorman and Ellis, we attempted to make the treatment (i.e. ME sessions) more engaging by using students' own written production during these sessions (as in Bouffard & Sarkar, 2008) and varying the delivery method for the practice phase based on Lyster (2007, 2015). Thirdly, all participants were explicitly encouraged to produce grammatically accurate texts as part of the instructions for the dictogloss task, in order to avoid situations reported in Gorman and Ellis (2019), where learners were so focused on reconstructing content that it prevented them from drawing upon the explicit knowledge gained during treatment.

4.4. Instruments and materials

4.4.1. Dictogloss

The collaborative task that students had to carry out in pairs consisted in a dictogloss task (Wajnryb, 1990), where learners had to replicate (i.e. recall and reconstruct) a text in written form that they had previously listened to. This task has been characterised as a focused task as it is designed with the intention of drawing learners' attention to specific language features in a meaning-based context (Benati,

2021; García Mayo, 2002) and has been previously used when targeting the possessive determiners *his/her* (Calzada, 2021). For this study, three different texts were developed and recorded by an L1 English speaker. The texts provided familiar situations to our participants as they narrated a family's routines in a typical day in the garden, a day that they spend together on holidays and a day deciding and buying what they want after having won the lottery. The second dictogloss can be found in Appendix 1 as an illustration.

Considering the limitations acknowledged in previous studies examining production data and the acquisition of possessive determiners (Calzada, 2021), as well as the instructional context in which the study was conducted, these texts were kept as short as possible (107 words each) to avoid learners' attention focusing excessively on meaning rather than on form. Moreover, the number of instances of the target structure appearing in each text has been increased with respect to previous studies: each dictogloss contained 8 instances of possessive determiners, half of which were gender-matched and the other half gender-mismatched. These were obligatory contexts (OCs), where the possessive determiners could not be substituted or omitted. Moreover, the possessives in each of those contexts were split into an equal number of *his* and *her*. This resulted in a total of 2 gender-matched *his*, 2 gender-matched *her*, 2 gender-mismatched *his* and 2 gender-mismatched *her* in each text. Additionally, the texts were syntactically simple, making minimal use of embedded clauses and favouring coordination, and lexical items contained in the texts were appropriate for the level of our participants.

Each dictogloss was accompanied by a worksheet for each student where they could take notes. Unlike in previous studies, these worksheets contained 4 keywords that appeared in the texts to avoid omissions of the target structure or substitution by the definite article *the*, which has been observed in previous research when an explanatory image is included (Calzada, 2021). These keywords also contributed to reducing participants' cognitive load when trying to remember the content of the story.

4.4.2. MLAT

To measure participants' language-analytic ability, part II of the MLAT-ES in pen-and-paper format was used (Stansfield & Reed, 2005). The MLAT-ES is a modified version of the MLAT-E (Carroll & Sapon, 2002) for children specifically geared towards L1 Spanish young learners between the ages of 8 and 13. This test has proved reliable with children in a variety of settings (Kiss & Nikolov, 2005; Roehr-Brackin & Tellier, 2019; Tellier & Roehr-Brackin, 2013), including L1 Spanish children (Gesa & Suárez, 2022; Muñoz, 2014). Part II of the test intends to measure students' sensibility to grammatical structures avoiding explicit reference to grammatical terminology. To

do so, participants had to recognise the function of a given word in a sentence in Spanish and identify which word in a second sentence had the same function. This part of the MLAT-E has been used to measure language-analytic ability in previous studies (Roehr-Brackin & Tellier, 2019).

4.5. Procedure

Prior to the beginning of the study, all participants completed a background questionnaire that gathered data about the languages they knew and their participation in any extracurricular activities in English, such as private lessons.

Before starting with the sequence of dictogloss tasks, all participants completed the MLAT-ES. It contained 4 training items that were used to make sure all students had understood the instructions correctly. After this, they were given 30 minutes to complete the test.

In each dictogloss task, the dictogloss worksheet described above was handed out to each student. Before listening to the text, they were asked to read aloud the four keywords presented at the top of the page. If a participant did not understand the meaning of a keyword, they were told the meaning of the word in Spanish to facilitate comprehension of the text. After this, they listened to the text twice. The first time, they simply listened to it, whilst the second time they were allowed to take notes individually in their worksheets, although they were encouraged to avoid writing full sentences. Then, each pair was given 10 minutes to reconstruct the story using the keywords and sharing their notes. Both measures could reduce learners' cognitive load, facilitating the recall of the content (Leeming, Aubrey, & Lambert, 2022), and thus helping learners focus on formal aspects of the language during the reconstruction phase. They were also told that they had to produce a grammatically correct text.

The ME treatment consisted in 15-minute mini-lessons that took place in class immediately before starting dictogloss 2 and 3, and were delivered by one member of the research team. Based on Bouffard and Sarkar (2008) and Lyster (2007, 2015), each mini-lesson was structured into the same three phases: noticing, awareness-raising and practice. In the noticing phase, students were shown some sentences from the previous dictogloss that contained, at least, an error related to the possessive determiners. The purpose of including participants' own production was to foster their engagement in the treatment. The researcher asked the whole class if they could identify any mistakes in those sentences and elicited answers from different students. They were also asked to provide a correction for those mistakes.

After having gone through the examples, the awareness-raising phase began by posing questions to students about how possessive determiners work in English (e.g., *why do we say 'his' and not 'her' in some of the sentences we have corrected?*) as well as about the reason why they may be difficult for Spanish speakers (e.g., *why do you think we make these mistakes?*). During this phase, they were guided by the researcher towards a general rule to explain gender agreement in English, which was explicitly stated and shown to students before continuing to the practice phase.

This last phase consisted in a cloze activity, where students had to select the correct possessive determiner to complete several sentences. In the first mini-lesson (before dictogloss 2), students had to raise a flashcard with the correct possessive determiner to complete the sentences that they were shown in the whiteboard, whereas in the second mini-lesson (before dictogloss 3), the activity was carried out through *Kahoot!*, an audience response tool, using students' tablets. We varied the way of delivering the practice in an attempt to make the activity more fun and dynamic for young learners. Except for the practice phase in the ME mini-lessons, the dictogloss+ME group did not receive any feedback on their production or any feedback related to the possessive determiners.

In the dictogloss+self-correction group, participants did not receive any specific instruction on the possessive determiners. Likewise, this group did not receive any feedback on their dictogloss production or on their self-correction of the text, as in Gorman and Ellis (2019). To compensate for the extra time that the dictogloss+ME group spent engaging with their production from the previous dictogloss, the dictogloss+self-correction group was also given 15 minutes to self-correct their previous writing. The self-correction session also took place immediately before starting a new dictogloss task to maintain comparability between the groups.

4.6. Data coding and analysis

Participants' change in the accuracy of the use of the possessive determiners was obtained in the following way. First, the OCs for the target structure were identified and counted in students' texts. When the correct determiner was provided in an OC, it was counted as an instance of correct use of the target structure. If a possessive determiner was incorrectly provided in a context where it was not needed, it was coded as an instance of overproduction. With this information, Pica's (1983) formula was used to obtain a percentage for the target-like use (TLU) of the possessive determiner for each participant in each dictogloss:

$$\text{“TLU”} = \frac{\text{“instances of correct use in an OC”}}{(\text{“number of OCs”} + \text{“instances of overproduction”})}$$

The TLU is considered a valid measure of accuracy in both written and oral production (Ellis & Barkhuizen, 2005) and has been used in previous studies with this purpose (e.g. Akbaş & Ölçü-Dinçer, 2021; Frear & Chiu, 2015; Vraciu, 2020).[2]

Finally, the difference between the TLU in dictogloss 3 and the TLU in dictogloss 1 for each participant was obtained as a measure of the development in students' accuracy in the use of the target structure. To control for the influence of prior knowledge students may have, gains were used instead of raw scores when comparing the dictogloss+self-correction and dictogloss+ME groups (File & Adams, 2010; Xu & Li, 2021; Zimmerman & Williams, 1982). Although we acknowledge that participants with higher accuracy in the pre-test can obtain smaller gains, there were no statistically significant differences in the distribution of high and low scores in the pre-test (z=-0.606, p=0.62, d=0.311) between the two groups.

As regards the MLAT-ES, a point was awarded for each correct answer. The maximum score that could be obtained in part II of this test was a total of 30 points. There was no difference in the distribution of MLAT scores between the two groups (z=0.911, p=0.381, d=0.319).

Although the Kolmogorov-Smirnov tests for normality were not significant for some variables in each group, exploration of the histograms and Q-Q plots revealed deviance from normality in the residuals. Therefore, non-parametric tests were used in all instances. Cohen's *d* effect size values were calculated, and subsequently interpreted following the specific benchmarks for L2 acquisition proposed by Plonsky and Oswald (2014) for both intergroup comparisons (values around 0.4 were considered small, medium around 0.7 and large if around 1.0) and intragroup comparisons (values around 0.6 were considered small, medium around 1.0 and large if around 1.4).

To answer the first research question, which looked into the effect of ME on the accurate use of *his/her* in gender-matched and gender-mismatched contexts, we analysed the difference in the accuracy of the possessive determiners in dictogloss 1 (i.e. pre-test) and dictogloss 3 (i.e. post-test) in each group, as well as the difference in gains between the groups. To do so, we calculated the mean TLU for each group using the TLU for each participant and implemented a Wilcoxon test or a Mann-Whitney U test to reveal significant differences in global accuracy in the use of the possessive determiners.[3] Results were also split by the context (i.e. gender-matched and gender-mismatched) but only descriptive statistics are reported due to the low number of instances produced in some cases, which prevents the use of inferential analyses.

To answer the second research question, which focused on the interplay between language-analytic ability and FFI, Spearman's rho was calculated in each group to

test the association between students' language-analytic ability and the gains they obtained in global accuracy measured via their differences in TLU between dictogloss 1 and 3. Associations between participants' language-analytic ability and their gains in both gender-matched and gender-mismatched contexts were analysed by visual inspection of the scatterplots and identification of trends as the number of data points was inadequate for inferential statistics.

5. Results

The first analysis aimed at answering the first research question by exploring the effectiveness of ME combined with collaborative dictogloss tasks on learners' accuracy in the use of the possessive determiners *his/her* as compared to the combination of the same tasks with self-correction (see Tables 2 to 4). Note that accuracy can only be measured in those pairs that produced OCs, and that the TLU was calculated for each pair that produced OCs and then averaged out to obtain a mean TLU for each group in dictogloss 1 and in dictogloss 3, rather than aggregating the data from all pairs yielding one single TLU for each group with no SD. Doing so allows for a more accurate representation of each pair's performance and their variability within each group.

Table 2: Global descriptives of obligatory contexts (OCs), correct provisions of the target feature, overproduction and target-like use (TLU) for dictogloss 1 and 3 in both groups

		Dictogloss+self-correction group (n=7 pairs)		Dictogloss+ME group (n=9 pairs)	
		Dictogloss 1	Dictogloss 3	Dictogloss 1	Dictogloss 3
OCs (N)		21 (7 pairs)	23 (7 pairs)	26 (7 pairs)	24 (6 pairs)
Correctly produced (N)		13	10	14	18
Overproduction (N)		0	0	0	0
TLU (%)	Mean	70.2	44.0	64.3	90.6
	SD	22.5	14.2	26.2	10.4
	Min	50	25.0	33.3	80
	Max	100	66.7	100	100
	Range	50	41.7	66.7	20

Table 3: Descriptives of obligatory contexts (OCs), correct provisions of the target feature, overproduction and target-like use (TLU) in gender-matched contexts for dictogloss 1 and 3 in both groups

		Dictogloss+self-correction group (n=7 pairs)		Dictogloss+ME group (n=9 pairs)	
		Dictogloss 1	Dictogloss 3	Dictogloss 1	Dictogloss 3
OCs (N)		11 (5 pairs)	18 (7 pairs)	10 (5 pairs)	10 (5 pairs)
Correctly produced (N)		8	7	9	8
Overproduction (N)		0	0	0	0
TLU (%)	Mean	90.0	36.9	80.0	88.3
	SD	22.4	34.6	44.7	16.2
	Min	50	0	0	66.7
	Max	100	100	100	100
	Range	50	100	100	33.3

Table 4: Descriptives of obligatory contexts (OCs), correct provisions of the target feature, overproduction and target-like use (TLU) in gender-mismatched contexts for dictogloss 1 and 3 in both groups

		Dictogloss+self-correction group (n=7 pairs)		Dictogloss+ME group (n=9 pairs)	
		Dictogloss 1	Dictogloss 3	Dictogloss 1	Dictogloss 3
OCs (N)		10 (6 pairs)	5 (4 pairs)	16 (6 pairs)	14 (6 pairs)
Correctly produced (N)		5	3	9	13
Overproduction (N)		0	0	0	0
TLU (%)	Mean	50.0	75.0	52.8	95.8
	SD	45.9	50.0	32.3	10.2
	Min	0	0	0	75
	Max	100	100	100	100
	Range	100	100	100	25

In the dictogloss+self-correction group, all 7 pairs produced OCs in dictogloss 1 (M=3.0, SD=1.83) and in dictogloss 3 (M=3.29, SD=1.38). As regards their accuracy in the use of the possessive determiners, displayed in Table 2, students achieved a TLU of 70.2% in dictogloss 1 (SD=0.225), which fell down to 44.0% in dictogloss 3 (SD=0.142). A Wilcoxon's signed-rank test shows that the scores in post-test are significantly lower (z=-2.032, p=0.042), with a large effect size (d=1.294).

When the context of the possessive determiners is considered, some differences emerge between gender-matched and gender-mismatched contexts in both the production of OCs and the accuracy of use. In gender-matched contexts (Table 3), 5 pairs produced OCs (M=2.20, SD=2.17) with an average accuracy of 90.0% (SD=0.224) in dictogloss 1, whereas all 7 pairs produced OCs in dictogloss 3 (M=2.57, SD=1.9), with only a 36.9% accuracy (SD=0.346). In gender-mismatched contexts (Table 4), 6 pairs produced OCs in dictogloss 1 (M=1.67, SD=1.03) with an average accuracy of 50% (SD=0.459), whereas only 4 pairs produced OCs in dictogloss 3 (M=1.25, SD=0.5), with a TLU of 75% (SD=0.5). More pairs produced OCs in dictogloss 3 than in dictogloss 1 in matched contexts, but the opposite trend is observed in mismatched contexts. Moreover, accuracy decreased in gender-matched contexts and increased in gender-mismatched contexts.

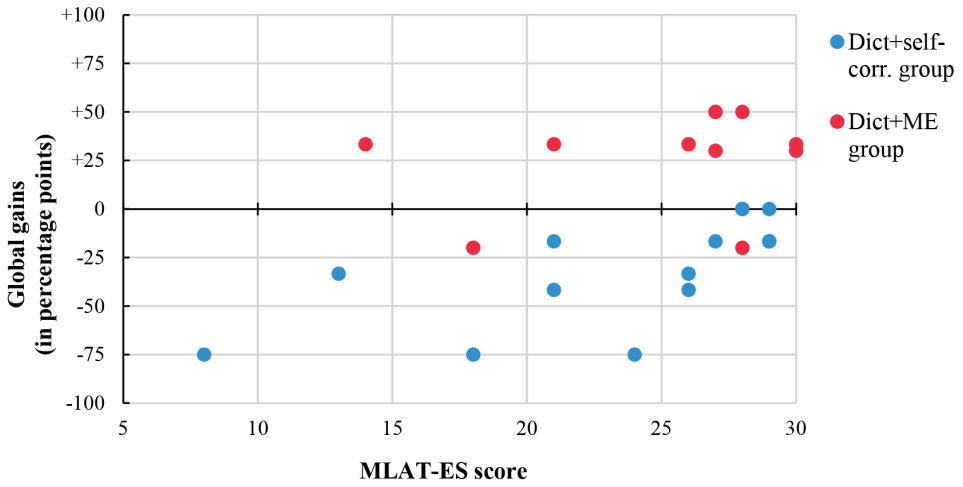
In the dictogloss+ME group, 7 out of 9 pairs produced OCs in dictogloss 1 (M=3.71, SD=2.56) and 6 pairs in dictogloss 3 (M=4.0, SD=1.55). As regards their accuracy in the use of the possessive determiners, displayed in Table 2, students achieved an average TLU of 64.3% in dictogloss 1 (SD=0.262). Although a Wilcoxon's signed-rank test shows that the post-test scores were not significantly higher than the pre-test scores (z=-1.761, p=0.078), accuracy rose to 90.6% in dictogloss 3 (SD=0.104), a result that could be reinforced by the existence of a large effect size (d=1.341).

Some differences can also be observed in the dictogloss+ME group when the context of the possessive determiners is considered. In gender-matched contexts (Table 3), 5 out of the 7 pairs produced OCs (M=2.0, SD=2.24) with an average accuracy of 80.0% (SD=0.447), and 5 pairs produced OCs in dictogloss 3 (M=2.0, SD=1.41), with an 88.3% accuracy (SD=0.162). In gender-mismatched contexts (Table 4), 6 pairs produced OCs in dictogloss 1 (M=2.67, SD=0.82) with an average accuracy of 52.8% (SD=0.323), and 6 pairs produced OCs in dictogloss 3 (M=2.33, SD=1.03), with a TLU of 95.8% (SD=0.102). Contrary to the dictogloss+self-correction group, there was no increase or decrease in the number of pairs that produced OCs in any of the contexts in the dictogloss+ME group. This was accompanied by a slight improvement in accuracy in gender-matched contexts and a more prominent increase in accuracy in mismatched contexts, achieving highly accurate production of the possessive determiners in both contexts.

As regards the impact of the metalinguistic explanations on the gains in the use of the possessive determiners *his/her*, a Mann-Whitney U test on the gains for the pairs that had produced OCs in *both* dictogloss shows that the gains in the dictogloss+ME group are significantly greater than the gains in the dictogloss+self-correction group (z=2.204, p=0.03, d=1.635), with a large effect size. In particular, the pairs in the dictogloss+self-correction group that produced OCs in both dictogloss tasks had an average gain of -26.2 percentage points (SD=0.265), whereas the pairs that received the ME treatment gained on average +25.3 percentage points (SD=0.265).

When we break it down considering the context of the possessive determiners, we can see that the average gains in the dictogloss+self-correction group were -53.3 (SD=0.361) percentage points in matched contexts and +11.1 (SD=0.839) in mismatched contexts. In the dictogloss+ME group, the average gains were +22.2 (SD=0.385) in matched contexts and +31.7 (SD=0.325) in mismatched contexts. In the dictogloss+self-correction group, negative gains are observed in matched contexts and positive gains in mismatched contexts, although these are lower than the dictogloss+ME group, where gains are always positive as well as more stable across contexts.

Graph 1: Scatterplot with the individual scores of the MLAT-ES in the horizontal axis, and the gains from dictogloss 1 to dictogloss 3 (irrespective of the context) in the vertical axis

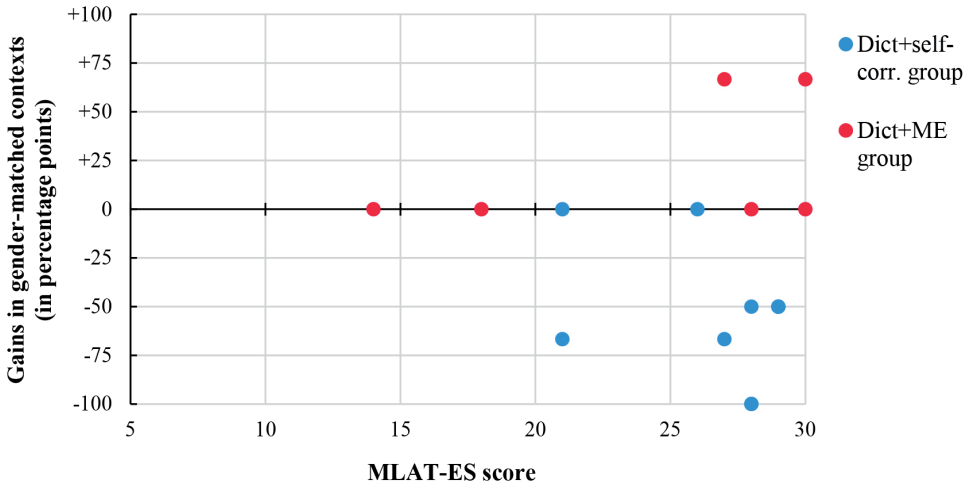


The second research question probed into the role of language-analytic ability in participants' gains in accuracy (or lack thereof). The Spearman's rho test reveals a strong and positive association between language-analytic ability and the gains

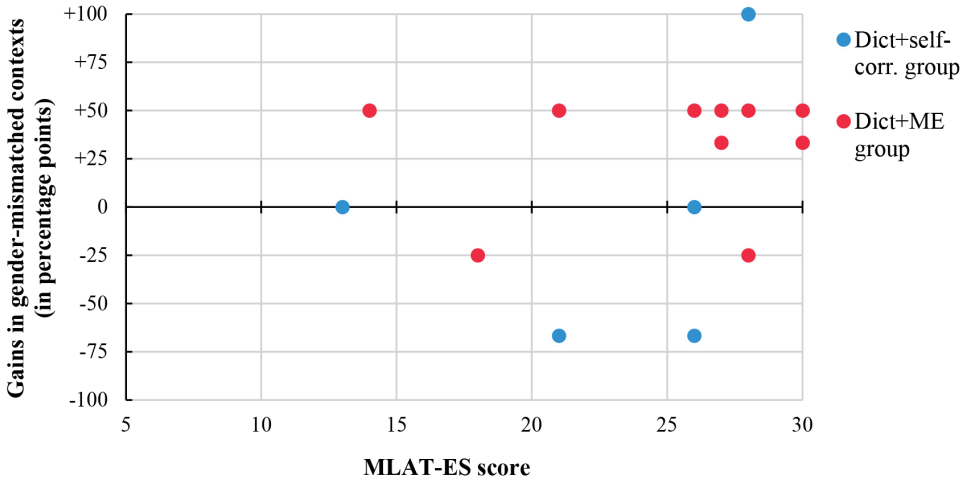
obtained in the dictogloss+self-correction group ($r_s(13)=0.755$, $p=0.001$), whereas no such association is found in the dictogloss+ME group ($r_s(8)=0.019$, $p=0.958$). As can be seen in Graph 1, participants in the dictogloss+self-correction group had smaller losses when their language-analytic ability was greater. In contrast, participants in the dictogloss+ME group generally obtained positive gains regardless of their language-analytic ability.

These trends seem to be driven by the influence of language-analytic ability on students' awareness of mismatched contexts. As shown in Graphs 2 and 3, we can observe that language-analytic ability seems to have a lesser impact in gender-matched contexts as gains are stable in both groups, although mostly negative in the dictogloss+self-correction group and mostly positive in the experimental group. In the mismatched context, however, gains remain stable and positive in the dictogloss+ME group, but they seem to follow a slightly positive association with their language-analytic ability in the dictogloss+self-correction group, although data points in this context are scattered and scarce. In the dictogloss+self-correction group, language-analytic ability seems to have some influence on accuracy in gender-mismatched contexts, or at least in their increased awareness of possessive determiners in this context due to the task.

Graph 2: Scatterplot with the individual scores of the MLAT-ES in the horizontal axis, and the gains from dictogloss 1 to dictogloss 3 in matched contexts in the vertical axis



Graph 3: Scatterplot with the individual scores of the MLAT-ES in the horizontal axis, and the gains from dictogloss 1 to dictogloss 3 in gender-mismatched contexts in the vertical axis



6. Discussion

In this section, results will be discussed according to the two research questions addressed. With respect to the first research question (*Does FFI in the form of metalinguistic explanations after the performance of a focused task lead to greater accuracy in the use of the target feature selected in reconstructed texts than self-correction after the performance of a focused task?*), the selected dictogloss task has been effective in drawing learners' attention to the possessive determiners in matched and mismatched contexts, the target features selected that have been found to pose difficulty for students with Romance L1s (White et al., 2007). Dictogloss tasks are focused tasks, characterised as explicit (Benati, 2021), which have proved efficient in drawing learners' attention to formal aspects of the language in a variety of contexts, with both adults and children (Alegría de la Colina & García Mayo, 2007; Basterrechea & García Mayo, 2014; Calzada, 2021; Lim & Jacobs, 2001).

Its efficiency, though, has been reported to depend on the type of target selected, yielding better results in the case of features such as the third persons singular *-s* morpheme than in the possessive determiners *his/her* (see Calzada, 2021).[4] In this regard, the fact that a greater number of targets for the different conditions were included in the three dictogloss tasks administered, the use of keywords rather than pictures to help the learners in their understanding of the text and to foster greater

attention to the language instead of the content, and the control for text length and vocabulary, allowing for more focused discussions, may have contributed to increase learners' attention to these features and to promote the production of OCs. In other words, the modifications introduced to solve the limitations observed in prior research, that were pilot-tested before the study began, have made the dictogloss an effective instrument for the elicitation of the possessive determiners *his/her*.[5]

The dictogloss task has been specifically effective in gender-mismatched conditions in both groups, which entail greater difficulty to learners with Romance L1s (White et al., 2007). In the case of the dictogloss+self-correction group, just performing the task without having received metalinguistic explanations can raise these learners' awareness of gender-mismatched contexts (see also Li, Ellis, & Kim, 2018), generating greater accuracy in this condition than in gender-matched contexts, where losses were observed. In this last respect, a restructuring process probably induced by the hypothesis-testing of language forms during the reconstruction phase may have taken place (Izumi, 2003; Swain, 1997; Swain & Lapkin, 1995), destabilising the possessive determiners in gender-matched contexts that generally seem to entail less difficulty for L1 Spanish speakers. In contrast, ME increases learners' awareness in both contexts, leading to an upsurge in accuracy in both gender-mismatched and gender-matched conditions and to an integration of both language forms. The mini-lessons of noticing, awareness and practice provided to our participants revolved around gender-matched and gender-mismatched contexts, and occurred after their production, preparing them for the next output-producing session. In other words, learners could have resorted to the knowledge gained during the ME sessions in their subsequent metalinguistic reflection (Swain, 1995) while reconstructing the texts, contributing to greater accuracy in the use of possessive determiners during production (Johns, 2003; Mitchell, 2000; Thepseenu & Roehr, 2013) than the participants in the self-correction group.

In this case, gains for the ME group could be considered robust enough in the light of the large effect size found when being compared to those of the self-correction group. Thus, unlike Gorman and Ellis (2019), explicit FFI has been found to be effective in the case of young learners for a feature not novel to them but that entailed learning difficulty due to the interaction between crosslinguistic influence and the developmental route for possessive determiners. We tentatively argue that adding a more engaging treatment with a noticing and an awareness phase followed by practice, giving precise instructions to focus on form, and asking them to perform the task collaboratively could foster the effectiveness of this combination of explicit forms of FFI.

Likewise, in contrast to other investigations in which explicit FFI has been found to be effective, low accuracy scores notwithstanding (below 50%) (e.g., Calzada, 2021; Serrano, 2011), in the present study, accuracy rose to 90% on average in the third

dictogloss task in the dictogloss+ME group. Thus, the addition of ME in the form of noticing and awareness sessions followed by practice proves as a better option to maximise the learning potential of these focused tasks than those employed in prior studies, a result reinforced by the examination of language development via a grammaticality judgement task administered to the same participants two weeks after the intervention. In particular, the group receiving ME has been found to show an average improvement of 12 points in the grammaticality judgement task when the global gains from the pre-test to the post-test were considered (Dictogloss+self-correction = +1.2; Dictogloss+ME = +12.3), as well as when examining gender-matched (Dictogloss+self-correction = +2.3; Dictogloss+ME = +11.3) and gender-mismatched contexts (Dictogloss+self-correction = +0.0; Dictogloss+ME = +13.2) separately (see Iglesias-Diéguez & Martínez-Adrián, in press). Further analyses await to determine whether the explicit knowledge gained thanks to the treatment received will last over time, considering the short duration of the treatment and the fact that explicit knowledge may be acquired faster but atrophies easily (Shintani & Eltis, 2013). This line of research will also respond to the call made in the literature by different authors (Calzada & García Mayo, 2020; Roehr-Brackin, 2018) for more studies on the durability of the benefits of FFI treatments that incorporate delayed post-tests. In this regard, recent studies conducted with children have unveiled benefits of FFI treatments in delayed post-tests (Li et al., 2018).

With respect to the second research question (*How do different forms of FFI (dictogloss+metalinguistic explanations; dictogloss+self-correction) mediate the relationship between language-analytic ability and use of the target feature selected?*), in the absence of a more guided treatment (i.e. dictogloss+self-correction group), this ID was found to be relevant, those students with a higher language-analytic ability attaining smaller losses (Roehr-Brackin & Tellier, 2019). In contrast, the more explicit the treatment received, the less relevant this ID was. In other words, learners with lower language-analytic ability may compensate their deficits thanks to the ME treatment received, a result in line with prior research conducted with both adults (Suzuki, 2019; Suzuki et al., 2022) and adolescents (J. White & Ranta, 2002). Further studies await to see whether this compensation pattern could be ascribed to an increase in the learners' language-analytic ability thanks to the treatment received, taking into account that this ID is not stable and could develop under FFI conditions (Roehr-Brackin & Tellier, 2019). Despite so, the present investigation has contributed to aptitude-treatment interaction research, a line of study that has started to thrive in the last years (DeKeyser, 2019; Suzuki, 2022). These results also seem to align with prior research in which this ID was more relevant with implicit types of FFI and less relevant with more explicit treatments (Li, 2013), all of which makes a step further as to "when and how L2 grammar instruction should be provided to learners with different strengths and weaknesses profiles of their cognitive abilities" (Suzuki, 2022, p. 297).

7. Conclusion

This paper has contributed to shed more light on the implementation of FFI in young learners in input-limited settings. More specifically, it has shown how FFI consisting of ME after the performance of a dictogloss task including a noticing and an awareness phase followed by practice along Bouffard and Sarkar (2008), and Lyster (2007, 2015) has led to greater awareness and increased accuracy in the use of the possessive determiners *his/her*. Likewise, results seem to indicate that the benefits of the FFI treatment received are independent of participants' language-analytic ability, blurring individual differences in this respect.[6]

For further research, it would be convenient to gather data from a bigger sample and to conduct the study in other learning contexts so as to increase the generalisability of the results obtained. Likewise, even if the combination of dictogloss tasks+ME has been effective in the promotion of accuracy, measures that could lead to a higher production of OCs in the case of possessive determiners would be worth exploring. Unlike the third person singular *-s* morpheme, features such as articles and possessive determiners have been found to be more difficult to elicit via focused tasks, as reported in previous research (Calzada & García Mayo, 2020, 2023; Collins & White, 2014). Adding a pre-task stage on family and friends' relationships could boost the task potential to elicit a greater number of forms and, in turn, to strengthen the value of the production data obtained. Similarly, considering that language-analytic ability is an *individual* difference, future studies including collaborative tasks could control for this variable by matching dyad members in terms of the scores obtained in the MLAT-ES test.[7] However, it should also be borne in mind that some studies have suggested that, regardless of the matching criterion, collaborative tasks can still be beneficial to young learners in a number of ways (Oliver & Bogachenko, 2019; see Oliver, Philp & Duchesne, 2017 for an example) and, thus, students need not be paired up by every single individual difference under study. All in all, since previous research has already unveiled the benefits of having proficiency-matched pairs (Basterrechea & Gallardo-del-Puerto, 2020; García Mayo & Imaz Agirre, 2019), future investigations, like the present study with intact classes, could focus on the effects of individual differences in proficiency-matched pairs as well, possibly in a classroom setting. This will facilitate the implementation of the pedagogical implications by practitioners as such pairings are arguably easier to carry out in a real classroom with the usual tests that students have to complete in class and, therefore, more ecologically valid.

Additionally, the incorporation of tasks tapping explicit and implicit knowledge would result in a more robust investigation of FFI effects in young learners. Qualitative measures such as retrospective interviews that could complement the examination

of LRE production also deserve attention as they could shed more light on covert metalinguistic activity (Gutiérrez, 2013).[8] Finally, longitudinal studies encompassing a wider array of FFI techniques and linguistic features (i.e. novel or already known but posing learning difficulty to the learners) are also desirable.

Despite the existing limitations, two main pedagogical implications are worth mentioning and discussing. First, primary school teachers might be ready to adopt focused tasks such as dictogloss if they want to draw learners' attention to linguistic features that pose greater difficulty for the English learner as for example the possessive determiners *his/her*. We want to stress the potential of these tasks that have already proved effective in adults and adolescent learners (Storch, 2016; Zhang & Plonsky, 2020), and are gaining momentum in the literature on child learners (Ellis, Li, & Zhu, 2019; Calzada & García Mayo, 2020, 2023). In addition, these tasks have been shown to be appealing to young EFL learners (Kopinska & Azkarai, 2020), all of which upholds their feasibility and suitability in this learning context. Furthermore, adding ME in the form of noticing and awareness sessions after the performance of dictogloss tasks in which real examples from the learners' production are shown can maximise the learning potential of these tasks, as they can enhance and induce the attention to linguistic features seeded in the text and enrich learners' metalinguistic reflection during text-reconstruction. In this vein, this is congruent with recent research that has confirmed the efficacy of explicit instruction within task conditions (Michaud & Ammar, 2023; Quinn, 2014), as learners can use the information provided during the treatment sessions in the follow-up tasks. Second, the addition of ME treatments is even more important in this learning context if we consider their potential to benefit students equally, regardless of their language-analytic ability. Thus, not only can the incorporation of this type of treatments promote learning opportunities, but also make them a reality to a greater number of students.

Acknowledgements

We would like to thank Andrea Robin Ruthven for recording the audio for the texts employed in the dictogloss tasks. This work was supported by the Spanish Ministry of Science and Innovation (PID2020-113630-GB-I00), the Basque Government (IT1426-22) and the University of the Basque Country (PIF20/300).

8. References

Akbaş, E., & Ölçü-Dinçer, Z. (2021). Accuracy order in L2 grammatical morphemes: Corpus evidence from different proficiency levels of Turkish learners of English. *Studies in Second Language Learning and Teaching*, *11*(4), 607-627. https://doi.org/10.14746/ssllt.2021.11.4.6

Alegría de la Colina, A., & García Mayo, M. P. (2007). Attention to form across collaborative tasks by low-proficiency learners in an EFL setting. In M. P. García Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 91-116). Bristol: Multilingual Matters. https://doi.org/10.21832/9781853599286-008

Basterrechea, M., & Gallardo-del-Puerto, F. (2020). Language-related episodes and pair dynamics in primary school CLIL learners: A comparison between proficiency-matched and student-selected pairs. *Studies in Second Language Learning and Teaching*, *10*(3), 423-447. https://doi.org/10.14746/ssllt.2020.10.3.2

Basterrechea, M., & García Mayo, M. P. (2014). Dictogloss and the production of the English third person '-s' by CLIL and mainstream EFL learners: A comparative study. *International Journal of English Studies*, *14*(2), 77-98. https://doi.org/10.6018/j.177321

Benati, A. G. (2021). *Focus on form*. New York: Cambridge University Press. https://doi.org/10.1017/9781108762656

Bouffard, L. A., & Sarkar, M. (2008). Training 8-year-old French immersion students in metalinguistic analysis: An innovation in form-focused pedagogy. *Language Awareness*, *17*(1), 3-24. https://doi.org/10.2167/la424.0

Calzada, A. (2021). *Dictogloss in the primary school EFL classroom: Investigating the process, product and perceptions of collaborative writing.* [Unpublished doctoral dissertation, University of the Basque Country [UPV/EHU], Vitoria, Spain].

Calzada, A., & García Mayo, M. P. (2020). Child EFL grammar learning through a collaborative writing task. In W. Suzuki, & N. Storch (Eds.), *Languaging in language learning and teaching: A collection of empirical studies* (pp. 19-40). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.55.01cal

Calzada, A., & García Mayo, M. P. (2023). Do task repetition and pretask focus on form instruction impact collaborative writing performance? Evidence from young learners. In M. Li, & M. Zhang (Eds.), *L2 collaborative writing in diverse learning contexts* (pp. 80-108). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.59.04cal

Cambridge Assessment English. (2018). *Pre A1 Starters, A1 Movers and A2 Flyers: Sample papers for exams from 2018* (Vol. 1). Cambridge: Cambridge University Press.

Carroll, J. B., & Sapon, S. M. (2002). *Modern Language Aptitude Test - Elementary: Manual, 2002 edition*. Rockville, MD: Language Learning and Testing Foundation.

Cenoz, J. (2009). *Towards multilingual education: Basque educational research from an international perspective*. Bristol: Multilingual Matters. https://doi.org/10.21832/9781847691941

Collins, L., & Ruivivar, J. (2020). Form-focused instruction. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1-4). Hoboken, NJ: Wiley. https://doi.org/10.1002/9781405198431.wbeal0432.pub2

Collins, L., & Ruivivar, J. (2021). Research agenda: Researching grammar teaching and learning in the second language classroom. *Language Teaching*, *54*(3), 407-423. https://doi.org/10.1017/S0261444821000070

Collins, L., & White, J. (2014). The quantity and quality of language practice in typical interactive pair/group tasks. *TESL Canada Journal*, *31*(8), 47-67. https://doi.org/10.18806/tesl.v31i0.1186

DeKeyser, R. M. (2019). Aptitude-treatment interaction in second language learning: Introduction to the special issue. *Journal of Second Language Studies*, *2*(2), 165-168. https://doi.org/10.1075/jsls.00007.int

DeKeyser, R. (Ed.). (2021). *Aptitude-treatment interaction in second language learning*. Amsterdam, the Netherlands: John Benjamins. https://doi.org/10.1075/bct.116

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

Ellis, R., & Barkhuizen, G. P. (2005). *Analysing learner language*. Oxford: Oxford University Press.

Ellis, R., Li, S., & Zhu, Y. (2019). The effects of pre-task explicit instruction on the performance of a focused task. *System*, *80*, 38-47. https://doi.org/10.1016/j.system.2018.10.004

Erlam, R. (2013). Effects of instruction on learners' acquisition of metalinguistic knowledge. In K. Roehr, & G. A. Gánem-Gutiérrez (Eds.), *The metalinguistic dimension in instructed second language learning* (pp. 71-94). London: Bloomsbury. https://doi.org/10.5040/9781472542137.ch-004

Ferman, S., & Karni, A. (2010). No childhood advantage in the acquisition of skill in using an artificial language rule. *PLoS ONE*, *5*(10), Article e13648. https://doi.org/10.1371/journal.pone.0013648

Fernández Dobao, A. (2012). Collaborative writing tasks in the L2 classroom: Comparing group, pair, and individual work. *Journal of Second Language Writing*, *21*(1), 40-58. https://doi.org/10.1016/j.jslw.2011.12.002

Fernández Dobao, A., & Blum, A. (2013). Collaborative writing in pairs and small groups: Learners' attitudes and perceptions. *System*, *41*(2), 365-378. https://doi.org/10.1016/j.system.2013.02.002

Fernández Dobao, A. (2020). Collaborative writing in mixed classes: What do heritage and second language learners think? *Foreign Language Annals*, *53*(1), 48-68. https://doi.org/10.1111/flan.12446

File, K. A., & Adams, R. (2010). Should vocabulary instruction be integrated or isolated? *TESOL Quarterly*, *44*(2), 222-249. https://doi.org/10.5054/tq.2010.219943

Frear, D., & Chiu, Y. (2015). The effect of focused and unfocused indirect written corrective feedback on EFL learners' accuracy in new pieces of writing. *System*, *53*, 24-34. https://doi.org/10.1016/j.system.2015.06.006

García Mayo, M. P. (2002). The effectiveness of two form-focused tasks in advanced EFL pedagogy. *International Journal of Applied Linguistics*, *12*(2), 156-175. https://doi.org/10.1111/1473-4192.t01-1-00029

García Mayo, M. P., & Imaz Agirre, A. (2019). Task modality and pair formation method: Their impact on patterns of interaction and LREs among EFL primary school children. *System*, *80*, 165-175. https://doi.org/10.1016/j.system.2018.11.011

Gesa, F., & Suárez, M. del M. (2022). Individual differences and young learners' L2 vocabulary development: The case of language aptitude and exposure to subtitled TV series. In V. De Wilde, & C. Goriot (Eds.), *Second language learning before adulthood: Individual differences in children and adolescents* (pp. 65-86). Berlin: De Gruyter. https://doi.org/10.1515/9783110743043-004

Goo, J., Granena, G., Yilmaz, Y., & Novella, M. (2015). Implicit and explicit instruction in L2 learning: Norris & Ortega (2000) revisited and updated. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 443-482). Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.48.18goo

Gorman, M., & Ellis, R. (2019). The relative effects of metalinguistic explanation and direct written corrective feedback on children's grammatical accuracy in new writing. *Language Teaching for Young Learners*, *1*(1), 57-81. https://doi.org/10.1075/ltyl.00005.gor

Gutiérrez, X. (2013). Metalinguistic knowledge in language-related episodes with covert metalinguistic activity. In K. Roehr, & G. A. Gánem-Gutiérrez (Eds.), *The metalinguistic dimension in instructed second language learning* (pp. 147-169). London: Bloomsbury. https://doi.org/10.5040/9781472542137.ch-007

Hakes, D. T. (1980). *The development of metalinguistic abilities in children*. Berlin: Springer. https://doi.org/10.1007/978-3-642-67761-8

Hanan, R. E. (2015). *The effectiveness of explicit grammar instruction for the young foreign language learner: A classroom-based experimental study*. [Doctoral dissertation, University of York, York, England].

Harley, B. (1998). The role of focus-on-form tasks in promoting child L2 acquisition. In C. J. Doughty, & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 156-173). Cambridge: Cambridge University Press.

Harley, B. (1989). Functional grammar in French immersion: A classroom experiment. *Applied Linguistics*, *10*(3), 331-360. https://doi.org/10.1093/applin/10.3.331

Iglesias-Diéguez, K., & Martínez-Adrián, M. (in press). At the crossroad between form-focused instruction (FFI) and language-analytic ability: A study on metalinguistic explanations and the acquisition of explicit knowledge of the L3 English possessive determiners *his/her*. In M. P. García Mayo (Ed.), *Investigating attention to form and individual differences: Research with EFL children*. Cham: Springer.

Imaz Agirre, A., & García Mayo, M. P. (2013). Gender agreement in L3 English by Basque/Spanish bilinguals. *Linguistic Approaches to Bilingualism*, *3*(4), 415-447. https://doi.org/10.1075/lab.3.4.02ima

Izumi, S. (2003). Comprehension and production processes in second language learning: In search of the psycholinguistic rationale of the output hypothesis. *Applied Linguistics*, *24*(2), 168-196. https://doi.org/10.1093/applin/24.2.168

Johns, A. M. (2003). Genre and ESL/EFL composition instruction. In B. Kroll (Ed.), *Exploring the dynamics of second language writing* (pp. 195-217). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139524810.014

Kachinske, I., & DeKeyser, R. (2019). The interaction between timing of explicit grammar explanation and individual differences in second language acquisition. *Journal of Second Language Studies*, *2*(2), 197-232. https://doi.org/10.1075/jsls.19003.kac

Kim, Y. (2008). The contribution of collaborative and individual tasks to the acquisition of L2 vocabulary. *The Modern Language Journal*, *92*(1), 114-130. https://doi.org/10.1111/j.1540-4781.2008.00690.x

Kiss, C., & Nikolov, M. (2005). Developing, piloting, and validating an instrument to measure young learners' aptitude. *Language Learning*, *55*(1), 99-150. https://doi.org/10.1111/j.0023-8333.2005.00291.x

Kopinska, M., & Azkarai, A. (2020). Exploring young EFL learners' motivation: Individual versus pair work on dictogloss tasks. *Studies in Second Language Learning and Teaching*, *10*(3), 607-630. https://doi.org/10.14746/ssllt.2020.10.3.10

Kuiken, F., & Vedder, I. (2012). Syntactic complexity, lexical variation and accuracy as a function of task complexity and proficiency level in L2 writing and speaking. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 143-169). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.32.07kui

Lado, B., & Sanz, C. (2021). Feedback, aptitude, and multilingualism. In H. Nassaji, & E. Kartchava (Eds.), *The Cambridge handbook of corrective feedback in second language learning and teaching* (pp. 689-712). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108589789.033

Leeming, P., Aubrey, S., & Lambert, C. (2022). Collaborative pre-task planning processes and second language task performance. *RELC Journal*, *53*(3), 534-550. https://doi.org/10.1177/0033688220953225

Li, S. (2013). The interactions between the effects of implicit and explicit feedback and individual differences in language analytic ability and working memory. *The Modern Language Journal*, *97*(3), 634-654. https://doi.org/10.1111/j.1540-4781.2013.12030.x

Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition*, *38*(4), 801-842. https://doi.org/10.1017/S027226311500042X

Li, S., Ellis, R., & Kim, J. (2018). The influence of pre-task grammar instruction on L2 learning: An experimental study. *Studies in English Education*, *23*(4), 831-857. https://doi.org/10.22275/SEE.23.4.03

Li, S., Ellis, R., & Zhu, Y. (2019). The associations between cognitive ability and L2 development under five different instructional conditions. *Applied Psycholinguistics*, *40*(3), 693-722. https://doi.org/10.1017/S0142716418000796

Lichtman, K. (2021). What about fluency? Implicit vs. explicit training affects artificial mini-language production. *Applied Linguistics*, *42*(4), 668-691. https://doi.org/10.1093/applin/amaa054

Lim, W. L., & Jacobs, G. M. (2001). *An analysis of students' dyadic interaction on a dictogloss task*. ERIC (ED456649).

Loewen, S., & Sato, M. (2018). Interaction and instructed second language acquisition. *Language Teaching*, *51*(3), 285-329. https://doi.org/10.1017/S0261444818000125

Lyster, R. (2007). *Learning and teaching languages through content: A counterbalanced approach*. Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.18

Lyster, R. (2015). Using form-focused tasks to integrate language across the immersion curriculum. *System*, *54*, 4-13. https://doi.org/10.1016/j.system.2014.09.022

Martínez-Adrián, M., & Gallardo-del-Puerto, F. (2021). Task modality and language-related episodes in young learners: An attempt to manage accuracy and editing. *Language Teaching Research, 28*(6), 2300-2325. https://doi.org/10.1177/13621688211052808

Michaud, G., & Ammar, A. (2023). Explicit instruction within a task: Before, during, or after? *Studies in Second Language Acquisition*, *45*(2), 442-460. https://doi.org/10.1017/S0272263122000316

Mitchell, R. (2000). Applied linguistics and evidence-based classroom practice: The case of foreign language grammar pedagogy. *Applied Linguistics*, *21*(3), 281-303. https://doi.org/10.1093/applin/21.3.281

Muñoz, C. (2014). The association between aptitude components and language skills in young learners. In M. Pawlak, & L. Aronin (Eds.), *Essential topics in applied linguistics and multilingualism: Studies in honor of David Singleton* (pp. 51-68). Cham: Springer. https://doi.org/10.1007/978-3-319-01414-2_4

Nassaji, H. (2016). Research timeline: Form-focused instruction and second language acquisition. *Language Teaching*, *49*(1), 35-62. https://doi.org/10.1017/S0261444815000403

Norris, J. M., & Ortega, L. (2001). Does type of instruction make a difference? Substantive findings from a meta-analytic review. *Language Learning*, *51*(s1), 157-213. https://doi.org/10.1111/j.1467-1770.2001.tb00017.x

Oliver, R., & Azkarai, A. (2017). Review of child second language acquisition (SLA): Examining theories and research. *Annual Review of Applied Linguistics*, *37*, 62-76. https://doi.org/10.1017/S0267190517000058

Oliver, R., & Bogachenko, T. (2019). Children and tasks. *Language Teaching Research Quarterly*, *12*, 48-65.

Oliver, R., Philp, J., & Duchesne, S. (2017). Children working it out together: A comparison of younger and older learners collaborating in task based interaction. *System*, *69*, 1-14. https://doi.org/10.1016/j.system.2017.08.001

Pawlak, M. (2021). Corrective feedback, developmental readiness, and language proficiency. In H. Nassaji, & E. Kartchava (Eds.), *The Cambridge handbook of corrective feedback in second language learning and teaching* (pp. 733-753). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108589789.035

Philp, J., Walter, S., & Basturkmen, H. (2010). Peer interaction in the foreign language classroom: What factors foster a focus on form? *Language Awareness*, *19*(4), 261-279. https://doi.org/10.1080/09658416.2010.516831

Pica, T. (1983). Adult acquisition of English as a Second Language under different conditions of exposure. *Language Learning*, *33*(4), 465-497. https://doi.org/10.1111/j.1467-1770.1983.tb00945.x

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

Quinn, P. (2014). *Delayed versus immediate corrective feedback on orally produced passive errors in English*. [Doctoral dissertation, University of Toronto, Toronto, Canada].

Roehr, K. (2008). Metalinguistic knowledge and language ability in university-level L2 learners. *Applied Linguistics*, *29*(2), 173-199. https://doi.org/10.1093/applin/amm037

Roehr-Brackin, K. (2018). *Metalinguistic awareness and second language acquisition*. New York: Routledge. https://doi.org/10.4324/9781315661001

Roehr-Brackin, K. (2022). Investigating explicit and implicit L2 knowledge and learning: Replications of Erlam (2005) and Roehr-Brackin and Tellier (2019). *Language Teaching*, *55*(2), 271-283. https://doi.org/10.1017/S0261444820000415

Roehr-Brackin, K. (2024). Explicit and implicit knowledge and learning of an additional language: A research agenda. *Language Teaching*, *57*(1), 68-86. https://doi.org/10.1017/S026144482200026X

Roehr-Brackin, K., & Tellier, A. (2019). The role of language-analytic ability in children's instructed second language learning. *Studies in Second Language Acquisition*, *41*(5), 1111-1131. https://doi.org/10.1017/S0272263119000214

Rosa-González, D. (2011). *The effects of aptitude and motivation on the acquisition of EFL in young learners.* [Master's thesis, University of Barcelona, Barcelona, Spain].

Serrano, R. (2011). From metalinguistic instruction to metalinguistic knowledge, and from metalinguistic knowledge to performance in error correction and oral production tasks. *Language Awareness*, *20*(1), 1-16. https://doi.org/10.1080/09658416.2010.529911

Shintani, N., & Ellis, R. (2013). The comparative effect of direct written corrective feedback and metalinguistic explanation on learners' explicit and implicit knowledge of the English indefinite article. *Journal of Second Language Writing*, *22*(3), 286-306. https://doi.org/10.1016/j.jslw.2013.03.011

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69-93). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.2.06ske

Spada, N. (1997). Form-focussed instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching*, *30*(2), 73-87. https://doi.org/10.1017/S0261444800012799

Spada, N. (2023). Form-focused instruction. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1-4). Hoboken: Wiley. https://doi.org/10.1002/9781405198431.wbeal0432.pub3

Stansfield, C. W., & Reed, D. J. (2005). *Modern Language Aptitude Test - Elementary: Spanish version, 2005 edition*. Rockville, MD: Language Learning and Testing Foundation.

Storch, N. (1999). Are two heads better than one? Pair work and grammatical accuracy. *System*, *27*(3), 363-374. https://doi.org/10.1016/S0346-251X(99)00031-7

Storch, N. (2016). Collaborative writing. In R. M. Manchón, & P. K. Matsuda (Eds.), *Handbook of second and foreign language writing* (pp. 387-406). Berlin: De Gruyter. https://doi.org/10.1515/9781614511335-021

Suzuki, Y. (2019). Individualization of practice distribution in second language grammar learning: The role of metalinguistic rule rehearsal ability and working memory capacity. *Journal of Second Language Studies*, *2*(2), 169-196. https://doi.org/10.1075/jsls.18023.suz

Suzuki, Y. (2022). Individual difference factors for second language grammar. In S. Li, P. Hiver, & M. Papi (Eds.), *The Routledge handbook of second language acquisition and individual differences* (pp. 294-309). New York: Routledge. https://doi.org/10.4324/9781003270546-25

Suzuki, Y., & DeKeyser, R. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning*, *67*(4), 747-790. https://doi.org/10.1111/lang.12241

Suzuki, Y., Yokosawa, S., & Aline, D. (2022). The role of working memory in blocked and interleaved grammar practice: Proceduralization of L2 syntax. *Language Teaching Research*, *26*(4), 671-695. https://doi.org/10.1177/1362168820913985

Swain, M. (1995). Three functions of output in second language learning. In G. Cook, & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 125-144). Oxford: Oxford University Press.

Swain, M. (1997). Collaborative dialogue: Its contribution to second language learning. *Revista Canaria de Estudios Ingleses*, *34*, 115-132.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, *16*(3), 371-391. https://doi.org/10.1093/applin/16.3.371

Tellier, A., & Roehr-Brackin, K. (2013). The development of language learning aptitude and metalinguistic awareness in primary-school children: A classroom study. *Essex Research Reports in Linguistics*, *62*(1), 1-28.

The Douglas Fir Group. (2016). A transdisciplinary framework for SLA in a multilingual world. *The Modern Language Journal*, *100*(S1), 19-47. https://doi.org/10.1111/modl.12301

Thepseenu, B., & Roehr, K. (2013). University-level learners' beliefs about metalinguistic knowledge. In K. Roehr, & G. A. Gánem-Gutiérrez (Eds.), *The metalinguistic dimension in instructed second language learning* (pp. 95-118). London: Bloomsbury. https://doi.org/10.5040/9781472542137.ch-005

Vraciu, A. (2020). Production accuracy of verb morphology in early EFL: Does primary school CLIL make a difference? *Linx*, *81*, 1-26. https://doi.org/10.4000/linx.7112

Wajnryb, R. (1990). *Resource book for teachers: Grammar dictation*. Oxford, England: Oxford University Press.

White, J., Muñoz, C., & Collins, L. (2007). The *his/her* challenge: Making progress in a 'regular' L2 programme. *Language Awareness*, *16*(4), 278-299. https://doi.org/10.2167/la419.0

White, J., & Ranta, L. (2002). Examining the interface between metalinguistic task performance and oral production in a Second Language. *Language Awareness*, *11*(4), 259-290. https://doi.org/10.1080/09658410208667060

White, L., Spada, N., Lightbown, P. M., & Ranta, L. (1991). Input enhancement and L2 question formation. *Applied Linguistics*, *12*(4), 416-432. https://doi.org/10.1093/applin/12.4.416

Xu, J., & Li, C. (2021). The effects of the timing of form-focused instruction on EFL learners' learning of difficult and easy grammatical features: A comparative study. *System*, *101*, Article 102612. https://doi.org/10.1016/j.system.2021.102612

Zhang, M., & Plonsky, L. (2020). Collaborative writing in face-to-face settings: A substantive and methodological review. *Journal of Second Language Writing*, *49*, Article 100753. https://doi.org/10.1016/j.jslw.2020.100753

Zimmerman, D. W., & Williams, R. H. (1982). Gain scores in research can be highly reliable. *Journal of Educational Measurement*, *19*(2), 149-154. https://doi.org/10.1111/j.1745-3984.1982.tb00124.x

Appendix 1 (Dictogloss 2)

Victor and Janire's family is on holiday in France. Victor speaks French very well. His father is French and his mum also speaks French. For Janire, Victor's sister, it is difficult to speak French. She constantly asks her brother to help her with the language. In the mornings, Victor goes surfing with his uncle and some new friends. Janire prefers riding horses so she rides with her aunt every day. In the afternoons, Victor likes to fish with his grandmother in their new boat. Janire prefers driving the boat with her mum. And in the evenings, her grandfather prepares a nice dinner in the boat for everyone!

[1] As conceived by Roehr (2008), metalinguistic knowledge can be defined as "a learner's explicit knowledge about the syntactic, morphological, lexical, phonological, and pragmatic features of the L2" (p. 179). It can be measured through metalinguistic tasks (e.g. grammaticality judgements, description and explanation tasks). It has also been investigated by examining learners' interaction while performing collaborative tasks in a L2 (operationalised through LRE production and accurate use of the target features selected) (Gutiérrez, 2013; Thepseenu & Roehr, 2013). As part of a bigger project on the effect of different forms of FFI treatments after dictogloss tasks on possessive determiners *his/her* and the third person singular *-s* morpheme (ME for the former and written corrective feedback for the latter), the present study will tackle the dimension of accuracy in the use of possessive determiners in written production, while the performance of a metalinguistic task is reported in Iglesias-Diéguez & Martínez-Adrián (in press) and LRE production will be explored in a forthcoming investigation.

[2] Note that this measure is also being used in the case of the studies within the same project examining FFI in the form of written corrective feedback and the third person singular *-s* morpheme.

[3] Notice that, as in previous studies analysing production data, the percentages obtained for each participant rely on a different number of OCs produced since these cannot be controlled for and depend on students' own production (e.g. Calzada & García Mayo, 2023).

[4] Note also that other focused tasks did not work as planned in other investigations with the possessive determiners *his/her* (Collins & White, 2014).

[5] The same dictogloss tasks have also proved effective in another forthcoming study conducted with a different sample in which both an experimental group receiving pre-task grammar instruction and a task-only group significantly improved from a pre- to a post-test phase.

[6] As noted by an anonymous reviewer, bilingual children, such as those that have taken part in this study, may have greater metalinguistic awareness than monolingual children. However, this is unlikely to have had any impact on the better results obtained by the group with ME as participants in both groups are bilingual. Note also that the blurring effect observed in this study has also been observed with Japanese monolingual adults (Suzuki et al., 2022).

[7] Even if future investigations should control for similar levels of MLAT-ES, it is worth mentioning that in a study with the same participants in which an *individual* grammaticality judgement task was incorporated, the trend observed in the current study as regards the blurring effect of ME has also been attested (see Iglesias-Diéguez & Martínez-Adrián, in press).

[8] Qualitative methodologies would have added further validity to the study, but as part of a bigger project, there were additional instruments that already took much time.

Navigating languages and love: Exploring the perceived emotional weight of *I love you* among Third Culture Kids in Denmark ——

Lydia Rodríguez-Bernal
Department of Culture and Language
University of Southern Denmark
lrb@sdu.dk

Abstract

This mixed-methods study contributes to understanding the emotional dynamics of a distinct, yet understudied, multilingual population: Third Culture Kids (TCKs). These individuals are typically raised in multilingual and nomadic households and are extensively socialised in English through international schools and expatriate communities. While prior research has focused on the emotional nuances of the phrase *I love you* in adult populations, a significant gap remains in understanding these nuances in other age groups, particularly in the context of transnational upbringing. This study addresses this gap by exploring the perceived emotional weight of the phrase in the linguistic repertoire of 276 TCKs enrolled in the International Baccalaureate Program by examining the emotional dynamics shaped by their first language (L1,mainly a heritage language) and their frequent exposure to English (LX, acquired after L1) in an environment influenced by American culture, where the use of *I love you* is a daily phenomenon. Unlike previous studies, results indicated a combined heightened emotional weight of L1 and LX, challenging the traditional centrality of L1. Furthermore, multinomial logistic regression models revealed TCKs' unique sociobiographical factors as the driving force behind their emotional perception. These findings contribute to understanding how sociobiographical features influence linguistic emotional perceptions and cross-cultural communication.

Keywords: Third Culture Kids; multilingualism; emotions; love; emotional weight.

Resumen

Este estudio de métodos mixtos contribuye a la comprensión de la dinámica emocional en una población multilingüe única y poco estudiada: los Niños de Tercera Cultura (NTCs), que suelen criarse en hogares multilingües y nómadas y son socializados en inglés a través de escuelas internacionales y comunidades de expatriados. Aunque investigaciones previas han explorado los matices emocionales de la frase "I love you" en adultos multilingües, este estudio se distingue al enfocarse en adolescentes. Además, amplía su enfoque para abarcar el fenómeno del translenguaje dentro del contexto influyente de la crianza transnacional. El estudio explora la intensidad emocional percibida de la frase en 276 NTCs en el Programa del Bachillerato Internacional en Dinamarca, examinando su primera lengua (L1, principalmente una lengua de herencia) y la exposición frecuente al inglés (LX, adquirida después de la L1). En contraste con estudios anteriores, los resultados un peso emocional combinado y elevado de L1 y LX, desafiando la centralidad tradicional de L1. Los modelos de regresión logística revelaron que los factores sociobiográficos únicos en los NTC son los principales impulsores de la percepción emocional. Estos hallazgos contribuyen a entender cómo los rasgos sociobiográficos influyen en la percepción emocional lingüística y en la comunicación intercultural.

Palabras clave: Niños de Tercera Cultura; multilingüismo; emociones; amor; peso emocional.

1. Introduction

Research on multilinguals and emotional perception has primarily compared the emotional resonance of the first language (L1) — generally (but not always) the language that elicits the strongest emotional response — and LX, any subsequent language (Dewaele, 2018) often regarded as emotionally distant. Factors influencing the emotional resonance shift from L1 to LX include early LX age of onset of acquisition (AoA), naturalistic LX context of acquisition (CoA), high LX proficiency, frequent use of LX, and extended and intense LX socialisation (e.g. Dewaele, 2008).

Research on the emotional impact of *I love you* has primarily focused on adult English bi-multilinguals in immersion/non-immersion settings, considering sociobiographical factors, such as age, gender, and education level. However, there is a gap in understanding early multilingual development and the enduring effects of transnational upbringing. Furthermore, exploring diverse forms of multilingualism, such as those from intercultural households (i.e. parents with different L1s) in expatriate communities where English is the lingua franca but reside in countries with different languages, requires further investigation.

Hence, this mixed-methods study aims to offer new insights into emotions in multilinguals by exploring the perceived emotional weight of *I love you* in a unique population: Third Culture Kids (TCKs). TCKs, raised in multilingual households and frequently relocating due to parental career choices, navigate the expatriate community (third culture), between their parent(s') and the host country's culture (Pollock et al., 2017), while attending international schools worldwide.

Love, enduring unlike other emotions (Bowers et al., 1994), is deeply influenced by one's linguistic and cultural background. Examining *love* within this cohort presents an opportunity to investigate the interplay between language and emotions over time, considering the past (linguistic history), present (linguistic competence), and various sociobiographical and sociolinguistic factors. This temporal approach enriches comprehension of how TCKs navigate emotions across languages during their developmental years.

2. Background

2.1. Theoretical framework

Love, a multifaceted and socially constructed phenomenon (Pavlenko, 2012), shapes relationships from birth to adulthood, embracing both universal and culturally specific societal norms, values, and expectations (Gareis & Wilkins, 2011). To understand the emotional impact of *I love you* on multilinguals, it is essential to consider the individual perspectives on love and how language preferences shape expression and perception. Thus, while this study focuses on verbal perception, it acknowledges its interdependence with expression.

Given that socialisation is crucial for TCKs, this study explores the family and friend dynamics in bi-multilingual contexts, where both L1(s) (mainly a heritage language, HL) and LX English (L2 in this sample) are used.[1] Grounded in the broader translanguaging framework, it considers the entire linguistic repertoire as an integrated entity, reflecting the versatile language practices of multilinguals (Li, 2018), and the socially constructed reality of TCKs (García & Li, 2014).

This social aspect is complemented by the cognitive and emotional-experiential dimensions of language acquisition.

The cognitive dimension follows the Complementary Principle (Grosjean, 1997), which characterises bi-multilinguals as those using multiple languages "for different purposes, in different domains of life, with different people" (p. 165), resulting in a language mode continuum (Grosjean, 1985) from monolingual (only one language

used) to bi-multilingual speech mode (two or more languages used). TCKs typically display uneven language distributions, with L1(s) predominantly used within the family and English extensively used across multiple domains. L1 acquisition is usually naturalistic (i.e. outside school), driven by parental input (Curdt-Christiansen, 2022), leading to skill asymmetries. Conversely, English (LX) dominates TCKs' daily lives at the international school, as the language of instruction, and during interactions with peers and friends within the expatriate community. TCKs' language processing spans from monolingual, common in households with a restrictive Family Language Policy (FLP), to bi-multilingual in households embracing a more flexible approach to language use. Consequently, TCKs are expected to emotionally blend their home language(s) with their English usage.

The emotional-experiential dimension adheres to Pavlenko's Theory of Language Embodiment (2005), placing TCKs' languages along a continuum from the "primary language acquisition" (typically emotional and context-bound) to the "foreign language acquisition" (typically devoid of emotion and context). Pavlenko attributes the heightened emotional significance of L1 to *conceptual development* (building emotional categories by integrating sensory information refined through socialisation) and *affective linguistic conditioning* (emotional meanings in words stem from connections to charged memories and experiences), rendering L1 an embodied language (Pavlenko, 2005). Conversely, foreign language learning often lacks emotional responses due to limited limbic system involvement (i.e. disembodied language). This study anticipates that English LX, acquired later in life and predominantly in a mixed context — naturalistic through peer interaction in the international school and the expatriate community, and instructed as the medium to deliver the International Baccalaureate (IB) curriculum — may align closer to L1 on the continuum due to its immersive exposure, potentially leading to an embodied perception among TCKs (Rodríguez-Bernal et al., 2023). Consequently, English is expected to actively shape TCKs' perception of *I love you*.

2.2. "I love you" in bi-multilinguals: Assessment clusters

Studies examining the emotional significance of the phrase *I love you* in both L1 and LX have yielded varying findings. To our knowledge, only three empirical studies have explored this expression: one involving a multilingual user base (Dewaele, 2008) and the other two involving English bilinguals from specific L1s (Jahangard & Holderread, 2013; Ożańska-Ponikwia, 2016). L1 consistently held greater emotional weight across all studies despite differing research contexts.

This research categorised factors influencing emotional differences between L1-LX into three clusters: sociobiographical, language profile, and sociolinguistic.

2.2.1. Sociobiographical

Previous research has not identified sociobiographical variables, such as gender, age, or education level to significantly impact the emotional perception of *I love you*. Studies conducted by Dewaele (2008) and Jahangard and Holderread (2013) involving 1459 adult multilinguals and 20 adult Irian English bilinguals, respectively, did not reveal any significant results. Nevertheless, Ożańska-Ponikwia's (2016) study of 72 adult Polish English bilinguals using correlations and stepwise regression found *length of immersion* in an English-speaking country to be a significant factor influencing LX emotional perception of the phrase. Socialisation in LX culture and the degree of LX use were strong predictors of LX emotional expression. Other studies have suggested that immersion can alter emotional perception of LX over time (e.g. Dewaele, 2011; Ożańska-Ponikwia, 2014), leading to *emotional acculturation* (De Leersnyder, 2017).

2.2.2. Language profile

Several studies have found that AoA, CoA, self-reported language dominance, and proficiency in LX influence emotionality in LX (e.g. Dewaele, 2010; Pavlenko, 2005, 2008). Specifically, in studies examining the perceived emotional weight of *I love you,* only self-reported language dominance consistently emerged as significant factor (Dewaele, 2008; Jahangard & Holderread, 2013; Ożańska-Ponikwia, 2016), whereas AoA, CoA, and self-reported oral proficiency only had marginal effects (Dewaele, 2008).

Research examining AoA typically differentiates between simultaneous and sequential bilinguals, with sequential bilinguals showing lower emotionality in later-acquired languages (Pavlenko, 2005). L1(s) is often perceived as more emotionally resonant, while earlier LX AoA aligns closely with L1 emotional perception (Harris, 2004). In fact, an early LX AoA has been found to act as a proxy for a more emotionally charged CoA. The "emotional contexts of learning" hypothesis (Harris et al., 2006) posits that the emotional richness of the learning context influences perceived emotionality. L1 is typically acquired in emotionally rich settings (outside school and at home through social interactions with caregivers), enhancing its emotional charge. However, LX(s) CoA varies, including both emotional and non-emotional contexts (i.e. purely instructed), potentially limiting emotional experiences (Jahangard & Holderread, 2013).

Language proficiency is necessary to accurately discern emotions (Dewaele & Nakano, 2013), with higher proficiency generally correlating with greater perception of emotional significance (Caldwell-Harris et al., 2011). Nevertheless, research on the influence of language proficiency on emotions has yielded diverse outcomes.

Dewaele's (2011) study involving 386 bi-multilingual adults maximally proficient in their L1 and LX revealed a preference for using L1 to communicate emotions, emphasising the context-driven aspect of language use. Other studies have highlighted the linguistic profile of LX participants and accessibility to LX in their upbringing context, further attenuating the overall impact of language proficiency (Dewaele et al., 2021) on perceived emotional intensity. Moreover, recent research suggests that high proficiency may not necessarily correlate with high perceived emotional intensity (e.g. Lorette & Dewaele, 2022), indicating that mastery is less important than the ability to process information. In fact, a novel scale developed to measure emotional resonance in LX compared to L1 in bi-multilinguals, separates proficiency from emotional resonance, recognising the relationship's dependence on factors like frequency of use and cultural exposure to LX (Toivo et al., 2023). Dewaele's et al. (2023) study involving 141 Arab bi-multilinguals, used this scale and found that the AoA and context and nature of exposure during primary education increased English emotional resonance. The current intensity and frequency of exposure to English enhanced language embodiment. Notably, proficiency did not necessarily imply language embodiment.

Individuals' cultural backgrounds also greatly impact their emotional expression and perception. Each culture provides a framework for conceptualising emotions and offers a distinct set of emotion scripts (Wierzbicka, 1999), shaping how emotions are interpreted and communicated. Consequently, *emotion concepts* (Pavlenko, 2008) serve as unique prototypical scripts that encapsulate and convey emotions based on society's collective experiences. Furthermore, diversity in emotional language use is linked to the presence or absence of emotion equivalents (Altarriba, 2003; Wierzbicka, 2004), explaining variations in usage among bi-multilinguals.

2.2.3. Sociolinguistic

The frequency of language use significantly impacts emotional perception in LX (e.g. Degner et al., 2011; Dewaele et al., 2023). A high frequency of language use (i.e. high level of language immersion) indicates strong socialisation, leading to a diverse network of interlocutors, increased proficiency, and a greater likelihood of expressing emotions in that language (Dewaele, 2010). Secondary socialisation (Bayley & Schecter, 2003) may result in a conceptual shift to LX for expressing and perceiving emotions (Pavlenko, 2004, 2005), especially for concepts lacking equivalence in L1 (Pavlenko, 2008).

Likewise, being bi-multilingual allows individuals to evaluate and interpret emotional experiences from multiple perspectives (Pavlenko, 2008). The perception of oneself and the world is shaped by the vocabulary provided by the L1/LX language(s) (Wierzbicka, 1999), making emotional experiences more than just comprehension.

Thoughts, memories, and emotions can be significantly influenced by the language in which events occur, leading to retrieval in that language when triggered (Marian & Neisser, 2000).

This research contributes to the literature on the phrase *I love you* by focusing on adolescents in the family and friendship domains, employing combined frequency analysis and inferential statistics (multinomial logistic regression) to analyse the impact of both individual and collective factors on emotional perception.

2.3. Assessment clusters in TCKs

To our knowledge, no studies have explored the perception of *I love you* in TCKs. The following section provides an overview of the clusters analysed in the study.

2.3.1. Sociobiographical cluster

TCK's perception of love is expected to be influenced by their transient lifestyle and emotional dynamics, centred around the family, international school, and expatriate community.

Their transient lifestyle often strengthens family cohesion (e.g. Lê et al., 2010), as the family becomes the primary source of stable socialisation across moves, resulting in a "family bubble" with parents and children, while the extended family is left behind (Schaetti & Ramsey, 1999). They often (not always) come from high socio-economic backgrounds and may access additional resources, such as domestic staff upon arrival in the host country (e.g. De Mejía, 2002), though this varies by destination. Early relocation age and interaction with domestic staff in a language different from parents may contribute to early bi-trilingualism (Eidse & Sichel, 2004), especially in intercultural families.

TCKs' families are connected to their community and international school through a sponsor (the caregiver's employer), which enhances their sense of belonging (Sichel et al., 2011). Families further strengthen their bonds by socialising within the "expatriate bubble", a closed social network formed through encounters with other expatriate families from their children's international schools, which limits interaction with locals and perpetuates linguistic reliance on English (Meyer, 2021). Similarly, the international school fulfils TCKs' academic and social needs, further isolating them from the host country (Benjamin, 2017). In Denmark, this study's context, English functions as an unofficial L2 (Lønsmann et al., 2022). This status is supported by the high proficiency of locals in English which permeates into daily life through original English-language films and TV shows.[2] This not only reinforces English as a lingua

franca in the country, but also gives TCKs and their families the impression that learning Danish is unnecessary, perpetuating the use of English.

This study anticipates that the aforementioned sociobiographical factors may play a crucial role in shaping the development of both TCKs' linguistic and sociolinguistic traits.

2.3.2. Language profile cluster

TCKs' language profile is shaped by the combined result of their linguistic history and current linguistic abilities, transcending mere linguistic significance to become markers of identity and national affiliation (Jeon, 2022). For instance, TCKs often associate their L1 with family ties and roots, as their concept of "home" constantly changes (Lijadi & Van Schalkwyk, 2017). This complexity increases when TCKs have parents in intercultural relationships. Consequently, this study considers the importance of the L1 in shaping TCKs' perception of *I love you*.

TCKs' early bi-trilingualism (Eidse & Sichel, 2004) is not always correlated with their age of mobility. Tannenbaum and Tseng's (2015) research on 54 adult TCKs found that language proficiency and dominance were not linked to AoA or mobility age, but rather to the frequency of language use. As English proficiency increased, it became the dominant language in the TCKs' linguistic repertoire, despite not being their L1 or the language of their current residence.

Research on TCKs' linguistic emotional use has yielded mixed results. Some studies suggest a reduced emotional perception of L1 in favour of English LX due to TCKs' nomadic component (Tannenbaum & Tseng, 2015), while others emphasise the emotional importance of L1 spoken at home, restricting English usage to the academic domain (Jeon, 2022).

The CoA also influences emotionality (Section 2.2.2). In this study, 72.1% of TCKs acquired their L2 (English for 68%) in a mixed or instructed context, potentially intensifying emotionality in their L2 (Section 2.1).

2.3.3. Sociolinguistic cluster

While exposed to their parents' L1, most TCKs lack significant exposure to their L1's country or the host country, resulting in a unique acculturation process within the "expatriate bubble" (Rodríguez-Bernal et al., 2023). Interaction with immediate family at home is the primary source of communication for TCKs due to distance from extended family (section 2.3.1), adhering to various FLPs, including: (1) intercultural parents adopting the "one parent one language approach" (OPOL); (2) parents with a shared

L1 adopting the "non-societal vs. societal approach" (i.e. parents' shared L1 vs. English), and (3) parents adopting a flexible approach where all shared languages are used, the so-called "Happylingual Approach" (Kopeliovich, 2013). English "leakage" (De Houwer & Bornstein, 2016) is expected due to its prevalence in international schools, parents' workplaces, and media exposure in Denmark (Dewaele et al., 2021). Consequently, this study examined TCKs with L1 Danish and at least one Danish parent.

At school, TCKs mingle with peers from similar backgrounds, fostering a sense of belonging where friendships become a vital source of emotional support (Lijadi & Van Schalkwyk, 2017). This environment is deeply rooted in the Anglophone world (Rydenvald, 2018), especially in the IB program, following an American-based curriculum (Carder, 2007), and characterised by an American-centric approach (Meyer, 2021). Therefore, TCKs must navigate conflicting emotional cues and cultural scripts regarding expressions of love. What might be considered a personal matter in their household culture could differ from American norms, where the phrase *I love you* is frequently verbalised in contexts and relationships deemed inappropriate in other cultures (e.g. Caldwell-Harris et al., 2011).

This study contributes to TCK literature by examining a larger and well-defined sample of TCKs (Tan et al., 2021), still considered "kids", minimising biases associated with retrospective methodologies. Unlike previous studies it focuses on a linguistic analysis exploring the emotional use of TCKs' language repertoire, considering the interplay between their socialisation, distinct sociobiographical features, and the impact of their current host country.

3. Research questions

The study examined the following research questions (RQs):

RQ1. Does the emotional impact of the phrase *I love you* vary across TCKs' linguistic repertoire, which language(s) elicit the strongest emotional response?

H1. It is hypothesised that the emotional impact varies across languages and that LX English contributes significantly to this perception.

RQ2. To what extent do TCKs' sociobiographical, language profile, and sociolinguistic clusters favour the variation in the perception of the phrase *I love you* across the TCKs' linguistic repertoire?

H2. It is hypothesised that, while individual clusters may contribute independently to the outcome, the optimal model arises from the interplay of variables primarily originating from the sociobiographical cluster, where L1 is anticipated to have a significant impact.

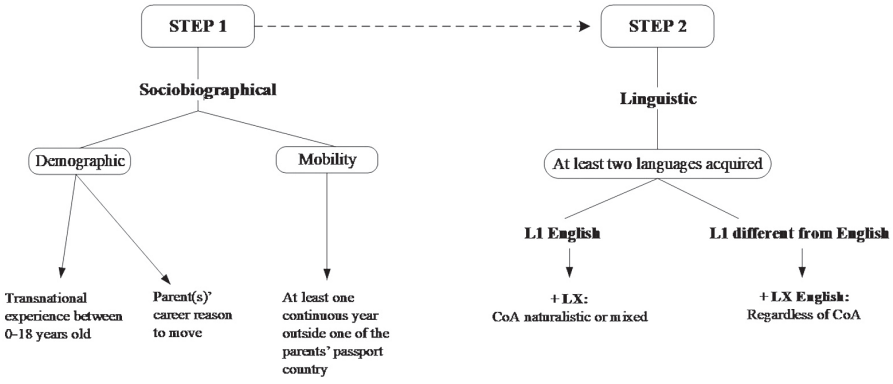
4. Methodology

The study employed a sequential mixed-methods design (Johnson & Christensen, 2020), combining quantitative data from an online questionnaire with qualitative insights from participants' open-ended questionnaire responses and semi-structured interviews. Qualitative data complemented the quantitative findings, allowing participants to share their personal perspectives. A broader qualitative analysis was beyond the scope of this study.

4.1. Participants

The study applied the following exclusion criteria (Figure 1):

Figure 1: TCK-sampling requirements



Sociobiographical criteria followed the traditional TCK definition (Pollock et al., 2017), while linguistic criteria were established considering factors influencing emotionality in an LX (Section 2.2.2). Although data from five languages were collected, L3, L4, and L5 were excluded due to their primarily instructed acquisition context, lower oral proficiency, and infrequent use, common in school-based language learning.

The study involved 276 students (123 males, 146 females, and seven non-binaries), aged 15-19 (M = 17; SD = 0.9), enrolled in the IB program across three international schools in Denmark. Participants exhibited significant international mobility, with 34.8% relocating between ages 0-2, spending an average of 9.6 years outside at least one parent's passport country (SD = 4.3). A majority (67.4%) had never resided in an English-speaking country, moving to an average of three different countries (SD = 0.9), and spending approximately half their lives in international schools (M = 8, SD = 4) due to parental professions. Additionally, 39.1% of households were intercultural couples, with 83% being LX English users (Appendix I, Sociobiographical cluster).

The participants represented 58 countries, 68 languages, and 52 different L1s. There were 31 bilinguals, 92 trilinguals, 98 quadrilinguals and 55 pentalinguals. Danish represented the largest L1 (19.9%), followed by English (12.7%). English was the second most common L2 (68%), followed by Danish (6.9%). 42.4% were bilingual L1 users and 8% were trilingual L1 users. About 45.7% reported dominance in their L1, while 39.1% reported dominance in two or more languages, including L1. Furthermore, 15.2% reported English LX dominance, which is a common outcome of international school enrolment (Appendix I, Language profile cluster).

The interviews involved seven participants (two males, four females, and one non-binary) organised into three pre-established groups: (1) strict FLP approach; (2) flexible FLP approach; and (3) LX English dominant with varying FLP approaches (Appendix I, Qualitative description of the sample).

4.2. Instruments and procedure

Quantitative data were collected via a voluntary online questionnaire completed outside school hours. The first section gathered sociobiographical information to validate the TCK profile. The second section used an adaptive version of the *Bilingualism and Emotions Questionnaire* (Dewaele & Pavlenko, 2001-2003) to collect language history data, including participants' self-reported emotional impact of the locution *I love you* across languages and the language(s) with the strongest emotional weight.

Factors of analysis were organised into three clusters.

1. **Sociobiographical:** Age of first move, countries TCK resided (at least for a year), years in international school, intercultural parents, and Danish parent(s) (Appendix I, TCKs with Danish parent[s]). Intercultural, and Danish parent(s) were coded as binary variables.

2. **Language profile:** Factors were divided into *linguistic history* and *linguistic competence*, enabling examination of the past (linguistic history) and present (current linguistic competence) influences on TCKs' perception of *I love you.*

    2.1. *Linguistic history:* AoA L2, CoA L1-L2 and subjective perception L1-L2 emotional. CoA was coded as naturalistic (outside school) or mixed (inside and outside school). L1-L2 subjective emotional perception was assessed through the statement *My L1-L2 is emotional,* recoded into three categories: high (*to a large extent, absolutely*), moderate (*more or less*), and low (*not at all, somewhat*).

2.2. *Linguistic competence:* Self-reported dominant language(s) and average self-reported oral skills L1-L2. Self-reported dominant language(s) were grouped into: L1, L1+LX, and LX. L1-L2 self-reported oral skills scores ranged from 1 (*least proficient*) to 5 (*fully proficient*); an arithmetic mean was calculated (self-rated score for speaking and listening divided by two).

3. **Sociolinguistic:** Network of interlocutors L1-L2, years living in Denmark, frequency of use of L1-L2, LX socialisation level, household linguistic mode, and English spoken at home (Appendix I, Sociolinguistic cluster). Network of interlocutors was categorised as general (*nobody, all, schoolmates, strangers*) or close (*family, friends*). Language frequency of use was classified as high (every day, several hours a day) or low (never, yearly, monthly, weekly). LX socialisation level was calculated by subtracting L1 and L2 frequency scores, with a negative result indicating higher LX socialisation. Household language modes were grouped into *monolingual* (one language) and *bi-multilingual* (two or more languages), *English spoken at home* was coded as binary (Appendix II, general FLP sample description).

One dependent variable was considered:

Language(s) with the strongest emotional impact were assessed through the question: "Does the equivalent of 'I love you' have the same emotional impact for you in your different languages, which language(s) does the equivalent of 'I love you' feel the strongest?". Multiple answers were allowed for a maximum of five languages and recoded into three groups: Group 1 (L1), Group 2 (L1+LX), and Group 3 (LX).

Interviews, averaging 45 minutes each, were conducted in English, tailored to participants' questionnaire responses (Brinkmann & Kvale, 2014). Recorded data were transcribed verbatim. Ethical approval was obtained from the university, school administrators, and adolescents prior to the data collection.[3]

4.3. Analyses

SPSS Version 28 (IBM Corp., 2021) was used for the analysis. For RQ1, frequency analysis was conducted for each language option. RQ2 employed a two-step approach: Firstly, forward stepwise multinomial logistic regression (MLR) analysis (Model 1) identified significant variables from the sociobiographical, language profile, and sociolinguistic clusters at an alpha of .05. Secondly, theory-driven factors were manually incorporated into the model, aiming for a lower *-2-Log Likelihood* compared to Model 1, considering indicators such as p-values, odd ratios, Nagelkerke R^2, and classification percentages. Non-significant variables were determined using a conservative fitting

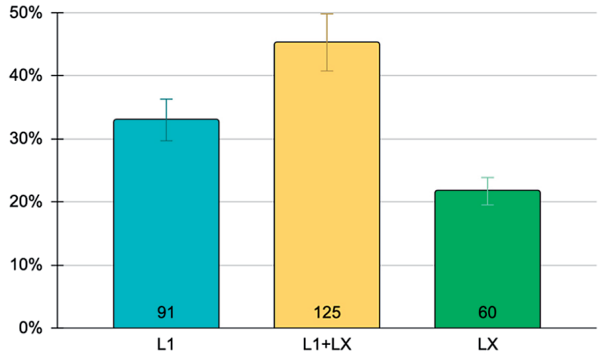
approach (p >.10), common in exploratory studies (Fisher, 1925). Parameter estimates, odds ratios, and 90% confidence intervals were used for each predictor (Model 2). A more robust predictive model was developed by combining data-driven and theoretically informed predictors (Freedman, 2009). The checks for *linearity in the logit* were satisfactory. However, high collinearity between *socialisation in L2* and *frequency of use L1* (r = -.792) and *frequency of use L2* (r = .817) was addressed, retaining the latter, due to their theoretical relevance and conceptual importance. Model fit was assessed using the *-2-LogLikelihood*, with lower scores indicating better fit and, while effect size was evaluated using Nagelkerke R^2 (Tabatchnick & Fidell, 2007).

5. Results

5.1. TCKs' emotional perception of the phrase "I love you" across languages: Identifying the language(s) that elicit the strongest emotional response

About 45.3% of TCKs perceived *I love you* to be more emotionally impactful in L1+LX, approximately one-third (33%) felt it was stronger in L1, and slightly less than a quarter (21.7%) perceived a greater impact in LX (Figure 2).

Figure 2: Percentage of the emotional impact of *I love you* across languages



Interviews and open-ended questionnaire responses supported the heightened emotional perception of L1+LX,[4] attributed to socialising and navigating emotional situations with close interlocutors in these languages.

> INF 265 (L1 Hindi, L2 English, L3 Danish, L4 German) *Hindi and English are the languages I use with family and friends at the school; I feel that I love you carries more emotional weight in them, whereas Danish and German are mainly used with people I've a weaker or no emotional attachment to.*

Some participants reported increased emotional weight in L1+LX (English) due to adapting to diverse linguistic demands at home, where L1 was used with parents and English prevailed among siblings, influenced by shared international schooling. A generational shift emerged, with parents prioritising their L1 transmission to the firstborn, but this trend declined with successive children, leading younger family members to predominantly use English, even in households with strict FLP, discouraging it.

> Mads (L1 Danish, L2 English, L3 Chinese, L4 Malay, L5 Spanish) *I´d say Danish and English are equally emotional. My dad spoke Danish to me since the beginning, but that changed with my younger sister; she's trying to learn Danish now that we live in Denmark, but we've always communicated in English.*

5.2. Sociobiographical, language profile and sociolinguistic clusters influencing variation in the perception of "I love you" across TCK's linguistic repertoire

Table 1 summarises the MLR results, identifying significant factors in the initial data-driven model and final theory-driven model for perceived emotional weight variations of *I love you* across TCKs' languages.

Table 1: Data-driven and final theory-driven MLR models: Factors influencing variations in the perceived emotional weight of *I love you* across TCKs' languages, with L1 as the reference category (N = 276)

		Model 1 Data-driven		Model 2 Theory-driven inclusion					
		L1+LX	LX	L1+LX	O.R.	90% CI O.R.	LX	O.R.	90% CI O.R.
Intercept		.68 (.65)	1.8 (.70)	.26 (.82)			.99 (.95)		
Sociobiographical cluster	Group								
Age of first move		.05* (.02)	-.02 (.03)	.07* (.03)	1.071	[1.017, 1.129]	.01 (.04)	1.010	[.943, 1.083]
Intercultural parents[a]				-.68+ (.37)	.505	[.271, .942]	.44 (.50)	1.556	[.683, 3.545]
Years in international school				.06 (.04)	1.063	[.992, 1.139]	.06 (.05)	1.065	[.974, 1.165]
Danish parent(s)[b]				-.01 (.38)	.987	[.523, 1.863]	-.98* (.45)	.375	[.177, .796]

		Model 1 Data-driven		Model 2 Theory-driven inclusion					
		L1+LX	LX	L1+LX	O.R.	90% CI O.R.	LX	O.R.	90% CI O.R.
Language profile cluster									
Linguistic history									
AoA L2[c]	0-2 years	.56 (.48)	-2.1*** (.62)	.34 (.50)	1.414	[.613, 3.264]	-2.2*** (.66)	.105	[.035, .315]
	3-7 years	-.04 (.46)	-.36 (.48)	-.04 (.48)	.957	[.429, 2.134]	-.36 (.52)	.692	[.294, 1.630]
CoA L1[d]	naturalistic			-.02 (.31)	.981	[.584, 1.647]	.44 (.40)	1.559	[.807, 3.011]
Perception L1 emotional[e]	medium			-.06 (.42)	.936	[.437, 2.001]	.46 (.53)	1.589	[.658, 3.838]
	low			.21 (.47)	1.240	[.570, 2.696]	.92 (.57)	2.531	[.984, 6.508]
Linguistic competence									
Self-reported dominant language(s)[f]	L1	-1.0* (.48)	-1.3*(.52)	-.86+ (.52)	.421	[.177, .999]	-1.0+ (.57)	.352	[.137, .905]
	L1 +LX	.26(.49)	-.20 (.51)	.52 (.51)	1.693	[.726, 3.948]	-.04 (.55)	.954	[.386, 2.356]
Sociolinguistic cluster									
Frequency use L1[g]	low			1.3+ (.72)	3.708	[1.121, 12.257]	.47 (.94)	1.605	[.341, 7.550]
Household linguistic mode[h]	monolingual	-.95** (.31)	-1.2** (.39)	-.87* (.44)	.417	[.200, .868]	-.55 (.57)	.578	[.226, 1.475]
English spoken at home[i]				.10 (.45)	1.111	[.529, 2.335]	-.90 (.57)	.406	[.158, 1.045]
-2LogLikelihood, p-value		506.117, p <.001		478.535, p <.001					
Nagelkerke R^2		.277		.355					

Note: Numbers in parentheses are standard errors. + $p \leq 10$ *p <.05. **p <.01. ***p <.001. [a] Reference Cat. is parents with different L1s. [b] Reference Cat. is TCK with Danish parent(s). [c] Reference Cat. is AoA learnt 8-18 years old. [d] Reference Cat. is CoA mixed and instructed. [e] Reference Cat. is high subjective emotional perception. [f] Reference Cat. is Dominant LX. [g] Reference Cat. is high frequency of use. [h] Reference Cat. is bi-multilingual language model. [i] Reference Cat. English spoken at home.

Table 2 shows the case processing of predictors in the final logistic regression model.

Table 2: Case-processing summary final logistic regression model (N = 276)

Predictor	Group	N	Marginal percentage
Sociobiographical cluster			
Intercultural parents	shared L1	168	60.9%
	different language L1s	108	39.1%
Danish parent(s)	yes	82	29.7%
	no	194	70.3%
Language profile cluster			
Learning history			
AoA L2	0-2 years old	117	42.4%
	3-7	115	41.7%
	8-18	44	15.9%
CoA L1	naturalistic	137	49.6%
	mixed and instructed	139	50.4%
Perception L1 emotional	high	190	68.8%
	medium	43	15.6%
	low	43	15.6%
Linguistic competency			
Self-reported dominant language(s)	L1	126	45.7%
	L1+LX	108	39.1%
	LX	42	15.2%
Sociolinguistic cluster			
Frequency use L1	high	256	92.8%
	low	20	7.2%
Household linguistic mode	monolingual	149	54.0%
	bi-multilingual	127	46.0%
English spoken at home	yes	131	47.5%
	no	145	52.5%

Model 1

A data-driven model (Model 1) using forward stepwise MLR identified predictors influencing the emotional weight of *I love you* across languages within the three clusters. Noteworthy predictors included sociobiographical: *age of first move* (β =.05, p = .050); language profile: *AoA L2* (β = -2.1, p < .001) and *self-reported dominant language(s)* for L1+LX (β = -1.0, p = 0.3) for LX (β = -1.3, p = .013), and sociolinguistic: *household linguistic mode* for L1+LX (β = -.95, p = .002) and LX (β = -1.1, p = .003).

The *-2-LogLikelihood* ratio test yielded statistical significance (p < .001), indicating its adequacy in explaining 27.7% of the variance (Nagelkerke R^2 = .277) in the perceived emotional weight of *I love you* across languages (Table 2, Model 1).

Model 2

1. The significant data-driven predictors per cluster were:

Sociobiographical

Age of first move: The odds ratio (OR) of 1.071 (p = .030, CI_{90} [1.017, 1.129]) indicated that for each one-year increase in age, the odds of perceiving *I love you* in L1+LX as stronger than in L1 increased by 7.1%.

Language profile

AoA L2: The OR = .105 (p < .001, CI_{90} [.035, .315]) suggested that TCKs who learned L2 between 0-2 years of age had an 89.5% decrease in the odds of perceiving *I love you* in LX as stronger than in L1, compared to those who acquired L2 between 8-18 years of age.

Self-reported dominant language(s): The OR = .421 (p = .10, CI_{90} [.177, .999]) suggested that TCKs reporting L1 as dominant were 57.9% less likely to perceive *I love you* in L1+LX as stronger than in L1, compared to LX-dominant peers. This trend persisted when perceiving *I love you* in LX as stronger than in L1: OR = .352 (p = .070, CI_{90} [.137, .905]), reducing the odds to 64.8% for LX having the highest emotional weight compared with L1.

Sociolinguistic

Household linguistic mode: The OR = .417 (p = .050, CI_{90} [.200, .868]) indicated that TCKs with a monolingual mode at home were 58.3% less likely to perceive *I love you* in L1+LX as stronger than in L1, compared to TCKs in bi-multilingual mode at home.

2. The theory-driven predictors incorporated into Model 1 per cluster were:

Sociobiographical

Intercultural parents: The OR = .505 (p = .071, CI_{90} [.271, .942]) indicated that TCKs with shared parental L1 were 49.5% less likely to perceive *I love you* in L1+LX as stronger than in L1, compared to TCKs whose parents had different L1s.

Danish parent(s): The OR = .375 (p = .032, CI_{90} [.177, .796]) indicated that TCKs without Danish parent(s) were 62.5% less likely to perceive *I love you* as more emotional in LX than in L1, compared to those TCKs with at least one Danish parent.

Years in international school: Although non-significant, this variable was retained in the final model (Tables 3 and 4).

Language profile

Although non-significant, *CoA L1* and *Perception L1 emotional* were retained in the final model (Tables 3 and 4).

Sociolinguistic

Frequency of use of L1: The OR= 3.708 (p = .071, CI_{90} [1.121, 12.257]) indicated that TCKs with low L1 frequency of use were 270% more likely to perceive *I love you* as more emotional in L1+LX than in L1 alone, compared to those with a high L1 frequency of use.

English spoken at home: Although non-significant, this variable was retained in the final model (Tables 3 and 4).

3. Non-significant theory-driven predictors retained in the model:

Table 3: Change in Model fit

Model comparison			
Model	*-2-Log Likelihood*	Nagelkerke R^2	Classification Accuracy
Full Model with all predictors	478.535	.355	59.1%
Reduced Model (without non-significant predictors)	427.972	.323	54.3%

Table 3 illustrates the collective impact of incorporating non-significant predictors. Despite raising the *-2-LogLikelihood*, their inclusion enhanced the comprehensiveness and accuracy of the model, as evidenced by the increased Nagelkerke R^2 (from .323 to .355) and classification accuracy (from 54.3% to 59.1%).

Table 4 presents the impact of each non-significant predictor included in the model.

Table 4: Contribution of theory-based non-significant predictors to Model 2 fit

Non-significant theory-based predictors						
Predictor	*-2-LogLikelihood* (full model)	*-2-LogLikelihood* (model without variable)	Change in *-2-LogLikelihood*	Nagelkerke R^2 (without variable)	*p*	Classification (without variable)
Perception L1 emotional	478.535	481.977	- 3.442	.345	.48	55.1%
Years international school		475.463	+ 3.072	.348	.29	58.7%
CoA L1		476.001	+ 2.534	.350	.44	58.7%
English spoken at home		480.615	- 2.08	.345	.17	59.4%

Perception L1 emotional: The *-2-LogLikelihood* decreased (-3.442), indicating an improved model fit, supported by enhanced Nagelkerke R^2 (from .345 to .355) and classification accuracy (from 55.1% to 59.1%).

Years in international school: Despite an increase in *-2-LogLikelihood* (+3.072), including it improved the overall model fit, reflected in the increased Nagelkerke R^2 (from .348 to .355) and enhanced classification accuracy (from 58.7% to 59.1%).

CoA L1: Despite a slight increase in *-2-LogLikelihood* (+ 2.534), Nagelkerke R^2 also increased (from .350 to .355), positively impacting the model's explanatory ability.

English spoken at home: Despite a decrease in *-2-LogLikelihood* (- 2.08), the model fit improved, supported by an increase in Nagelkerke R^2 (from .345 to .355).

4. Overall Model 2 accuracy:

Table 5: Classification accuracy for both models

Model 1 Data-driven					**Model 2 Theory-driven inclusion**			
Observed	L1	L1+LX	LX	% Correct	L1	L1+LX	LX	% Correct
L1	37	40	14	40.7	46	36	9	50.5
L1+LX	22	80	23	64.0	23	89	13	71.2
LX	10	22	28	46.7	12	20	28	46.7
Overall %	25.0	51.4	23.6	52.5	29.3	52.5	18.1	59.1

Table 5 compares the accuracy percentages of both models in predicting the three outcomes. Overall, Model 2 enhanced the accuracy (from 52.5% to 59.1%). Significant improvements were observed for L1 (from 40.7% to 50.5%) and L1+LX (from 64% to 71.2%). However, the accuracy of the LX level remained constant at 46.7%.

Results across the three clusters aligned with the insights from the interviews and open-ended questionnaire responses regarding the perceived emotional weight of *I love you.*

Sociobiographical cluster

Several respondents described a significant rise in English use after relocation, attributed to forming connections with new interlocutors and interacting with peers at the international school, indicating that *language frequency of use* outweighs AoA in emotional resonance within this context.

> Freja (L1 Danish, L2 English, L3 French, L4 Arabic, L5 German; 8 years old at first move) *Abroad, we resided in these communities, we barely went out, you live in a compound, a gated area. People who lived there were all international(...) In India, we had a maid because that was very normal, and I spoke English to her(...) plus attending international school every day, so it was always English.*

Participants' testimonies further supported the heightened emotional perception of L1+LX in the phrase *I love you* for TCKs with parents in an intercultural relationship, where English became the primary mode of communication.

> Mads (L1 Danish, L2 English, L3 Chinese, L4 Malay, L5 Spanish) *They* (Danish, English, and Malay) *all have the same emotional weight for me since I grew up with them; my parents mainly communicated in English since my mother* (Malaysian) *could not speak Danish* (father's language).

Having at least one Danish parent unexpectedly found greater emotional resonance in English (LX) compared to their L1 (Danish), indicating a potential "emotional erosion" of Danish, particularly notable among those with one Danish parent (Appendix I, Descriptive statistics of TCKs with Danish parents), possibly due to increased English use displacing emotional perception in Danish.

> Morten (L1 Danish, L2 English, L3 French; Strict FLP: L1) *I've noticed over time that when I speak Danish to express my deepest feelings, I won't get as emotional and I'll remain emotionless, whereas with English I can really evoke and perceive deep feelings.*

Language profile cluster

Influenced by a nomadic lifestyle, some TCKs prioritised English over their L1, potentially leading to early emotional integration of English alongside L1, despite sequential English acquisition, as noted in the quantitative analysis.

> Elin (L1 Swedish, L2 English, L3 Spanish) *In Egypt I always spoke English to the driver and the nanny (...) when my Swedish aunt visited, I spoke English to her and she'd respond in Swedish (...)'cause I was always around English.*

Other TCKs highlighted the emotional significance of L1, describing how using it in intimate, emotionally charged situations with familiar interlocutors evoked childhood memories and emotions.

> INF 209 (L1 Russian, L2 Romanian, L3 English, L4 Danish) *To me, strong emotions are associated with certain languages because of my past experiences. For example, the first time I heard I love you was in Russian; therefore, I mainly associate love with Russian.*

The relationship between perceived emotional weight and language dominance in TCKs is complex. While the quantitative analysis highlighted the significance of language dominance in the emotional perception of *I love you,* qualitative findings uncovered a more nuanced picture. Even with high L1 dominance, English dominance seemed crucial in TCKs' expression of love, with participants underscoring the linguistic versatility of English.

> Freja (L1 Danish, L2 English, L3 French, L4 Arabic, L5 German) *English is more emotional to me, it's easier to express myself, there's a lot of synonyms, whereas in Danish I feel there's one word, the right word; in English you have all these adjectives to make it emotional, whereas in Danish you need to state the truth.*

Moreover, English emotional preference seemed to transcend proficiency, with some TCKs reporting a unique sense of affinity, self-perception, identity connection, and well-being unparalleled compared to their L1.

> Sergei (L1 Russian, L2 English, L3 Spanish, L4 Danish) *Although I grew up with Russian, it doesn't accommodate for certain parts of my identity...I'm non-binary and I can communicate in English freely; gender neutrality isn't a thing in Russian, I have to conjugate every verb to the gender...it makes me feel uncomfortable.*

Other questionnaire responses revealed heightened emotional resonance despite limited linguistic abilities, suggesting associations with past experiences (memories, childhood), the influence of specific interlocutors in intimate contexts, and emotional resonance with a particular language. Moreover, limited language knowledge can facilitate connection between TCKs and their extended family.

> INF269 (L1 Tagalog, L2 English, L3 Danish, L4 German) *I don't speak much Tagalog...only bits and pieces which I learned from my parents, but greatest emotional weight is in Tagalog. For me is the power behind the words that are backed by my culture, what keeps me connected to my family.*

Sociolinguistic cluster

TCKs' testimonies echoed quantitative findings, emphasising English use prevailing over strict FLP, to the detriment of their L1(s). School and expatriate community interactions, coupled with emotional exposure to English through the media, further reinforced the importance of frequent and intense language exposure in emotional resonance.

> INF53 (L1 German, L2 French, L3 English, L4 Spanish. Strict FLP: L1-L2) *It just feels right in English as it's the language I'm most comfortable and exposed to when it comes to emotions, like my friends at school, significant other, and such; it's what I'm used to, at home too. Also shows, movies in which these subjects come up are all in English.*

Other participants expressed mixed views on the emotional weight of *I love you*, with some corroborating its widespread use, implying reduced authenticity, while others saw it as a convenient expression lacking the seriousness of their L1 equivalent, indicating that for some TCKs, L1 retained greater emotional authenticity, especially in households with strict FLPs.

> Lena (L1 German, L2 English, L3 Spanish. Flexible FLP: L1-L2-L3) *At school, 'I love you' is almost said to everyone and almost all the time, so the impact is a bit less than when you say it in German or Spanish.*

> Renata (L1 Portuguese, L2 Spanish, L3 English, L4 French. Strict FLP: L1) *For me to say 'I love you' to a friend or to my parents, it's easier than in Portuguese, which is very strong. I'm not verbal with my emotions, so I tend to stick to English because I find it much easier, less strong.*

Furthermore, participants noted discrepancies between the emotional impact of the phrase in English and its corresponding emotion script in their L1, prompting them to resort to English for expressing emotions considered too intense or unconventional in their L1.

> Elin (L1 Swedish, L2 English, L3 Spanish. Strict FLP: L1) *As a family, we're not that touchy, saying 'I love you' in Swedish is so wrong. My parents will have an entire conversation in Swedish and they'll say 'I love you' in English at the end, never in Swedish, in Swedish it doesn't come naturally, it's feels old fashioned, it's just weird.*

6. Discussion

6.1. "I love you" across TCKs' languages and the language(s) that elicit the strongest emotional response

The first research hypothesis, that the emotional impact of the phrase *I love you* varies across TCKs' linguistic repertoire, with LX English contributing to this perception, was corroborated. Nearly half of the participants reported the strongest perceived emotional response with the L1+LX (English) combination, which differs from previous studies (Dewaele, 2008; Jahangard & Holderread, 2013; Ożańska-Ponikwia, 2016) where L1 was deemed the most emotionally significant.

This outcome reflects the societal habits of TCKs, emphasising language use and resulting in a blend of L1 and LX English, ultimately forming a harmonious linguistic partnership that embodies translanguaging (García & Li, 2014).

TCKs' socialisation revolves around their family and the expatriate community, where the international school is integrated. On the one hand, the household operates as a "family bubble" (Schaetti & Ramsey, 1999) under a predefined FLP, which can vary in flexibility. Qualitative data demonstrated occasional "leakage" (De Houwer & Bornstein, 2016) of English in households with a strict FLP or blended use with the L1(s) in those following a "Happylingual Approach" (Kopeliovich, 2013).

It is noteworthy that younger siblings, primarily communicating in English despite a strict FLP at home, influenced this phenomenon (De Houwer, 2020). Moreover, TCKs whose parents were in an intercultural relationship and had adopted English as a language of communication in the absence of a shared L1 were observed. While comprising only 5% of the study (Appendix II, general FLP sample description), this indicated that these TCKs were raised in English as an LX at home, corroborating the need to include parental use of the LX to investigate the combined effects of AoA and CoA on emotional resonance (Dewaele et al., 2023).

On the other hand, the "expatriate bubble" and international school significantly influence the emotional development of TCKs, where friends play an essential role (Lijadi & Van Schalkwyk, 2017), and the frequency and intensity of exposure to English are maximised, given its status as lingua franca.

Consequently, the convergence of the household context, expatriate community, and international school explains why *I love you* in L1+LX is perceived with heightened emotionality by TCKs, as they have experienced emotional situations in these languages with diverse interlocutors. This not only validates the necessity for both languages to fulfil TCKs' emotional needs, in accordance with the Complementary Principle (Grosjean, 1997), but also explains TCKs' emotional dynamics in a language mode continuum (Grosjean, 1985), spanning from monolingual situations (e.g. parents at home or friends at school) to bilingual scenarios (e.g. intercultural parents or siblings at home). This adjustment in language use shows TCKs' ability to adapt to varying contexts and interlocutors in emotional situations. It also demonstrates the fluidity and interconnectedness of languages within these TCKs, consistent with the principles of translanguaging where linguistic boundaries are blurred (García & Li, 2014).

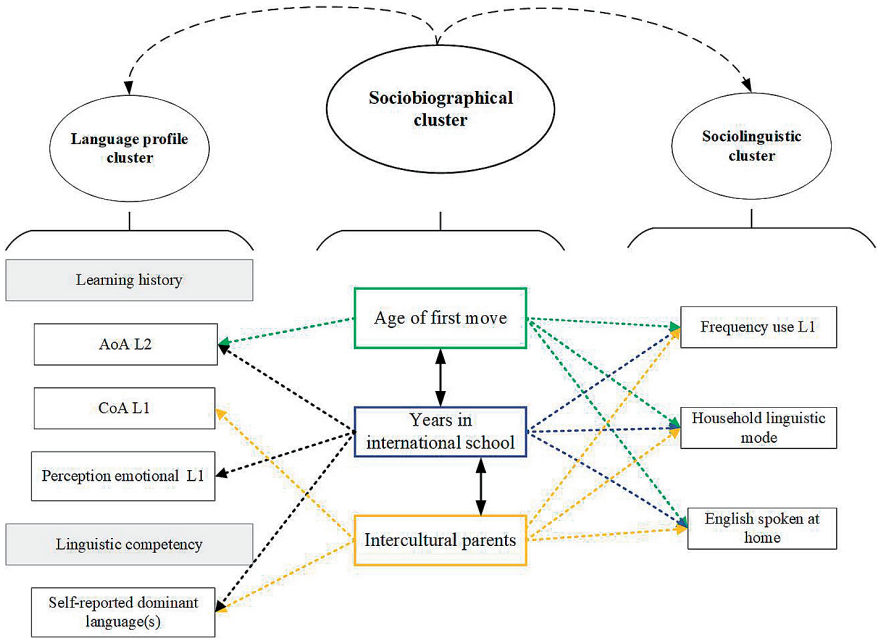
Nonetheless, it is worth exploring how TCKs maintain the emotional perception of their L1(s) in an English-dominant environment despite relocation challenges, such as reduced exposure to diverse input and limited interaction opportunities with diverse interlocutors due to geographical distance from extended family. Moreover, it is intriguing to investigate how English becomes effortlessly incorporated into TCKs' emotional linguistic repertoire. The next section explores these nuances further.

6.2. Influence of the sociobiographical, language profile and sociolinguistic cluster on the perceived emotional weight of "I love you" across TCKs' linguistic repertoire

The second research hypothesis, that the optimal model for the perceived emotional weight of *I love you* arises from integrating variables primarily from the sociobiographical cluster with an anticipated impact of L1, was supported.

Figure 3 illustrates the central role of the sociobiographical cluster, representing the driving force behind the results within the language profile and sociolinguistic clusters.

Figure 3: Scope of influence, sociobiographical cluster's impact on language and sociolinguistic dynamics



6.2.1. Influence of the sociobiographical cluster on the language profile

Figure 3 depicts the bidirectional relationship between sociobiographical factors, *age of first move*, and *years in international school*, resulting from TCKs' relocation. It emphasises the onset and cumulative duration of English exposure and the depth of immersion in the expatriate community during schooling. Despite their interdependence, these factors independently influenced the language profile cluster.

While the AoA L2 (LX English) variable aligned with expectations, indicating that an early acquisition of LX predicted a closer emotional perception to L1 due to its naturalistic CoA (Harris, 2004), the *age of first move* seemed to mitigate its effect on the perceived emotional weight of *I love you*, regardless of whether the acquisition was simultaneous or sequential. Essentially, the perception of English as a language with high emotional resonance would develop over time, rather than being influenced by

the age at which TCK began moving. Research has shown the significance of language proficiency in accurately perceiving emotions (Dewaele & Nakano, 2013; Caldwell-Harris et al., 2011). In this study, AoA was not associated with acquiring dominance (cf. Pavlenko, 2005), aligning with TCKs' unique immigration patterns diverging from the conventional correlations between AoA and dominance (Tannenbaum & Tseng, 2015). Instead, the current study proposes that *self-reported language dominance* and heightened perceived emotional perception in LX (English) originated from *years in international school,* indirectly resulting from relocation (black arrows in Figure 3). The naturalistic CoA of LX English (Harris et al., 2006) for TCKs, fostered through interactions at the international school and within the expatriate community, strengthens its dominance and perceived emotional resonance over time (Dewaele et al., 2021), establishing it as an embodied language (Dewaele et al., 2023; Pavlenko, 2005; Rodríguez-Bernal et al., 2023). The significance of self-reported language dominance on the perceived emotional weight of *I love you* aligns with previous findings (Dewaele, 2008; Jahangard & Holderread, 2013; Ożańska-Ponikwia, 2016).

However, adding L1-related predictors (*CoA L1* and *perception emotional L1*) significantly improved the model's predictive capacity, underscoring the crucial role of L1 in the emotional linguistic repertoire of TCKs. Their emotional perception of *I love you* in L1 is tied to their past, childhood memories (Marian & Neisser, 2000), roots, and a naturalistic CoA characterised by its high emotional resonance (Harris et al., 2006). Moreover, L1 maintains influence in the present by being the primary language at home and connecting TCKs to their extended family (Tannenbaum, 2005). This usage preserves HL as an emotional language, irrespective of self-reported proficiency. Qualitative data evidenced the complex interplay between language proficiency, love expression, and perception within this group. While the link between love expression and language proficiency was evident, the connection to the perceived emotional weight of *I love you* was less apparent. Participants emphasised English's emotional versatility in expressing love, contrasting it with their L1. Some even suggested that English transcended linguistic significance, contributing to personal identity and well-being (De Houwer, 2020) beyond their L1.

Conversely, some interviewees perceived *I love you* to hold equal emotional weight in LX English (self-reported dominant language) and L1, despite limited dominance in L1. This showcases the transcendence of linguistic significance beyond mere language mastery (Dewaele, 2011; Lorette & Dewaele, 2022) and suggests that the ability to perceive emotional intensity may develop separately from linguistic proficiency (Dewaele et al., 2023; Toivo et al., 2023). The cognitive and emotional aspects of perceiving *I love you* in this population may be influenced by their unique past experiences, including childhood memories, interlocutors, and the language of emotional events (Marian & Neisser, 2000).

Nonetheless, the results demonstrate that TCKs can fit both emotion scripts harmoniously (De Leersnyder et al., 2020), utilising languages in a complementary manner (Grosjean, 1997) that effectively meets their needs. By drawing on diverse linguistic resources, TCKs can uniquely express and perceive love. This helps explain intra- and inter-variability in TCKs' use of their emotional repertoire when transitioning from monolingual to bi-multilingual speech modes (Grosjean, 1985). Both their L1(s) and LX English provide unique frameworks for conceptualising emotions and a set of emotion concepts (Pavlenko, 2008) that may or may not overlap (Altarriba, 2003; Wierzbicka, 2004), as indicated by the interview participants. Interestingly, this contrasts with previous findings emphasising either emotional resonance (Tannenbaum & Tseng, 2015) or relegation of LX English to the academic domain (Jeon, 2022) in this population.

6.2.2. Influence of the sociobiographical cluster on the sociolinguistic cluster

The relationship between the sociobiographical and language profile differs from that of the sociolinguistic cluster. TCKs experience a bidirectional connection between their socialisation contexts: the third culture (expatriate community during their *years enrolled in international school*) and the domestic environment, represented by parents, whether engaged in an intercultural relationship (Figure 3).

When a TCK relocates (*age of first move*), it often involves enrolling in an international school (*years in international school*) and establishing new English-speaking networks, including interactions with domestic staff (in some cases) and within the school and expatriate community. These experiences result in frequent English use across different domains, which can have a ripple effect on TCKs' FLP, shaping the acceptance of English usage and ultimately determining the frequency of use of L1 (Figure 3).

TCKs generally perceive their L1(s) as embodied (Pavlenko, 2005); the intriguing part is the potential for LX English to also be considered embodied, given that 83% of parents are not L1 English users and nearly seven out of ten have never lived in an English-speaking country. This suggests that emotional input stems from TCKs' deep immersion in the "expatriate bubble" and its linguistic reliance on English, fostering both higher proficiency and emotional use of LX (Degner et al., 2011; Dewaele, 2010). Frequent and intense exposure to English among TCKs enhanced familiarity with the *I love you* emotion script, rendering LX English an embodied language. This aligns with previous studies (Dewaele, 2008; Ożańska-Ponikwia, 2016), emphasising the relationship between strong socialisation in LX and emotional expression (LX), as well as how the intensity and frequency of exposure heightens English emotional resonance (Dewaele et al., 2023). Therefore, despite collinearity removal, *degree of*

socialisation in LX influence persists indirectly, evident in its implicit representation through *frequency of language use*.

Furthermore, this extensive societal integration prompts adjustments in TCKs' emotional linguistic repertoire (Pavlenko, 2012), potentially leading to the incorporation of LX English into their emotional linguistic repertoire or manifesting as *emotional acculturation* (De Leersnyder, 2017). This study found that most participants incorporated LX English into the emotional domain, while some TCKs with Danish parent(s) exhibited emotional acculturation, prioritising LX emotional perception over L1 Danish. This suggests a process of secondary socialisation (Bayley & Schecter, 2003), where TCKs L1 Danish users experienced a conceptual shift (Pavlenko, 2004) in their perception of *I love you* towards LX English regardless of the FLP at home (Appendix II, Danish FLP sample description). This aligns with studies indicating that extensive immersion can gradually reshape the emotional perception of LX over time (Dewaele, 2011; Ożańska-Ponikwia, 2014).

Additionally, interview excerpts highlighted the disparities in the cultural scripts (Pavlenko, 2008) of *I love you* between some L1(s) and its frequent use in the expatriate community. TCKs manage these cultural tensions by capitalising on their use of this phrase, which they perceive as offering more flexibility and less seriousness compared to their L1(s) in certain contexts. Moreover, the unofficial L2 English status in Denmark (Lønsmann et al., 2022) and its widespread use in Danes' daily lives (Muñoz et al., 2018) may have reinforced a heightened perception of *I love you* in English, a notion supported by participants citing media influence, such as TV and films, amplifying its emotional sensitivity (Dewaele et al., 2021).

7. Conclusion

The phrase *I love you* carries heightened emotional weight for TCKs as a blend of L1+LX English, reflecting an ideal fusion aligned with an international mindset, where tensions between maintaining L1 identity and adopting a cultural hybrid are absent, showcasing TCKs' translanguaging identity (Rodríguez-Bernal et al., 2023). TCKs effortlessly navigate the emotion scripts of their home language and those of the third culture, indicating a dynamic emotional construction shaped by past and present experiences within these cultural contexts (De Leersnyder & Pauw, 2022). The ongoing need for both L1 and LX influences the constantly evolving linguistic system. TCKs' extensive exposure to English not only impacts their perception and use of L1, but also fosters a blending of the caregivers' culture with that of the expatriate community (third culture). The heightened emotional perception of *I love you* in LX English reflects emotional acculturation (De Leersnyder, 2017), a product of sustained contact within the "third culture" community. This group-centric acculturation

contributes to the process's uniqueness (Rodríguez-Bernal et al., 2023). TCKs can be viewed as bi-multicultural, combining features from various cultures (Grosjean, 2019). This blending leads TCKs to navigate a situational continuum from monocultural (mostly at home) to bi-multicultural behaviour, triggered by interactions with other bi-multicultural individuals at the international school and expatriate community, resulting in translanguaging spaces (Li, 2018).

In conclusion, TCKs' emotional language use exemplifies the fusion of past experiences with the present situation. The heightened emotional resonance of *I love you* attached to L1 and LX (English) signifies the convergence of two emotional spheres: the world shaped by their nomadic English-dominated lifestyle and the world anchored in their L1, connecting them to their past, family, and cultural heritage. This embodies the essence of TCKs: a harmonious blend of linguistic worlds coexisting within them.

The study's findings may not generalise to all TCK populations due to the small sample size of two out of three schools and the stringent eligibility criteria focused on traditional TCKs. Furthermore, the use of MLR models, which require ample observations for accurate predictions, may further constrain generalisability.

Future research on TCKs' linguistic dynamics could benefit from longitudinal studies exploring diachronic variation within the family domain and the application of new instruments to measure emotional resonance in bi-multilinguals (Toivo et al., 2023). Given the central role that family plays in TCKs' linguistic development, expanding studies on the phrase *I love you* should involve collecting data from TCKs' parents and siblings. Additionally, examining the various communication modes and contexts for conveying the love expression is crucial. Further longitudinal and multi-site research on TCKs and expatriate families could provide valuable insights into the lasting effects of nomadic lifestyles on linguistic adaptation.

---

[1] L2 represents the order of language acquisition.

[2] Source Education First: https://www.ef.com/assetscdn/WIBIwq6RdJvcD9bc8RMd/cefcom-epi-site/reports/2022/ef-epi-2022-english.pdf

[3] In Denmark, the minimum consent age for research participation is 15.

[4] Interview excerpts use pseudonyms in italics for direct quotes, whereas survey responses are identified by an informant (INF) number and quotation marks.

Acknowledgements

I would like to express my gratitude to all the students who completed the questionnaire and agreed to be interviewed for their participation, wherever they are. I extend my sincere appreciation to the anonymous reviewers for their invaluable comments and suggestions, which contributed to enhancing the quality of this manuscript.

8. References

Altarriba, J. (2003). Does *cariño* equal "liking"? A theoretical approach to conceptual non-equivalence between languages. *The International Journal of Bilingualism*: Cross-Linguistic Studies of Language *Behavior,* 7(3), 305-322. https://doi.org/10.1177/13670069030070030501

Bayley, R., & Schecter, S. R. (2003). Introduction: Language socialization in bilingual and multilingual societies. In R. Bayley, & S. R. Schecter (Eds.), *Bilingual education and bilingualism* (pp.1-6). Clevedon: Multilingual Matters.

Benjamin, S. (2017). "*People who don't live what we live, don't understand*": *Youths' experiences of hypermobility*. [Doctoral dissertation, University of Helsinki, Turku, Finland. Migration Institute of Finland].

Bowers, J., Metts, S., & Duncanson, T. W. (1994). Emotion in interpersonal communication. In M. Knapp, & G. R. Miller (Eds.), *Handbook of interpersonal communication* (pp. 500-550). Beverly Hills, CA: Sage.

Brinkmann, S., & Kvale, S. (2014). *Interviews: Learning the craft of qualitative research interviewing* (3rd ed.). Los Angeles: Sage.

Caldwell-Harris, C. L., Tong, J., Lung, W., & Poo, S. (2011). Physiological reactivity to emotional phrases in Mandarin—English bilinguals. *The International Journal of Bilingualism: Cross-Disciplinary, Cross-Linguistic Studies of Language Behavior, 15*(3), 329-352. https://doi.org/10.1177/1367006910379262

Carder, M. (2007). *Bilingualism in international schools: A model for enriching language education*. Clevedon, UK: Multilingual Matters.

Curdt-Christiansen, X. L. (2022). Family language policy and school language policy: Can the twain meet? *International Journal of Multilingualism, 19*(3), 466-475. https://doi.org/10.1080/14790718.2022.2050242

De Houwer, A. (2020). Why do so many children who hear two languages speak just a single language? *Zeitschrift Für Interkulturellen Fremdsprachenunterricht, 25*(1), 7.

De Houwer, A., & Bornstein, M. H. (2016). Bilingual mothers' language choice in child-directed speech: Continuity and change. *Journal of Multilingual and Multicultural Development, 37*(7), 680-693. https://doi.org/10.1080/01434632.2015.1127929

De Leersnyder, J. (2017). Emotional acculturation: A first review. *Current Opinion in Psychology, 17,* 67-73. https://doi.org/10.1016/j.copsyc.2017.06.007

De Leersnyder, J., Kim, H. S., & Mesquita, B. (2020). My emotions belong here and there: Extending the phenomenon of emotional acculturation to heritage culture fit. *Cognition and Emotion, 34*(8), 1573-1590. https://doi.org/10.1080/02699931.2020.1781063

De Leersnyder, J., & Pauw, L. (2022). Emotions in social situations. In G. L. Schiewer, J. Altarriba, & B. C. Ng (Eds.), *Language and Emotion* (pp. 825-854). Berlin: De Gruyter Mouton.

De Mejia, A-M. (2002). *Power, prestige, and bilingualism: International perspectives on elite bilingual education*. Clevedon: Multilingual Matters. https://doi.org/10.21832/9781853595929

Degner, J., Doycheva, C., & Wentura, D. (2011). It matters how much you talk: On the automaticity of affective connotations of first and second language words. *Bilingualism Language and Cognition, 15*(1), 181-189. https://doi.org/10.1017/S1366728911000095

Dewaele, J-M. (2008). The emotional weight of I love you in multilinguals' languages. *Journal of Pragmatics, 40*(10), 1753-1780. https://doi.org/10.1016/j.pragma.2008.03.002

Dewaele, J-M. (2010). *Emotions in multiple languages*. Basingstoke: Palgrave Macmillan. https://doi.org/10.1057/9780230289505

Dewaele, J-M. (2011). Self-reported use and perception of the L1 and L2 among maximally proficient bi-and multilinguals: A quantitative and qualitative investigation. *International Journal of the Sociology of Language, 2011*(208), 25-51. https://doi.org/10.1515/ijsl.2011.011

Dewaele, J-M. (2018). Why the dichotomy 'L1 versus LX user' is better than 'Native versus non-native speaker'. *Applied Linguistics, 39*(2), 236-240. https://doi.org/10.1093/applin/amw055

Dewaele, J-M., Alsuhaibani, Y., Altalhab, S., & Alghamdi, W. (2023). How frequency and intensity of exposure to a foreign language boosts its emotional resonance. *International Journal of Multilingualism, ahead-of-print,* 1-15. https://doi.org/10.1080/14790718.2023.2233551

Dewaele, J-M., Lorette, P., Rolland, L., & Mavrou, I. (2021). Differences in emotional reactions of Greek, Hungarian, and British users of English when watching television in English. *International Journal of Applied Linguistics, 31*(3), 345-361. https://doi.org/10.1111/ijal.12333

Dewaele, J-M., & Nakano, S. (2013). Multilinguals' perceptions of feeling different when switching languages. *Journal of Multilingual and Multicultural Development, 34*(2), 107-120. https://doi.org/10.1080/01434632.2012.712133

Dewaele, J-M., & Pavlenko, A. (2001-2003). *Web questionnaire bilingualism and emotions*. https://www.academia.edu/2134800/Web_questionnaire_bilingualism_and_emotion

Education First. (2022). EF English Proficiency Index: *A Ranking of 111 Countries and Regions by English Skills*. https://www.ef.com/assetscdn/WIBIwq6RdJvcD9bc8RMd/cefcom-epi-site/reports/2022/ef-epi-2022-english.pdf

Eidse, F., & Sichel, N. (2004). Introduction. In F. Eidse, & N. Sichel (Eds.), *Unrooted childhoods: Memoirs of growing up global* (pp. 1-6). London: Hachette.

Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh, Scotland: Oliver and Boyd.

Freedman, D. (2009). Issues in Statistical Modelling: In D. Freedman (Ed.), *Statistical models: Theory and practice* (pp. 209-218). New York: Cambridge University Press.

García, O., & Li, W. (2014). *Translanguaging: Language, bilingualism, and education*. London, UK: Palgrave Macmillan. https://doi.org/10.1057/9781137385765

Gareis, E., & Wilkins, R. (2011). Communicating love A sociocultural perspective. *Annals of the International Communication Association, 35*(1), 199-239. https://doi.org/10.1080/23808985.2011.11679117

Grosjean, F. (1985). The bilingual as a competent but specific speaker-hearer. *Journal of Multilingual and Multicultural Development, 6*(6), 467-477. https://doi.org/10.1080/01434632.1985.9994221

Grosjean, F. (1997). The bilingual individual. Interpreting (Amsterdam), *2*(1/2). 163-187. https://doi.org/10.1075/intp.2.1-2.07gro

Grosjean, F. (2019). *A journey in languages and cultures: The life of a bicultural bilingual.* New York: Oxford University Press.

Harris, C. L. (2004). Bilingual speakers in the lab: Psychophysiological measures of emotional reactivity. *Journal of Multilingual and Multicultural Development, 25*(2-3). 223-247. https://doi.org/10.1080/01434630408666530

Harris, C. L., Gleason, J. B., & Ayçiçeği, A. (2006). When is a first language more emotional? Psychophysiological evidence from bilingual speakers. In A. Pavlenko (Ed.), *Bilingual minds: Emotional experience, expression, and representation.* (pp. 257-283) Clevedon, UK: Multilingual Matters.

IMB Corp. (2021). *IBM SPSS Statistics for Windows, Version 28.0.* IBM Corp.

Jahangard, A. & Holderread, S. (2013). The Emotional Value of the Phrase "I love you" for Iranian Bilinguals. *Journal of Language, Culture, and Translation, 2*(1), 39-63.

Jeon, A. (2022). Growing up (un)bounded: Globalization, mobility and belonging among Korean third culture kids. *International Multilingual Research Journal, 16*(1). 65-77. https://doi.org/10.1080/19313152.2021.1951941

Johnson, R. B., & Christensen, L. B. (2020). *Educational research: Quantitative, qualitative, and mixed approaches* (7th ed.). Los Angeles, CA: SAGE.

Kopeliovich, S. (2013). Happylingual: A family project for enhancing and balancing multilingual development. In M. Schwartz, & A. Verschik (Eds.), *Successful family language policy: Parents, children, and educators in interaction* (pp. 249-275). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-7753-8_11

Lê, J. K., Tissington, P. A., & Budhwar, P. (2010). To move or not to move–a question of family? *The International Journal of Human Resource Management, 21*(1), 17-45. https://doi.org/10.1080/09585190903466848

Li, W. (2018). Translanguaging as a practical theory of language. Applied Linguistics, *39*(1), 9-30. https://doi.org/10.1093/applin/amx039

Lijadi, A. A., & Van Schalkwyk, G. J. (2017). Place identity construction of third culture kids: Eliciting voices of children with high mobility lifestyle. *Geoforum, 81,* 120-128. https://doi.org/10.1016/j.geoforum.2017.02.015

Lønsmann, D., Mortensen, J., & Thøgersen, J. (2022). Is English still a foreign language in Denmark? A question of collective linguistic identity. NyS, *Nydanske Sprogstudier, 1*(61). 126-179.

Lorette, P., & Dewaele, J-M. (2022). Interpersonal perception of emotional intensity by English first (L1) and foreign (LX) language users in audio(visual) communication. *International Journal of Multilingualism*, *21*(2), 1116-1133. https://doi.org/10.1080/14790718.2022.2144326

Marian, V., & Neisser, U. (2000). Language-dependent recall of autobiographical memories. *Journal of Experimental Psychology: General, 129*(3), 361-368. https://doi.org/10.1037/0096-3445.129.3.361

Meyer, H. A. (2021). *The global imaginary of international school communities*. Cham: Springer International Publishing.

Muñoz, C., Cadierno, T., Casas, I., & Casas, I. (2018). Different starting points for English language learning: A comparative study of Danish and Spanish young learners: Different starting points. *Language Learning, 68*(4), 1076-1109. https://doi.org/10.1111/lang.12309

Ożańska-Ponikwia, K. (2014). The influence of immersion in the L2 culture on perception of the L1 culture-specific emotion of tęsknota. *International Journal of Bilingualism, 20*(2), 116-132. https://doi.org/10.1177/1367006914537728

Ożańska-Ponikwia, K. (2016). Expression and perception of emotions by Polish-English bilinguals *I love you* vs. *Kocham Cię*. *International Journal of Bilingual Education and Bilingualism, 22*(4), 493-504. https://doi.org/10.1080/13670050.2016.1270893

Pavlenko, A. (2004). 'Stop Doing That, Ia Komu Skazala!': Language Choice and Emotions in Parent-Child Communication. *Journal of Multilingual and Multicultural Development, 25*(2-3), 179-203. https://doi.org/10.1080/01434630408666528

Pavlenko, A. (2005). *Emotions and multilingualism.* Cambridge: Cambridge University Press.

Pavlenko, A. (2008). Emotion and emotion-laden words in the bilingual lexicon. *Bilingualism: Language and Cognition, 11*(2), 147-164. https://doi.org/10.1017/S1366728908003283

Pavlenko, A. (2012). Multilingualism and emotions. In M. Martin-Jones, A. Blackledge, & A. Creese (Eds.), *The Routledge handbook of multilingualism* (pp. 454-469). Abingdon, UK: Routledge. https://doi.org/10.4324/9780203154427

Pollock, D. C., Van Reken, R., & Pollock, M.V. (2017). *Third Culture Kids: Growing up among worlds*. Boston, MA: Nicholas Brealey Publishing.

Rodríguez-Bernal, L., Cadierno, T., Doquin de Saint-Preux, A., & Lauridsen, J. T. (2023). Third Culture Kids in Denmark: Exploring code-switching patterns according to interlocutors and topic emotionality. *Journal of Multilingual and Multicultural Development*, 1-18. https://doi.org/10.1080/01434632.2023.2277862

Rydenvald, M. (2018). Who speaks what language to whom and when – rethinking language use in the context of European Schools. *International Journal of the Sociology of Language*, *2018*(254), 71-101. https://doi.org/10.1515/ijsl-2018-0034

Schaetti, B., & Ramsey, S. (1999). "The expatriate family: Practicing practical leadership". *Mobility. Employee Relocation Council.* Retrieved from: https://transitiondynamics.wordpress.com/resources-and-products/articles-and-publications/the-expatriate-family-practicing-practical-leadership/

Sichel, N., Orr, E. N., & Eidse, F. (Eds.). (2011). *Writing out of limbo: International childhoods, global nomads and third culture kids.* Newcastle-upon-Tyne: Cambridge Scholars Publishing.

Tabatchnick, B. G., & Fidell, L.S. (2007). *Using Multivariate Statistics* (7th ed.). Boston, MA: Pearson.

Tan, E. C., Wang, K. T., & Cottrell, A. B. (2021). A systematic review of third culture kids empirical research. *International Journal of Intercultural Relations*, *82*, 81-98. https://doi.org/10.1016/j.ijintrel.2021.03.002

Tannenbaum, M. (2005). Viewing family relations through a linguistic lens: Symbolic aspects of language maintenance in immigrant families. *Journal of Family Communication*, 5(3), 229-252. https://doi.org/10.1207/s15327698jfc0503_4

Tannenbaum, M., & Tseng, J. (2015). Which one is Ithaca? Multilingualism and sense of identity among Third Culture Kids. *International Journal of Multilingualism, 12*(3), 276-297. https://doi.org/10.1080/14790718.2014.996154

Toivo, W., Scheepers, C., & Dewaele, J. (2023). RER-LX: A new scale to measure reduced emotional resonance in bilinguals' later learnt language. *Bilingualism: Language and Cognition,* 1-13. https://doi.org/10.1017/S1366728923000561

Wierzbicka, A. (1999). *Emotions across languages and cultures: Diversity and universals*. Cambridge, UK: Cambridge University Press.

Wierzbicka, A. (2004). Preface: Bilingual lives, bilingual experience. *Journal of Multilingual and Multicultural Development, 25*(2-3), 94-104. https://doi.org/10.1080/01434630408666523

Appendix 1: Background information on participants

Sociobiographical cluster

Predictor	Description	*n*	%	M	*SD*	Mode
Age of first move	0-2 years	96	34.7			
	3-7 years	77	27.9	6	5	0
	8-12 years	41	14.9			
	13-18 years	62	22.5			
Countries TCK resided		276		3	0.9	3
Years in international school		276		8	4	4
Intercultural parents	yes	108	39.1			
	no	168	60.9			
Parent(s) L1 Danish	yes	82	29.7			
	no	194	70.3			
Additional sociobiographical information						
School Identifier	International school 1	200	72.5			
	International school 2	39	14.1			
	International school 3	37	13.4			
Place of birth different from parents	yes	100	36.2			
	no	176	63.8			
Binational couple	yes	112	40.6			
	no	164	59.4			
Parent(s) L1 English	yes	47	17.0			
	no	229	83.0			
Resided English speaking country	yes	90	32.6			
	no	186	67.4			
Parent(s)' profession*	Multinational	160	58			
	Diplomats and Governmental	34	12.3			
	International education	26	9.4			
	NGOs/Humanitarian Aid	9	3.3			
	Other	45	16.3			

*Two participants did not respond to this question.

Language profile cluster

Predictor	Description	*n*	%	*M*	*SD*
Learning history					
AoA L2	0-2	117	42.4		
	3-7	115	41.7		
	8-18	44	15.9		
CoA L1	naturalistic	137	49.6		
	mixed and instructed	139	50.4		
CoA L2	naturalistic	77	27.9		
	mixed and instructed	199	72.1		
Perception emotional	L1	high	190	68.8	
		moderate	43	15.6	
		low	43	15.6	
	L2	high	148	53.6	
		moderate	59	21.4	
		low	69	25.0	
Linguistic competency					
Self-reported dominant language	L1	126	45.7		
	L1+LX	108	39.1		
	LX	42	15.2		
Average self-reported oral skills	L1	276		4.7	0.6
	L2	276		4.6	0.7
Additional linguistic information					
L1	Danish	55	19.9		
	English	35	12.7		
	French	21	7.6		
	German	16	5.8		
	Italian	13	4.7		

Predictor	Description	*n*	%	*M*	*SD*
L2	English	188	68.0		
	Danish	19	6.9		
	French	12	4.3		
	Spanish	12	4.3		
	German	7	2.5		
Languages acquired	276		4	0.9	
	bilinguals	31	11.2		
	trilinguals	92	33.3		
	quadrilinguals	98	35.5		
	pentalinguals	55	19.9		
Languages spoken at home	276		2	0.6	
Number of dominant language(s)	276		2	0.6	

Qualitative description of the sample

Name[a]	Age/ AFM[b]	Years int. school[c]	Countries resided for at least one year	Years in Denmark	Languages acquired	Language(s) spoken at home	Dominant languages
Group 1: strict FLP, societal-non societal approach							
Elin	16/1	14	Sweden Egypt Denmark	1	L1 Swedish L2 English L3 Spanish	Swedish	L2
Renata	16/5	7	Portugal Spain United States Belgium Denmark	4	L1 Portuguese L2 Spanish L3 English L4 French	Portuguese	L3
Group 2: flexible FLP, *Happylingual approach*							
Lena	17/8	9	United States China Germany Denmark	6	L1 German L2 English L3 Spanish	German English Spanish	L1 L2 L3

Name[a]	Age/ AFM[b]	Years int. school[c]	Countries resided for at least one year	Years in Denmark	Languages acquired	Language(s) spoken at home	Dominant languages
Sergei	18/7	6	Russia UAE Denmark	2	L1 Russian L2 English L3 Spanish L4 Danish	Russian English	L1 L2
Group 3: LX English dominant with varying FLP approaches							
OPOL approach							
Mads	17/0	16	Malaysia China Vietnam Thailand Denmark	1	L1 Danish L2 English L3 Chinese L4 Malay L5 Spanish	English Danish Malay	L1 L2 L4
Strict FLP, societal-non societal approach (both parents Danish)							
Morten	17/12	7	UK Denmark	11	L1 Danish L2 English L3 French	Danish	L2
Freja	18/8	8	India UAE Russia Denmark	10	L1 Danish L2 English L3 French L4 Arabic L5 German	Danish	L1 L2

[a] All names used are aliases.
[b] Age of first move.
[c] Years enrolled in international school.

Background information on participants with Danish parents

Descriptive statistics of TCKs with Danish parent(s) n = 82

Danish status	*n*	%	Variable	Description	*n*	%	*M*	*SD*	Mode
Both parents	33		Place of birth different from parents	yes	9	27.3			
				no	24	72.7			
			Age of first move	0-2	9	23.7			
				3-7	8	24.2	1.4	1.1	0
				8-12	7	21.2			
				13-18	9	27.3			
			L1	Danish	31	93.9			
			L2	English	29	87.9			
			Simultaneous AoA	L1 + English	7	21.2			
			Years international school				8.3	4.1	4
			Countries resided				2.8	1	2
			Years living in Denmark				8.6	4.1	3
			English is spoken at home	yes	9	27.3			
				no	24	72.7			
			Self-reported dominant language	L1	9	27.3			
				L1+LX	15	45.5			
				LX	9	27.3			
One parent (no L1 English)	33		Place of birth different from parents	yes	18	54.5			
				no	15	45.5			
			Age of first move	0-2	14	42.4			
				3-7	10	30.3	1.0	1.1	0
				8-12	2	6.1			
				12-18	7	21.2			
			L1	Danish	12	36.4			

Danish status	*n*	%	Variable	Description	*n*	%	*M*	*SD*	Mode
One parent (no L1 English)	33		L2	English	17	51.5			
			Simultaneous AoA	Ln[1] + English	20	60.6			
			Years international school				7.6	4.3	4
			Countries resided				3	0.7	3
			Years living in Denmark				4.7		2
			English is spoken at home	yes	20	60.6			
				no	13	39.4			
			Self-reported dominant language	L1	14	42.4			
				L1+LX	12	36.4			
				LX	7	21.2			
One parent (L1 English)	16		Place of birth different from parents	yes	6	62.5			
				no	10	37.5			
			Age of first move	0-2	9	56.3			
				3-7	–	–			
				8-12	3	18.8			
				13-18	4	25.0			
			L1	Danish	9	56.3			
				English	6	37.5			
			L2	English	10	62.5			
				Danish	4	25.0			
			Simultaneous AoA	Ln[1] + English	9	56.3			
			Years international school				8.1	4.6	5
			Countries resided				2.2	0.5	2

Danish status	*n*	%	Variable	Description	*n*	%	*M*	*SD*	Mode
One parent (L1 English)	16		Years living in Denmark				5.3	3.6	2
			English is spoken at home	yes	15	6.3			
				no	1	93.8			
			Self-reported dominant language	L1	10	62.5			
				L1+LX	6	37.5			
				LX	–	–			

[1] Mother's or father's language.

Sociolinguistic cluster

Predictor	Description	*n*	%	M	SD
Network interlocutors	L1	general	156	56.5	
		close	120	43.5	
	L2	general	175	63.4	
		close	101	36.6	
Years living in Denmark	276		5.4	4.1	
Frequency of use	L1	high	256	92.8	
		low	20	7.2	
	L2	high	233	84.4	
		low	43	15.6	
Degree of socialisation	LX	High	48	17.4	
		Equal to L1	146	52.9	
		Low	81	29.3	
Household linguistic mode	monolingual	149	54.0		
	bi-multilingual	127	46.0		
English spoken at home	yes	144	52.2		
	no	132	47.8		

Appendix II: FLP background information on participants

General FLP sample description n = 276

Language mode	*n*	%	Description	*n*	%	Parents' English status	*n*	%	% sample
Linguistic mode									
Monolingual	149	54.0	Only English	32	11.6	L1	18	56.3	6.5
						LX	14	43.7	5.1
			Only parents' L1 (shared language)	117	42.4	–	–	–	–
Bilingual	127	46.0	English + parents' L1	95	34.4	L1	28	29.4	10.1
						LX	67	70.6	24.3
			Only L1x[1] + L1y[2]	14	5.1	–	–	–	–
Multilingual			English + L1x +L1y	18	6.5	L1	1	5.5	0.4
						LX	17	94.5	6.1
English spoken at home									
Yes	145	52.5	Only English	32	11.6	L1	18	56.3	6.5
						LX	14	43.7	5.1
			English + parents' L1	95	34.4	L1	28	29.4	10.1
						LX	67	70.6	24.3
			English + L1x +L1y	18	6.5	L1	1	5.5	0.4
						LX	17	94.5	6.1
No	131	47.5	Only parents' L1 (shared language)	117	42.4	–	–	–	–
			Only L1x + L1y	14	5.1	–	–	–	–

[1] Mother's language, [2] Father's language

Danish FLP sample description n =82

	n	Language mode at home	*n*		%	Description	Parents' English status	*n*	%
Danish status									
Both parents	33 (40.2%)	Monolingual	25 (75.7%)	24	72.7	Only DA[1]	–	–	–
				1	3.0	Only EN [2]	LX	1	3.0
		Bilingual	7 (21.2%)	7	21.2	DA + EN	LX	7	21.2
		Multilingual	3 (3.0%)	1	3.0	DA+ EN+ Ln[3]	LX	1	3.0
One parent	49 (59.7%)	Monolingual	14 (29.1%)	1	7.1	Only DA	–	–	–
				3	21.4	Only EN	L1	3	21.4
				3	21.4		LX	3	21.4
				7	29.2	Only Ln	–	–	–
		Bilingual	29 (58.3)	7	24.2	DA+Ln	–	–	–
				19	65.5	DA +EN	L1	13	68.4
							LX	6	31.6
				3	10.3	Ln + EN	LX	3	10.3
		Multilingual	6 (12.5%)	6	100	DA+ EN+Ln	LX	6	100

[1] DA= Danish, [2] EN = English, [3] Ln = Mother's or father's language.

Trainees' beliefs about the use of cinema as a tool for EFL

Estefanía Sánchez-Auñón
Department of Didactics of Modern Languages and Literatures
University of Murcia
estefania.sanchez8@um.es

Pedro Antonio Férez Mora
Department of Didactics of Modern Languages and Literatures
University of Murcia
paferez@um.es

Fuensanta Monroy Hernández
Department of Research Methods in Education
University of Murcia
fuensanta.monroy@um.es

Abstract

Despite the growing presence of cinema in the classroom and the multiple benefits it has been proved to have for the teaching of English as a foreign language (EFL), the implementation of films in the English lessons is still an under-researched area, particularly, as regards instructors' views on the adequacy of this material. Thus, the present study delves into trainees' beliefs (N = 121) regarding the potential of cinema for EFL before and after analysing a film-based lesson plan. Questionnaires, open-ended questions, and focus groups were employed to obtain quantitative and qualitative data on the perceived linguistic, intercultural, and motivational advantages of films. The results reveal that, at the outset of the research, trainees held cinema especially useful for eliciting linguistic and intercultural knowledge. However, they were unaware of how to exploit this textual modality and considered some of its characteristics problematic for class use. In the post-analysis stage, the participants assigned similar rates of importance to the three analysed dimensions, reflecting an even more favourable attitude towards this resource. This positive change trend seems to stem from a broadened awareness of the proper exploitation of film after examining the lesson plan provided. Pedagogical and research implications are also defined.

Keywords: EFL; cinema; benefits; trainees' opinion; teacher education.

Resumen

A pesar de la creciente presencia del cine en el aula y de los numerosos beneficios que tiene para la enseñanza del inglés como lengua extranjera (ILE), la implementación de películas es un tema poco investigado, especialmente, en cuanto a las percepciones de los docentes sobre la adecuación de este material. Así, el presente estudio explora las opiniones del profesorado de inglés en formación (N = 121) sobre el potencial del cine para la enseñanza de ILE antes y después de analizar una unidad curricular basada en una película. Se emplearon cuestionarios, preguntas de respuesta abierta y grupos focales para obtener datos cuantitativos y cualitativos sobre las ventajas lingüísticas, interculturales y motivacionales de las películas. Antes del análisis, los participantes consideraban que el cine era especialmente útil para generar conocimiento lingüístico e intercultural. No obstante, no sabían cómo explotar esta modalidad textual, considerando problemática la implementación en clase de algunas de sus características. Tras el análisis, las tres dimensiones recibieron valoraciones similares, reflejando, así, una actitud todavía más favorable hacia este recurso. Este cambio positivo parece deberse a un mayor conocimiento sobre la correcta explotación de películas tras examinar la unidad curricular. Se establecen también implicaciones pedagógicas y de investigación.

Palabras clave: Enseñanza de inglés como lengua extranjera; cine; beneficios; opinión de futuros profesores; formación docente.

1. Introduction

Cinema is rarely exploited in the contemporary English as a foreign language class owing to time, cost, and syllabus constraints, the difficulty of film selection, and, particularly, teachers' lack of pedagogical training (Carney & Foss, 2008; Ismaili, 2013; Kabooha, 2016; Shahani et al., 2014). Most English instructors regard film screening as a backup activity, and, therefore, they simply play a whole film and let students watch it passively, as a time filler or as a reward for good behaviour (Albiladi et al., 2018; Hobbs, 2006). Learners have noticed this improper implementation of cinema and urge teachers to make a more frequent and comprehensive use of films in English lessons (Kabooha, 2016; Shahani et al., 2014; Tuncay, 2014; Yue, 2019).

Considering teachers' and students' interest in the cinema-based approach, the scarcity of research on this topic comes as a surprise. Much of the existing academic production on the use of films for EFL is theoretical and didactic, and the few empirical studies conducted to date are inconclusive due to multiple methodological limitations. This applies especially to the examination of teachers' and students' perceptions. Given the paucity of empirical research in the field, exploring trainees' views about the advantages existing research has conceptualised when it comes to the use of cinema

for EFL emerges as a relevant line of research. As future teachers, and, therefore, instrumental stakeholders in formal educational processes, trainees might contribute valuable insights into the viability of cinema-based EFL. In fact, their perceived advantages and disadvantages, some of which might also be misperceptions, might help develop a pedagogy of cinema-based EFL. Against this background, the present study explores EFL trainees' opinions on a cinema-based lesson plan. Questionnaires, open-ended questions and focus groups were used to this end.

2. Literature review

2.1. The advantages of using cinema for EFL

2.1.1. Linguistic advantages

It has been argued that films may help EFL learners to enhance their linguistic ability since they can practice the four English-language skills while learning grammar, vocabulary, and pragmatics (Canning-Wilson, 2000; Fernández-Guerra, 2013; Kaiser, 2011). According to several researchers, cinema allows students to hone their listening skills since the paralinguistic cues of films help them to better comprehend what they hear (Chapple & Curtis, 2000; Kabooha, 2016; Shahani et al., 2014), and also because learners can practice extensive listening as they are exposed to longer stretches of discourse than they are used to (Tuncay, 2014). English instructors have highlighted the great utility of the film-based approach to teaching listening skills (Aksu-Ataç & Günay-Köprülü, 2018), which has been confirmed by EFL learners in numerous studies (Chao, 2013; Chapple & Curtis, 2000; Shahani et al., 2014). Likewise, Chao's (2013) and Ismaili's (2013) investigations reported that, after watching films with subtitles, students' reading skills had improved as they had had the opportunity to read a large amount of text at normal speed.

EFL students have also acknowledged they have learnt to speak more fluently after attending cinema-based lessons (Aksu-Ataç & Günay-Köprülü, 2018; Chiu, 2012). This is because, in films, they can see native speakers interacting in authentic contexts and using varied dialects and accents (Argynbayev et al., 2014; Mahmoodi-Shahrebabaki, 2015; Tuncay, 2014). Furthermore, as students are interested in films, they are more willing to express their views on what they are watching, and even less-advanced and shy learners feel less anxious to participate in class (Chao, 2013; Kabooha, 2016). In Chapple and Curtis's research, students rated confidence when speaking in English as the most noticeable benefit of the film-based approach (2000). Academics have also supported the use of cinema to develop writing skills, mainly, on the grounds that it allows English teachers to design activities like writing a summary or a review of the film (Fluitt-Dupuy, 2001; Gallagher, 1988; Scacco, 2007). Besides, EFL students have

declared that, when cinema is used in the classroom, they can identify and reduce negative language transfer in writing tasks (Argynbayev et al., 2014).

Multiple scholars have alleged that cinema fosters lexical, grammatical, and pragmatic acquisition because films show students how words, linguistic structures, and discourse strategies are used in a real communicative situation, which facilitates meaning inference and memorisation (Albiladi et al., 2018; Ismaili, 2013; Kabooha, 2016; Shahani et al., 2014). Vocabulary learning is one of the advantages of the film-based approach EFL instructors and students value the most. They emphasise the varied range of English vocabulary in films, especially, colloquial expressions, which are not normally included in textbooks (Aksu-Ataç & Günay-Köprülü, 2018; Chapple & Curtis, 2000; Chiu, 2012; Christopley, 2017; Mahmoodi-Shahrebabaki, 2015). Similarly, Cetinavci's study (2019) reported how explicit instruction from dialogues in films, among other audio-visual materials, had led the Turkish EFL trainees in the experimental group to improve their capacity to interpret implied meaning in comparison to participants in the control group.

2.1.2. Intercultural advantages

Cinema also enhances intercultural awareness. As held by Argynbayev et al. (2014), Kabooha (2016), and Shahani et al. (2014), films show English-speaking people's behaviour, values, customs, social relationships, history, and food in a more visual and effective way than any other didactic material. Additionally, cinema may present complex cultural issues like racism and sexism, which helps EFL students to better comprehend the L2 society. A student in Albiladi et al.'s research stressed that she started to reflect more deeply on women's role in American society after watching American films featuring Julia Roberts (2018). When discovering such cultural factors in films, English students widen their knowledge of the target culture and feel more confident in intercultural exchanges (Chao, 2013).

Some academics affirm that cinema allows EFL students to become more conscious of their community. The links learners establish between what they see on the screen and the L1 culture result in a deeper understanding of the students' own context (Kaiser, 2011; Sturm, 2012). An English learner in Chao's study explained how comparing two similar tragedies in Poland and Taiwan while watching a film had helped them to better appreciate the current improved situation of their country (2013).

2.1.3. Motivational advantages

Most of the scant data on the film-based approach are related to the purported motivating nature of the seventh art. Experts in the field maintain that, since English

students are already familiar with this material and enjoy it frequently, films stimulate them to learn (Anas & Zakaria, 2019; Chapple & Curtis, 2000; Shahani et al., 2014). Christopley (2017) explains how watching cartoon films helped students to relax, and thus they were more enthusiastic about the vocabulary activities afterwards. EFL learners have confirmed this in several investigations (Chao, 2013; Ismaili, 2013).

Researchers have underlined additional reasons why cinema may be such a stimulating pedagogical tool in the EFL classroom: (i) the authenticity of the English language in films, contrasted with the adapted L2 input EFL learners normally receive in formal instruction (Albiladi et al., 2018; Anas & Zakaria, 2019; Chao, 2013; Fernández-Guerra, 2013; Stewart, 2006); (ii) the aesthetic pleasure students experience when watching films (Allan, 1985; Alwehaibi, 2015); (iii) the nontrivial topics cinema often depicts — like sexism, homophobia or environmental degradation, which encourage them to fight for a better world (Anas & Zakaria, 2019; Kaiser, 2011; Sturm, 2012); (iv) the development of creative and critical thinking skills (Kabooha, 2016; Lonergan, 1984; Shahani et al., 2014); and (v) the understanding of emotions, since students can analyse the actors' feelings and express their own affective responses (Berk, 2009; Jurkovič & Mertelj, 2015; Kaiser, 2011).

2.2. A need for more empirical research

By and large, the studies reviewed in the previous section have adopted a narrow focus for a series of reasons. First, the teaching sequences used to explore the participants' perceptions of cinema-based benefits were so fleetingly described that, in some cases, it was difficult to determine their purpose (Argynbayev et al. 2014; Chao, 2013; Chiu, 2012; Ismaili, 2013; Kabooha, 2016; Mahmoodi-Shahrebabaki, 2015; Shahani et al., 2014; Tuncay, 2014). Second, these investigations examined the participants' thoughts on only some of the benefits this method may have for EFL (Aksu-Ataç & Günay-Köprülü, 2018; Chiu, 2012; Christopley, 2017; Ismaili, 2013; Kabooha, 2016; Shahani et al., 2014). In contrast to the above-mentioned research, Sánchez-Auñón and Férez-Mora's (2021) study investigated the whole spectrum of advantages of cinema-based EFL as perceived by a high school teacher before and after observing a lesson plan whose contents were thoroughly detailed. The teacher's perception evolved from an anecdotal and minimal use of cinema in her lessons to a clear and nuanced awareness of the pedagogical potential of this resource along with a willingness to incorporate it into her teaching. Although the findings of the study were promising, its small sample leads to consider it as a pilot study.

Given these methodological shortcomings, further empirical research is needed on teachers' and students' views regarding the implementation of cinema in the EFL class. Targeting trainees' perceptions seems to be a priority because, among the sources

reviewed, only one study has targeted this population (Aksu-Ataç & Günay-Köprülü, 2018). However, the digital-native environment of these future teachers might bring up fresh and innovative insights on the use of film in EFL. At the same time, they might entertain misconceptions and limitations stemming from a still-forming grasp of what is feasible in a foreign language class (Goodnough et al., 2009). This study also intends to advance research in the field by: (i) tapping into the whole spectrum of cinema-based advantages as theoretically defined; (ii) using a pre- post-test design to move beyond one-shot studies, thus allowing for exploring change in the participants' perception; and (iii) using a combination of research instruments (a questionnaire, open questions and focus groups) which, to date, has not been applied in the field.

The following research questions guide this study:

1. What are trainees' views about the use of films in EFL?

2. Do trainees' views change after analysing a cinema-based lesson plan?

3. Method

3.1. Participants

The participants in the study were 121 trainee teachers (63% women and 37% men) pursuing a master's degree in EFL teacher training at a university in Southeast Spain. The sample included students from three groups (39, 40 and 41 students, respectively). The study was conducted in the innovation course of this academic programme in which there was a module on the use of film for EFL. The ages of the participants ranged from 22 to 39, with most in their mid-20s. All the participants had completed a degree in English or Translation Studies and had, at least, a C1 level according to the Common European Framework for Languages.

3.2. Data collection

The data collection instruments consisted of a questionnaire (Likert scale), two open-ended questions and a focus group conducted with volunteering participants. The three instruments were piloted with a group of 20 students at the same university. Participants did not express any concerns regarding the clarity of the instruments. This triangulation mixed-methods research design was selected on the grounds that its three-fold orientation leads to the collection of richer and more nuanced data from which robust findings can be derived (Plano Clark & Creswell, 2014).

The questionnaire (see Table 3 and Appendix 2) explored the trainees' beliefs concerning the potential of cinema as a resource for learning English before and after

the analysis of a cinema-based lesson plan. It contained 17 statements to be scored on a 4-point Likert scale, with 1 indicating complete disagreement and 4 full agreement. The questionnaire was adapted from that administered in Férez Mora et al. (2020), which tapped into students' perceptions regarding the use of a poem as a tool for EFL. These researchers concluded that the advantages of conducting EFL from literary texts could be grouped within a linguistic, an intercultural, and a motivational construct. That questionnaire was chosen to conduct this study because, with the exception of one of the items in the questionnaire, the advantages for literary texts surveyed in Férez Mora et al. (2020) overlapped with those which scholarly research had theoretically established for cinematic texts. Three adjustments were made to the original questionnaire. First, items 8 — "Poems elicit knowledge of the L2 culture" — and 9 — "Poems elicit knowledge of the L2 people" — were merged into one item given their similarity. Second, item 12 in Férez et al. (2020) — "Poems motivate students to learn EFL because they tend to be short texts" — was discarded as irrelevant to films. Third, another item was introduced which referred to the audio-visual nature of films (item 17: "Cinema is a useful resource for the teaching of English as a foreign language because it is audio-visual material").

Open Question 1 targeted trainees' raw views about the potential of cinema for EFL — "What is your view about the potential of cinema for EFL?" — while Open Question 2 focused on experienced change — "Has your perception of the potential of film for EFL changed after analysing the lesson plan?". Focus groups were initiated by a general prompt which was left as open as possible to foster free interaction (Dörnyei, 2007): "Once you have analysed the lesson plan provided, can you assess the usefulness of film as a tool for EFL?". Twelve trainees participated in the first group of students (one focus group), 20 in the second one (two focus groups) and 9 in the third one (one focus group).

The lesson plan these teacher candidates assessed was designed around the film *Bridge to Terabithia* (Csupó, 2007), which addresses school bullying (Appendix 1). The advantages present in the questionnaire were embedded in the lesson plan. The film, which was pedagogically exploited for a B1 level (see Appendix 1 for contents), was chosen following considerations of scholars' proposed criteria for film selection — the learners' needs and personal characteristics, their English language level, the instructor's pedagogical aims, and the film's ideological content (Asyidiq & Akmal, 2020; Tuncay, 2014; Wang, 2009). To be more specific, *Bridge to Terabithia* was selected, on the one hand, because, based on its unambiguous and straightforward plot, the film is an adequate platform from which to design activities targeting lexical and grammatical contents in line with those established in the Spanish Official Curriculum for Secondary Education EFL. On the other hand, the film deals with a topic of social justice that is of utmost importance in secondary education as bullying

is a worldwide scourge. Therefore, prompting trainees to reflect on a lesson plan that focuses on bullying can provide them with ideas and resources that might be useful in their future careers, and also with an example of socio-critical EFL (EFL that targets students' motivation to learn the FL by eliciting empathy, higher thinking evaluative skills, and personal growth) (Crookes, 2021; Fairclough, 2015; Pennycook, 2014).

To validate the research instruments, we sent the questionnaire, the two open questions, and the prompt for the focus group to three specialist experienced EFL professors; they gave their approval to these research instruments without requiring any modifications. Furthermore, the lesson plan was also sent to these professors to check its suitability.

After obtaining trainees' consent to participate in the study, data were gathered in three 90-minute sessions. Trainees devoted the first half of the first session to completing Open Question 1 to avoid contamination of participants' raw perceptions by the items in the questionnaire. Then, they completed the pre-test questionnaire. In the second session, trainees analysed the lesson plan. They were asked to find examples of activities which reflected each of the 17 advantages included in the questionnaire and to assess their pedagogical relevance (or lack thereof). No explanation of the advantages of cinema for EFL was provided. Therefore, any changes reported in Open Question 2, the focus groups, or the post-test questionnaire were due to the trainees' own perception following the analysis of the lesson plan. Finally, 60 minutes of the third session were allocated for trainees to complete the post-test questionnaire and Open Question 2. Focus groups were conducted that same day with volunteers (9 students participated in the first year, 8 in the second, and 11 in the third).

3.3. Data analyses

The quantitative sample was studied by means of parametric statistics. The mean frequencies and the standard deviations of the participants' responses to the questionnaires were calculated. Paired sample *t*-tests (with a significance level of 0.05) were run for pre- and post-test scores for the three components to indicate any statistically significant change in the trainees' beliefs. Size effects were also calculated. Before conducting the *t*-tests, the normality of the sample was confirmed through a Kolmogorov-Smirnov test (p>.005).

Responses to the open questions and the focus groups were transcribed and then analysed qualitatively by all three researchers separately. These responses were open-coded line by line (Corbin & Strauss, 2008) to identify the main actions in the sentences, which were coded with simple words reflecting these actions. Then, codes which referred to similar actions were grouped under concepts. Finally, related

concepts were collected within categories. Interrater agreement in this respect was found to be 85% and any disagreements were resolved by discussion between the three raters. Table 1 below presents the thematic categories which were identified in relation to the two Open Questions and the codes employed, while Table 2 provides an example of how coding was undertaken for both Open Questions:

Table 1: Examples of categories and codes for both open questions

Open Question	Category and Code
Trainees' views on the potential of cinema for EFL	(i) perceived usefulness of cinema-based EFL, especially for improving linguistic competence (USEF) (ii) diffuse awareness of the potential of film for EFL (AW) (iii) perceived problems of using film for EFL (PROB)
Change in trainees' perceptions after analysing the lesson plan	(i) heightened understanding of the advantages of cinema-based EFL (ADV) (ii) marked emergence of methodological awareness (METHOD) (iii) importance of seeing in the practice (PRACT)

Table 2: Examples of trainees' comments and coding for both open questions

Example
Open Question 1 (perceived potential of cinema): "It is just obvious that watching films helps enhance **students' listening skills". (USEF)**
Open Question 2 (change in perceptions): "Now I feel capable of using a film for the English class but only because **we saw how to do so** in the lesson plan we had to analyse"**. (PRACT)**

4. Results

4.1. Quantitative results

As shown in Table 3 and Figure 1, the participants' views about the usefulness of cinema for EFL were favourable in the pre-test and the post-test stages both as a whole (M = 3.11, SD = 0.28; M = 3.69, SD = 0.23) and for the three specific dimensions analysed: linguistic (M = 3.10, SD = 0.32; M = 3.67, SD = 0.24), intercultural (M = 03.23, SD = 0.46; M = 3.72, SD = 0.42), and motivational (M = 2.99, SD = 0.39; M = 3.68, SD = 0.69). Statistically significant increases in mean scores were reported in the post-test for all dimensions (p<0.001) and in general as well. Effect sizes were large

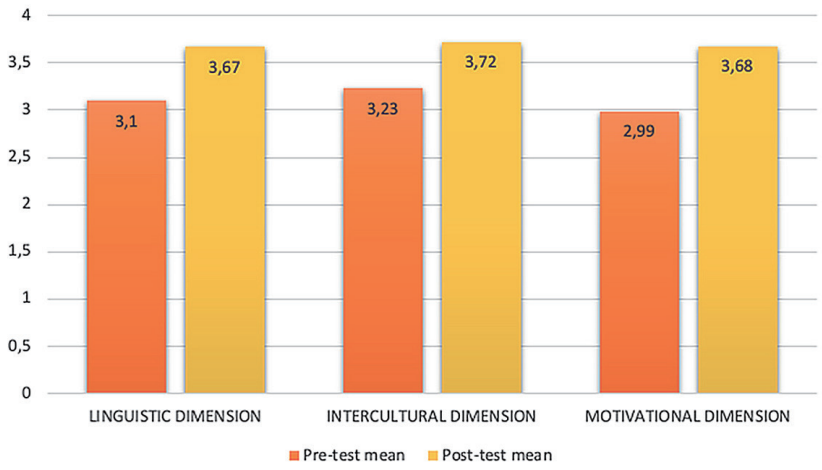
in all four cases, with values over 0.8. At both test times, the most valued was the intercultural dimension, followed by the linguistic and the motivational dimensions in the pre-test, and by the motivational and linguistic ones in the post-test. Mean scores for the three dimensions were less homogeneous in the pre-test than in the post-test since, in this latter stage, agreement rates were almost coincidental, ranging from 3.67 (linguistic dimension) to 3.72 (intercultural dimension). This means that, in the post-test, cinema-based EFL was perceived as almost equally useful for the development of linguistic and intercultural contents and for boosting students' motivation.

Table 3: Pre- and post-test results regarding trainees' views towards cinema-based EFL

DIMENSIONS AND ITEMS	Pre-test Mean	Post-test Mean	Mean change	t	*p* -value	d
Films can help improve students' linguistic competence:	**3.10 (0.32)**	**3.67 (0.24)**	**0.57**	-18.52	**< 0.001**	**-1.69**
Their writing (item 1)	2.23 (0.71)	3.30 (0.74)	1.07	-12.32	< 0.001	-1.12
Their speaking (item 2)	3.31 (0.71)	3.82 (0.41)	0.50	-7.37	< 0.001	-0.67
Their reading (item 3)	2.67 (0.81)	3.36 (0.71)	0.69	-8.47	< 0.001	-0.77
Their listening (item 4)	3.63 (0.52)	3.92 (0.31)	0.29	-5.33	< 0.001	-0.49
Their lexis (item 5)	3.64 (0.48)	3.92 (0.26)	0.28	-6.36	< 0.001	-0.58
Their grammar (item 6)	2.58 (0.73)	3.53 (0.65)	0.95	-12.72	< 0.001	-1.16
Their pronunciation (item 7)	3.63 (0.59)	3.84 (0.41)	0.21	-3.68	< 0.001	-0.34
Films can help improve students' intercultural competence:	**3.23 (0.46)**	**3.72 (0.42)**	**0.49**	-9.68	**< 0.001**	**-0.88**
Their knowledge of the L2 culture (item 8)	3.58 (0.56)	3.81 (0.40)	0.23	-4.14	< 0.001	-0.38
Their knowledge of the L1 culture (item 9)	2.88 (0.65)	3.62 (0.61)	0.74	-10.37	< 0.001	-0.95
Films can foster students' motivation to learn EFL:	**2.99 (0.39)**	**3.68 (0.28)**	**0.69**	-15.81	**< 0.001**	**-1.44**
Because they can express their own opinions and feelings (item 10)	2.87 (0.76)	3.68 (0.60)	0.81	-9.54	< 0.001	-0.87

DIMENSIONS AND ITEMS	Pre-test Mean	Post-test Mean	Mean change	t	*p* -value	d
Because they can understand others' opinions and feelings (item 11)	3.13 (0.60)	3.74 (0.48)	0.62	-9.19	< 0.001	-0.84
Because social justice can be promoted (item12)	2.84 (0.73)	3.59 (0.53)	0.75	-10.12	< 0.001	-0.92
Because high order thinking skills can be promoted (item 13)	3.18 (0.66)	3.72 (0.47)	0.54	-7.64	< 0.001	-0.70
Because it elicits aesthetic pleasure (item 14)	2.97 (0.87)	3.64 (0.56)	0.68	-7.82	< 0.001	-0.71
Because it is authentic material (item 15)	3.23 (0.71)	3.75 (0.45)	0.52	-7.18	< 0.001	-0.66
Because it deals with nontrivial topics (item16)	2.21 (0.88)	3.48 (0.61)	1.27	-13.73	< 0.001	-1.25
Because of its audio-visual condition (item17)	3.50 (0.61)	3.81 (0.42)	0.31	-4.86	< 0.001	-0.44
Total score	**3.11 (0.28)**	**3.69 (0.23)**	**0.58**	-18.99	< 0.001	**-1.7**

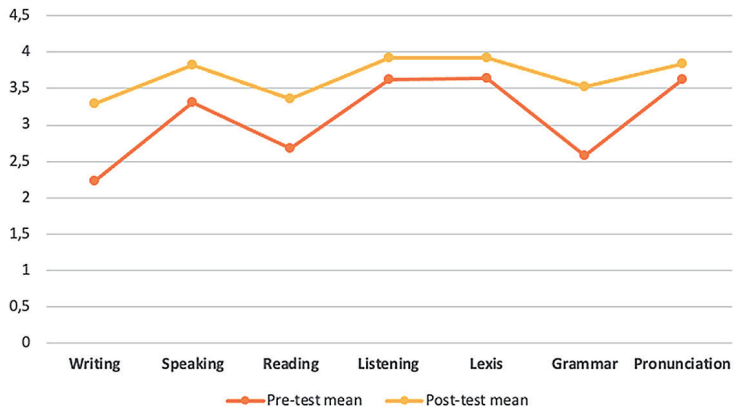
Figure 1: Pre- and post-test total mean values for the linguistic, intercultural and motivational dimensions



4.1.1. Linguistic dimension

Focusing on the linguistic dimension (See items 1-7 in Table 3 and Figure 2 below), the highest means in the pre-test were reported for vocabulary (item 5: M = 3.64, SD = 0.48), listening (item 4: M = 3.63, SD = 0.52), and pronunciation (item 7: M = 3.63, SD = 0.59). The lowest values were for writing (item 1: M = 2.23, SD = 0.71), grammar acquisition (item 6: M = 2.58, SD = 0.73), and reading (item 3: M = 2.67, SD = 0.81). This same trend was reported in the post-test: vocabulary (item 5: M = 3.92, SD = 0.26), listening (item 4: M = 3.92, SD = 0.31), and pronunciation (item 7: M = 3.84, SD = 0.41) received the highest values, while, once again, writing (item 1: M = 3.30, SD = 0.74), grammar acquisition (item 6: M = 3.53, SD = 0.65), and reading (item 3: M = 3.36, SD = 0.71) received the lowest. However, the gaps between the highest and the lowest rates were bridged in the post-test since all items were assigned values within the 3.30–3.92 spectrum, while pre-test means ranged from 2.23 to 3.64. The most accented mean changes between the pre-test and post-test were for writing (1.07) and grammar (0.95), and the least accented for pronunciation (0.21), vocabulary (0.28), and listening (0.29). All effect sizes were large, with *d* values between -1.16 and -0.49, with the exception of pronunciation, whose effect size was medium (*d* = -0.34).

Figure 2: Pre- and post-test mean values for each of the items within the linguistic dimension

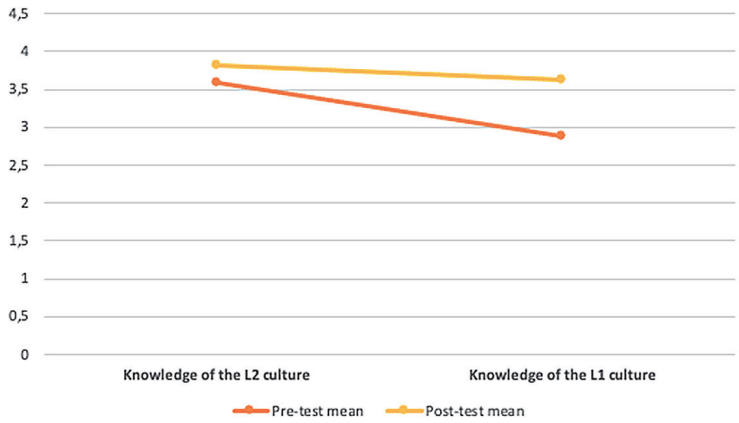


4.1.2. Intercultural dimension

As presented in Table 3 and Figure 3, the trainees' views were positive for the two items within the intercultural dimension at both test times. Change trends were statistically significant, with *p* values <0.001 for the two items (acquiring L2 cultural

knowledge [item 8: M = 3.58, SD = 0.56; M = 3.81, SD = 0.40]; and heightened awareness of one's own culture [item 9: M = 2.88, SD = 0.65; M = 3.62, SD = 0.61]). Effect sizes were large for item 9 (*d* = -0.95) and medium for item 8 (*d* = -0.38). Although item 8 obtained higher agreement rates than item 9 both in the pre-test and the post-test, the difference between these two items in the post-test was greatly reduced because item 9 experienced one of the most remarkable mean increases of all surveyed advantages (0.74).

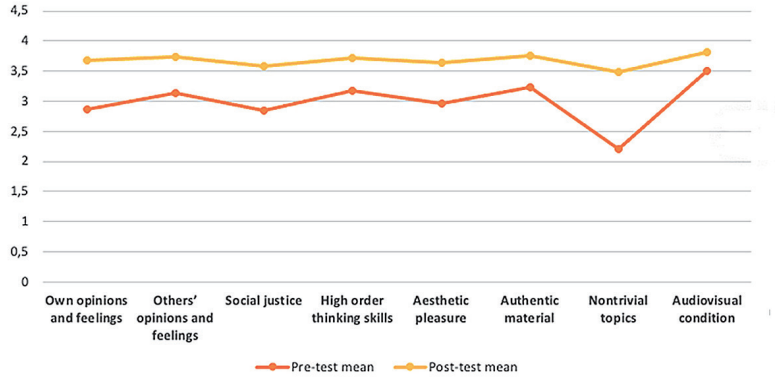
Figure 3: Pre- and post-test mean values for each of the items within the intercultural dimension



4.1.3. Motivational dimension

With regard to the motivational dimension, all the changes between the pre-test and the post-test were positive and statistically significant (p<0.001 for all items; see items 10-17 in Table 3 and Figure 4). All size effects were large, ranging from -0.44 to -1.25. The highest means in the pre-test and post-test stages were assigned to the audio-visual aspect of cinema (item 17: M = 3.50, SD = 0.61; M = 3.81, SD = 0.42) and the facts that films are authentic materials (item 15: M = 3.23, SD = 0.71; M = 3.75, SD = 0.45) and adequate tools for understanding alterity (item 11: M = 3.13, SD = 0.60; M = 3.72, SD = 0.48). The ability of films to deal with nontrivial topics (item 16: M = 2.21, SD = 0.88; M = 3.48, SD = 0.61) and introduce issues of social justice (item 12: M = 2.84, SD = 0.73; M = 3.59, SD = 0.53) were held as the least advantageous contributions of cinema-based EFL at the two test times. Despite obtaining the lowest scores, mean changes for these two items experienced the first and the third most important increases within the motivational dimension, 1.27 (item 16) and 0.75 (item 12).

Figure 4: Pre- and post-test mean values for each of the items within the motivational dimension



4.2. Qualitative results

4.2.1. Open Question 1

The trainees' responses to Open Question 1 — "What is your view about the potential of cinema for EFL?" — can be classified within three categories: perceived usefulness of cinema-based EFL, especially for improving linguistic competence; diffuse awareness of the potential of film for EFL; and perceived problems of using film for EFL.

4.2.1.1. Perceived usefulness of cinema-based EFL

Most comments in Open Question 1 referred to the perceived usefulness of film-based EFL for enhancing the linguistic competence. All 121 participants highlighted this usefulness, especially for listening, pronunciation, and vocabulary acquisition — mentioned by 121, 104, and 89 trainees, respectively. Eighteen trainees asserted that, with English subtitles, students could also improve their reading skills. However, output-based linguistic gains did not have such a good reception, with remarkably fewer trainees highlighting the potential of cinema to improve speaking (10 trainees), writing (8 trainees), or grammar acquisition (6 trainees). Example 1 illustrates the most common attitude towards the advantages of cinema-based EFL for improving the linguistic competence.

Example 1

I would say [cinema] is a very helpful tool as it can be used, especially, to develop listening skills and pronunciation, but also to learn vocabulary and practice reading if subtitles are shown.

Films' ability to illustrate cultural aspects of the L2 was mentioned by 63 trainees. In one participant's words, "films constitute a privileged window into the way of living of English-speaking people". Motivational factors were mentioned less often than linguistic and intercultural advantages. The participants considered film-based EFL might encourage students to learn for three reasons. First, because it is a novel approach that does not rely on textbooks (7 participants); second, because, as films are an authentic material, learners feel they are experiencing "the real thing" instead of "artificial" dialogues in textbooks (18 participants); and, third, due to cinema's audio-visual component, which learners enjoy because "video is simply their natural medium" (10 participants). No comments were reported for the other aspects included within the motivational dimension of cinema. Similar to the tendency reported for quantitative data, most of the participants' reflections were related to the intercultural and linguistic benefits.

4.2.1.2. Diffuse awareness of the potential of film for EFL

In their responses, 15 trainees evidenced an intuitive awareness that cinema could be useful for EFL in ways other than the enhancement of input-related aspects and the portrayal of the culture of English-speaking countries — the approach to language instruction they were most familiar with. However, they did not know what these other ways might be (Example 2). Twelve of these students also related their diffuse perceptions on this matter to the absence of cinema in their previous EFL schooling (5 participants, Example 3) or the lack of pedagogical purpose of the films screened (7 participants, Example 4).

Example 2

I think that cinema is a very powerful resource for foreign language teaching (FLT), but I'm not quite sure how to exploit it. A student can watch a film and learn some vocabulary but I am not aware of the potential activities that could be done.

Example 3

I am not very informed about the potential that cinema may have for EFL as I cannot recall having been taught English by means of a film. Therefore, this leads me to state that it may not have a remarkable influence on FLT.

Example 4

Cinema was always used as an extra activity to mark the end of the semester or as a sort of reward when the students behaved well. Thus, teachers never went further than playing a movie with subtitles in Spanish. They never exploited the film pedagogically to start a debate or whatever. So, watching a film always ended up being an activity to which students did not even pay attention.

4.2.1.3. Perceived problems of using film for EFL

Trainees also found the use of films for the English class might involve problems. Five reasons were identified. First, 21 participants considered a film was too long. As one of them asserted, "only watching a film would take, at least, two sessions, let alone exploiting it pedagogically". Second, 13 participants regarded the authentic nature of films as a hindrance to the teaching-learning process since extensive work with language might be necessary before the text could be used in class (Example 5). Third, they stated that finding a film that deals exactly with the target contents is highly unlikely (8 trainees). This, in the words of a trainee, "is time-consuming because the academic programme would have to be adjusted". The fourth problem, expressed by 7 trainees, was that using films in class would involve an extra effort on the part of the teacher, because there might be a shortage of teaching material and teachers would have to design their own activities (Example 6). While these four problems were related to time and the added pressure for teachers, the fifth problem emphasized the fact that learners might prefer TV series to films (5 participants). According to them, this might be because learners "are used to watching" TV series or because "films are too long to maintain teenagers' span of attention, no matter how interesting the film is".

Example 5

I don't think students can understand [the language in the film] unless they have a high level of English or they do many vocabulary activities before watching it to facilitate comprehension. I see it problematic in both cases. If the film is screened directly, learners might get frustrated. However, if vocabulary is practised previously, much time will be devoted to controlled tasks, which may demotivate them too.

Example 6

I have never seen a textbook which includes films and, consequently, teachers would have to do everything from scratch, whether finding a film and designing the activities. And this is too time-consuming.

4.2.2. Open Question 2

The analysis of responses to Open Question 2 — "Has your perception of the potential of film for EFL changed after analysing the lesson plan? Why? Why not?"— allowed us to attest that trainees' views changed positively following the analysis of the film-based lesson plan. Three categories explained these positive changes: heightened understanding of the advantages of cinema-based EFL, marked emergence of methodological awareness, and the usefulness of analysing a practical example of a cinema-based EFL lesson plan.

4.2.2.1. Heightened understanding of the advantages of cinema-based EFL

While in Open Question 1, the majority of participants commented on the linguistic and intercultural benefits of films, responses to Open Question 2 evidenced a marked increase in the number of the motivational advantages brought up, some of which had not been mentioned at all in the pre-analysis stage. The latter was the case for the two most frequently reported motivational incorporations. Sixty participants (49.5% of the total sample) considered that focusing the lesson plan on an issue of social justice, bullying in this case, might be highly motivational for students because, in such a way, the EFL class "becomes useful as bullying is a very relevant topic", "helps improve students' lives", or "contributes to making students more critical". Besides, 13 of these 60 trainees expressed their surprise at social justice not being more frequently present in EFL lessons:

Example 7

I don't understand why this approach is hardly implemented. It is a "win-win" because students learn English at the same time as they discover key concepts about bullying which may be highly useful.

Cinema's ability to activate high-order thinking skills, an aspect which emerged in trainees' reaction to the social justice component of the lesson plan on *Bridge to Terabithia*, was the second most common type of comment in Open Question 2 (49 trainees):

Example 8

I reckon that the activities in which students have to make hypotheses and create questionnaires are quite relevant because, apart from learning grammar, they are encouraged to analyse their own situation.

In contrast, the frequency of linguistic-related opinions, computed as a whole, experienced a marked decrease in the post-analysis stage (113 comments in the post-test versus 235 comments in the pre-test). The increase in qualitative motivational comments seems to be explained by a trade-off towards the items within this dimension which trainees might have perceived as more novel. Still, the heightened perception of motivational advantages did not imply trainees stopped noticing linguistic benefits; as reported before, in quantitative terms, these were still assigned the highest values at the post-test stage. What transpired was a levelling out of the linguistic, intercultural and motivational dimensions in the post-analysis stage, which, therefore, were perceived as almost equally important.

4.2.2.2. Marked emergence of methodological awareness

In Open Question 2, 81 trainees (66.9% of the total sample) referred to methodological aspects regarding the use of cinema for EFL. This implied a remarkable difference from Open Question 1, in which no such comments were recorded at all beyond the previously mentioned intuition that cinema could be used in other ways. In fact, six trainees explicitly expressed that, after the analysis of the lesson plan based on *Bridge to Terabithia*, they could finally understand and verbalise that marginal intuition they had in the pre-analysis stage that cinema could be used more profitably in the English class. This can be seen by comparing Example 2 with Example 9:

Example 9

The analysis of the lesson plan has widely expanded my views on using cinema for English teaching and learning. Thanks to this, I have realised that my views were too traditional since there are several ways to exploit a film of which I was totally unaware. I knew that cinema was a potentially powerful resource but I did not know how to exploit it.

This methodological awareness defused some of the problems of the film-based approach mentioned in the pre-analysis stage, to the extent that no problems were mentioned anymore in the post-test. A remarkable number of methodological comments (21 trainees, 17.3% of the total sample) highlighted the fact that the lesson plan had included short excerpts instead of the whole film. This was perceived by 14 of these 21 trainees to address the most common problem mentioned in responses to Open Question 1, namely, that films were unsuitably long for one-hour English classes. "My opinion has changed because using short fragments does really allow you to use films in class", one trainee said. Another methodological aspect which 19 trainees mentioned focused on the fact that, in the sample lesson plan, the fragments used had been pasted together by means of a narrative told orally by the teacher during the

sessions that provided the film with cohesion. In the trainees' own words, proceeding this way resulted in students not "getting lost in the plot" (5 participants) and feeling that "there was a beginning and an end to the plot instead of unconnected fragments, which would surely demotivate them to learn" (7 participants). This strategy was also observed by seven trainees as the solution to another problem stated in the pre-analysis comments: the concern that the authentic and, therefore, demanding nature of films could put off students. Through this narrative technique, "the teacher could adapt the film to whichever target level and direct the learners' attention to those specific excerpts of authentic language in the film from which [they want] to work".

Eighteen trainees (14.8% of the total sample) also mentioned that examining the lesson plan had made them aware of the fact that the exploitation of cinema did not necessarily rely on contents present "in" the film (e.g., vocabulary or grammatical patterns included in a specific scene). Instead, activities could be designed by establishing lines of flight "from" the film. Eight trainees explicitly stated that this "from" strategy made them see how films could be used for teaching writing or speaking skills, activating critical thinking, or exploring issues of social justice. As one participant said, "the film works as an excuse to deal with any content from it. Now, it is clear to me that a film can help students to practice their writing. This is as easy as asking students to react to any aspect in the film".

4.2.2.3. Usefulness of analysing a practical example of a cinema-based EFL lesson plan

Lastly, trainees asserted that the methodological gains detailed had been elicited from examining a practical application of the advantages of cinema-based EFL as developed in the sample lesson plan (29 participants, 23.9% of the total sample). One participant referred to this as follows: "if I hadn't seen a concrete and comprehensive example of how to use cinema for EFL, I would still be unaware of most of its advantages. While analysing the lesson plan, I have realised that films may be exploited beyond the typical listening or vocabulary tasks. I have missed the practical side in this master's degree. I can't understand how most of the contents are theoretical in a teacher training programme".

4.2.3. Qualitative results for focus groups

To respond to the question that guided the focus groups — "Once you have analysed the lesson plan provided, can you assess the usefulness of film as a tool for EFL?" — all four groups referred to arguments they had already expressed in Open Question 2. In fact, the contributions of Groups 1 and 3 did not add any new data and, therefore, they are not reproduced again. Interactions in Groups 2 and 4 offer some

new categories worth exploring: the importance of topics and the conflict between dealing with issues of social justice and parental authorisation.

4.2.3.1. The importance of topics

Groups 2 and 4 brought up the importance of topic selection for cinema to be successfully used in the EFL class on two different grounds: attractiveness and emotional resonances. As shown in Interaction 1, Group 2 related the well-functioning of cinema in EFL to films dealing with topics learners found interesting.

Interaction 1

T1: I think students feel attracted to films naturally because of their audio-visual component and the fact that they are not normally used as pedagogical material. However, if the topic of the film does not engage them, the class is not going to work, no matter how many films are screened.

T2: I could not agree more with you. The topics in the English class are not attractive at all. At least, I never found them attractive.

T3: I really think using topics such as bullying might engage EFL learners to a greater extent.

Group 4 also engaged in a discussion on the importance of the topics portrayed in films. Interaction 2 shows that this group considered that, while teachers' knowledge of methodological guidelines is essential to motivate students, it is not enough. The topic of the film is equally important. These trainees referred specifically to topics which appeal to students' emotions. They considered this so relevant that they even proposed that teachers should use a questionnaire to learn about their students' needs in this respect.

Interaction 2

T1: Knowing how to implement cinema in the English class is useless if the film topic is not chosen properly. Teachers must select subjects which strike a chord with learners. In textbooks, topics are so ridiculous that it is impossible to get emotionally involved with them.

T2: That is true! Bullying is something that touches us all either because we have experienced it or because we have seen it. A textbook would never include such a theme.

T3: I think a solution to this might be administering a questionnaire to ask students about the topics they are concerned with. In such a way, they could continue learning English in class but through appealing themes.

T2: Yes, but this is not just about textbooks. Teachers are not willing to deal with these topics either. They explain grammar and then leave the classroom. Generally speaking, instructors do not care about learners' emotional state. I mean, they do not even know your name, and they do not deign to ask how you are doing.

T4: What you mention is typical of the English classroom because, in other subjects like history, I did have teachers that triggered the students' emotional reactions through the historical facts they explained. They also made us establish links between historical events and our own lives and context.

4.2.3.2. Social justice and parental authorisation

Parental authorisation was brought up by both groups when talking about the topic of social justice present in *Bridge to Terabithia* (bullying). While the participants agreed that including such topics in the EFL class might make the lessons more interesting and more meaningful to students, they also expressed their concerns that targeting such contents might be controversial, especially, if they related to gender or sexual orientation. This was so on the grounds that the Education Authority of the Region of Murcia might consider that these learners are being indoctrinated and, consequently, the teacher might be in legal trouble. This concern elicited various reactions, the most numerous in Group 3, as reported next.

Interaction 3

T1: Tackling social matters is very interesting, but watch out! The students' parents may denounce you on the grounds that you are brainwashing their children. Some of them ask teachers to stop beating around the bush and focus on the content instead of on social issues. If this happens… you may get into trouble with inspection.

T2: So what? My mother is a teacher and, if the educational centre has a progressist policy and supports you, it is okay. Instructors have academic freedom. I am not going to leave these topics aside for the fear of a parent's complaint, because they help us to grow up as citizens.

T1: Well, you do what you want to! But I am not going to be in trouble however relevant I find these topics, especially at the beginning of my teaching career when I have just started at a high school and I am an underdog.

Group 4 also considered parental authorisation as a problem. The mention of parental authorisation was followed by an uncomfortable pause, which the conductor of the focus group perceived as participants' reluctance to express their opinions on this issue. Instead, the trainees debated the safeness of the topics included in textbooks and even in certification exams and how discouraging this safeness is when it comes to expressing proper opinions.

Interaction 4

T1: Personally, dealing with values in the EFL class makes me rather nervous because of parental authorisation! I think we all understand what I am talking about, don't we?

[Uncomfortable pause]

T2: The topics included in textbooks and even in certification exams like those offered by Cambridge are so silly that it is difficult to feel motivated to talk about them, which should be the objective.

T3: You are right! If topics were more charged and you had to personally engage with them, it would be easier to have more substantial opinions!

5. Discussion

The research questions that guided the present study are: RQ1. What are trainees' views about the use of films in EFL, and RQ2. Do these views change as a result of trainees analysing a cinema-based lesson plan?

The results showed agreement rates were positive at both test times, although significantly higher in statistical terms and closer to full agreement after students had analysed the lesson plan. The academic production that thus far has examined the advantages of film-based EFL as perceived by students, teachers, or trainees has also revealed positive results. Specifically, favourable views were mostly reported for the enhancement of learners' linguistic competence (Aksu-Ataç & Günay-Köprülü, 2018; Albiladi et al, 2018; Anas & Zakaria, 2019; Chapple & Curtis, 2000; Chiu, 2012; Christopley, 2017; Ismaili, 2013; Kabooha, 2016; Mahmoodi-Shahrebabaki, 2015; Shahani et al., 2014; Tuncay, 2014); followed by the motivational aspects of cinema, particularly, the depiction of authentic language and relevant topics (Albiladi et al., 2018; Anas & Zakaria, 2019; Chapple & Curtis, 2000; Tuncay, 2014); and the development of intercultural awareness (Albiladi et al., 2018; Argynbayev et al., 2014; Chao, 2013; Tuncay, 2014).

The methodological awareness acquired through the analysis of the lesson plan (using short excerpts of the film, narrating the rest of the plot, and working "in" and "from" the film) seems to be an important factor. It explains the levelling out in the post-intervention stage regarding the perception of the three dimensions under study and the fact that the agreement rate for the three dimensions in the post-intervention stage was closer to full agreement. The trainees reported that this heightened methodological awareness broadened their perceptions on how cinema could be used in EFL: as a tool to improve input-based skills and knowledge of the L2 culture, but also encompassing output-based skills and a wide array of motivational advantages. On the other hand, methodological awareness was also considered instrumental in deactivating the problems they had previously identified (the length of a film, the complex authentic nature of its language, or the shortage of cinema-based materials). On the whole, it seems that any doubts about the usefulness of cinema for EFL reported in the pre-intervention stage were assuaged by the clear emergence of methodological awareness following the analysis of the cinema-based lesson plan.

To date, existing research in the field had never reported participants' methodological insights, not even in the empirical studies which tapped into teachers' perceptions (Aksu-Ataç & Günay-Köprülü, 2018; Ismaili, 2013; Kabooha, 2016; Mahmoodi-Shahrebabaki, 2015; Sánchez-Auñón & Férez-Mora, 2021). This seems to be because most of this research did not explore change in participants' perceptions as it focused on the epistemological exploration of film-based EFL through one-shot designs. Although the latter is obviously a fundamental endeavour, the clear practical methodological guidelines on this method are equally important since such cues might help EFL teachers alleviate the difficulties which lead them to implement cinema as a time filler instead of a resource of its own right (Albiladi et al., 2018; Hobbs, 2006). As argued by Martínez Agudo (2017) and Peacock (2009), EFL teacher training programmes should prioritise the generation of methodological awareness since the innovative development of teaching/learning processes depends upon it. Unfortunately, it seems this aspect does not receive enough attention in these programmes, and trainees highlighted it as an important deficiency (Martínez Agudo, 2017; Peacock, 2009). Teachers have also emphasized the absence of specific training on the exploitation of film for EFL (Anas & Zakaria, 2019; Chapple & Curtis, 2000), which students have noticed as well (Kabooha, 2016; Shahani et al., 2014; Tuncay, 2014; Yue, 2019).

It is important to note that these trainees associated their enhanced methodological awareness with seeing a practical application of the advantages of cinema-based EFL, as presented in the lesson plan they had to analyse. In doing so, they are acknowledging a well-known principle of teacher training: the analysis of practical models fosters

willingness to learn, which, in turn, results in more solid learning processes (Mosley Wetzel et al., 2017). This led us to revise the lesson plans included in previous studies. As emphasized earlier, it seems existing scholarship had put up a limited application of the "in practice" factor. Aksu-Ataç & Günay-Köprülü (2018), Albiladi et al. (2018), and Anas and Zakaria (2019) tapped into English instructors' and students' perceptions on the use of cinema only in the abstract since no pedagogical intervention was included. In contrast, Argynbayev et al. (2014), Chao (2013), Chiu (2012), Kabooha (2016), Mahmoodi-Shahrebabaki (2015), Shahani et al. (2014), and Tuncay (2014) explored the participants' views after implementing a cinema-based lesson. However, they looked into a limited range of advantages and failed to provide the lesson plan used. This research gap (and weakness) in the field is worth attending to in future studies since the absence or very limited presence of lesson plans undermines the solidity of reported data in that it is unclear what the pedagogical intervention consisted of.

The changes reported in the participants' perceptions as regards the use of films in EFL highlight other pertinent areas which to date have also remained uninformed or under-informed in the field: socio-critical approaches to EFL and trainees' professional voice.

5.1. Trainees' views of EFL evolve from a traditional language-based to a more socio-critical understanding

Following the analysis of the lesson plan, film-based EFL was perceived as almost equally useful for enhancing linguistic, intercultural, and motivational dimensions. This seems to indicate that trainees' perceptions evolved from considering film as a tool which mainly generated linguistic and intercultural competences to a platform for a much broader range of contents related to a socio-critical agenda, including the exploration of one's own emotions and those of others, social justice, or creativity and critical thinking. English teachers' and students' thoughts on the impact of films as regards these last four contents have been minimally explored to date, albeit with very positive results: Anas and Zakaria (2019) tapped into social justice; and Anas and Zakaria (2019), Chapple and Curtis (2000), and Tuncay (2014) delved into creative and critical thinking skills. This shows that surveyed trainees were willing to understand EFL beyond linguistic instrumentality. In turn, their interest in the socio-critical agenda situated them in line with principles of critical applied linguistics. This approach purports FLT should not only aim at addressing the correct use of an abstract linguistic system but, instead, should be understood as social practice and target personal growth, critical awareness, self-empowerment, and engagement with social justice, which communication should activate (Crookes, 2021; Fairclough, 2015; Pennycook, 2014).

Lately, this socio-critical approach to EFL has been gaining momentum on the grounds that FL classrooms should connect "world and word" indissolubly (Safari & Rashidi, 2015). One of the most common lines along which this socio-critical approach to FL has materialized, and also the one adopted by the present study, has involved the design of and the instruction with lesson plans addressing topics of social justice. These socio-critically oriented lesson plans, towards which learners showed manifestly positive views, have tended to explore forms of alterity, whether based on gender (Hayik, 2015), sexual orientation (Férez Mora et al., 2022), race (Curtis & Romney, 2006), or social class (Vandrick, 2010). Given its focus on bullying, a practice which stems from the basic failure to accept alterity (Rivers, 2011), the present study fully connects with this socio-critical orientation to EFL, thus testifying to an increasing research interest in moving FL beyond a merely linguistic understanding.

5.2. Emergence of trainees' professional voice

In the open questions, and especially in the focus groups, these trainees' own voices as future teachers started emerging. They began to reflect about what EFL education is and should be, what sort of educators they want to become, what problems they might face as teachers and how they would address such problems. Some of the trainees did not even hesitate to critically approach current educational practices. The clearest instance of this was some trainees' conviction that they would include issues of social justice into their lessons, no matter the concern about parental authorisation.

As with most of the findings thus far defined, the emergence of trainees' professional voices had not been reported before in the field of film-based EFL. However, this is one of the most important objectives of teacher training programmes because, as held by Furlong and Maynard (2012), the consolidation of a voice of one's own is likely to determine whether pre-service teachers are ready to start teaching.

6. Conclusions

This study revealed that, before the pedagogical intervention, trainees had a limited understanding of the potential of cinema for EFL since they tended to conceive this textual modality as useful mostly for eliciting the linguistic and intercultural competences. However, following the analysis of a film-based lesson plan, trainees broadened their understanding, considering film useful not only for students' linguistic and intercultural enhancement but also for fostering their autonomy as thinkers and their awareness of issues of social justice — among other motivational variables. This amplification of trainees' views seemed to stem from the deactivation of perceived problems of the cinema-based approach and the rise of methodological awareness. As confirmed by the participants, both operations emerged from their first-

hand experience of the cinema-based lesson plan as it allowed them to develop a more holistic, critical, and personalised conception of EFL. Considering Lamie's assertion (2004) that instructors tend to apply new teaching strategies once they notice their positive impact, these future teachers are likely to introduce into their own practice everything they have learnt during this research on the use of film. Indeed, the critical understanding of EFL they have developed runs counter to traditional approaches which look at EFL as a mere instrumental endeavour.

The pedagogical implications of the present study seem to unfold from a specific insight: if the surveyed trainees' regard for cinema-based EFL was high, then why are films not a popular resource in the EFL class? Textbook-centred instructional practice might be one of the reasons as textbooks commonly fail to include cinema-based materials (Boufahja, 2019; Tajeddin & Teimournezhad, 2015). Another relevant reason seems to be the paucity of empirical research as EFL practitioners might feel the inclusion of cinema-based EFL does not sit at sound foundations. Likewise, the dearth of practice-oriented examples and guidelines, especially when it comes to designing lesson plans and not merely one-off activities, might hinder the incorporation of films into instructional practice. The lack of teacher training specifically aimed at designing cinema-based lesson plans also emerges as a significant factor because, without such training, teachers are likely to exclude film from their English lessons (or to use it unsuitably, as reported before).

The present study is not without its limitations. A larger sample of trainees including other Spanish and international universities would have contributed to more generalizable findings. Although the experimental design employed rendered positive results, implementing a longer lesson plan might have led to more fine-grained results as we could have undertaken a more nuanced exploitation of the advantages of film-based EFL. Also, conducting a delayed post-test questionnaire might have revealed the extent to which trainees' favourable opinions lasted over time, thus indicating the internalization of positive change. In addition to the research lines which unfold from the previously-mentioned limitations, other lines of enquiry stemming from the present study include: undertaking further studies with a training focus which incorporates comprehensive practical applications of the film-based approach; exploring the opinions of both experienced and novel EFL instructors on this textual modality and comparing them with that of English learners; analysing students' attitudes towards the motivational aspects of cinema, a dimension commonly left aside, as revealed by this study; and carrying out contrastive research on EFL learners' linguistic, inter/cultural, and motivational gains as a result of film-based instruction.

7 References

Aksu-Ataç, A. & Günay-Köprülü, S. (2018). The role of subtitle in foreign language teaching. *International Online Journal of Education and Teaching, 5*(3), 525-533.

Albiladi, W. S., Abdeen, F. H., & Lincoln, F. (2018). Learning English through movies: Adult English language learners' perceptions. *Theory and Practice in Language Studies, 8*(12), 1567-1574. http://dx.doi.org/10.17507/tpls.0812.01

Allan, M. (1985). *Teaching English with Video.* Harlow: Longman.

Alwehaibi, H. O. (2015). The impact of using YouTube in EFL classroom on enhancing EFL students' content learning. *Journal of College Teaching & Learning, 12*(2), 121-126. https://doi.org/10.19030/tlc.v12i2.9182

Anas, F. A., & Zakaria, N. Y. K. (2019) ESL learners' perceptions on English subtitled audio-visual materials. *C.reative Education, 10*, 2796-2803. https://doi.org/10.4236/ce.2019.1012206

Argynbayev, A., Kabykbekoba, D., & Yaylaci, Y. (2014). Teaching culture and identifying language interference errors through films. *English Language Teaching, 7*(9), 49-56. http://dx.doi.org/10.5539/elt.v7n9p49

Asyidiq, M. N., & Akmal, R. O. (2020). Animation short film on students' narrative text writing ability (a mixed method). *Universal Journal of Educational Research, 8*(5), 2022-2027. http://dx.doi.org/10.13189/ujer.2020.080540

Berk, R. A. (2009). Multimedia teaching with video clips: TV, movies, YouTube, and mtvU in the college classroom. *International Journal of Technology in Teaching and Learning, 5*(1), 1-21.

Boufahja, I. (2019). Incorporating audio-visual aids in the English language classroom in Tunisia: A case study. In S. Hidri (Ed.), *English Language Teaching Research in the Middle East and North Africa* (pp. 687-701). London: Palgrave Macmillan.

Canning-Wilson, C. (2000). Practical aspects of using video in the foreign language classroom. *Internet TESL Journal, 6*(11).

Carney, N., & Foss, P. (2008). Student-produced video: Two approaches. *English Teaching Forum, 46*(2), 14-19.

Cetinavci, U. R. (2019). The effects of explicit film-based instruction on EFL teacher trainees' interpretation of implied meanings. *European Journal of Educational Research, 8*(2), 581-605. https://doi.org/10.12973/eu-jer.8.2.581

Chao, T. (2013). A diary study of university EFL learners' intercultural learning through foreign films. *Language, Culture and Curriculum, 26*(3), 247-265. http://dx.doi.org/10.1080/07908318.2013.833936

Chapple, L., & Curtis, A. (2000). Content-based instruction in Hong Kong: Student responses to film. *System, 28*(3), 419-433. https://doi.org/10.1016/S0346-251X(00)00021-X

Chiu, Y. (2012). Can film dubbing projects facilitate EFL learners' acquisition of English pronunciation? *British Journal of Educational Technology, 43*(1), 24-27. http://dx.doi.org/10.1111/j.1467-8535.2011.01252.x

Christopley, C. (2017). Developing students' vocabulary through watching English cartoon films at SMP BHAKTI. *Journal of English Teaching, 3*(1), 11-26. https://doi.org/10.33541/jet.v3i1.696

Corbin, J. M., & Strauss, A. L. (2015). *Basics of Qualitative Research*. Thousand Oaks, Ca.: Sage.

Crookes, G. V. (2021). Critical language pedagogy: An introduction to principles and values. *ELT Journal, 75*(3), 247-255. https://doi.org/10.1093/elt/ccab020

Csupó, G. (Director). (2007). *Bridge to Terabithia* [Film]. Walt Disney Pictures and Walden Media.

Curtis, A., & Romney, M. (2006). *Color, Race, and English Language Teaching: Shades of Meaning*. Mahwah, NJ: Lawrence Erlbaum Associates.

Dörnyei, Z. (2007). *Research Methods in Applied Linguistics*. Oxford: Oxford University Press.

Fairclough, N. (2015). *Language and Power*. London: Routledge.

Férez Mora, P. A., Coyle, Y., & Dorado Otero, A. (2022). Foreign language students' views on FL and critical literacies. *Language* Awareness, *31*. https://doi.org/10.1080/09658416.2022.2042008

Férez Mora, P. A., Coyle, Y., & Maurandi López, A. (2020). The use of poetry for EFL: Exploring the beliefs of primary school trainees. *Revista Signos, 53*, 56-79. http://dx.doi.org/10.4067/S0718-09342020000100056

Fernández-Guerra, A. B. (2013). Using TV series as input source of refusals in the classroom. In P. Salazar-Campillo, & O. Martí Arnàndiz (Eds.), *Refusals in Instructional Contexts and Beyond* (pp. 5-23). Leiden: Brill Academic Publishers.

Fluitt-Dupuy, J. (2001). Teaching argumentative writing through film. *TESOL Journal, 10*(4), 10–15.

Furlong, J., & Maynard, T. (2012). *Mentoring Student Teachers: The Growth of Professional Knowledge*. London: Routledge.

Gallagher, B. (1988). Film study and the teaching of English: Technology and the future of pedagogy. *English Journal, 77*(7), 58-61. https://doi.org/10.2307/818940

Goodnough, K., Osmond, P., Dibbon, D., Glassman, M., & Stevens, K. (2009). Exploring a triad model of student teaching: Pre-service teacher and cooperating teacher perceptions. *Teaching and Teacher Education, 25*, 285-296.

Hayik, R. (2015). Addressing religious diversity through children's literature: An "English as a foreign language" classroom in Israel. *International Journal of Multicultural Education, 17*(2), 92-106.

Hobbs, R. (2006). Non-optimal uses of video in the classroom. *Learning, Media and Technology, 31*(1), 35-50. https://doi.org/10.1080/17439880500515457

Ismaili, M. (2013). The effectiveness of using movies in the EFL classroom—A study conducted at South East European University. *Academic Journal of Interdisciplinary Studies, 2*(4), 121-132. http://dx.doi.org/10.5901/ajis.2012.v2n4p121

Jurkovič, V., & Mertelj, D. (2015). Pedagogical uses of authentic video in ESP classrooms for developing language skills and enriching vocabulary. *Scripta Manent, 9*(2), 15-33.

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

Kaiser, M. (2011). New approaches to exploiting film in the foreign language classroom. *L2 Journal, 3*(2), 232-249. https://doi.org/10.5070/L23210005

Lamie, J. M. (2004). Presenting a model of change. *Language Teaching Research, 8*(2), 15-142. https://doi.org/10.1191/1362168804lr137oa

Lonergan, J. (1984). *Video in Language Teaching*. Cambridge: Cambridge University Press.

Mahmoodi-Shahrebabaki, M. (2015). Attitude of Iranian EFL learners and teachers towards using movie clips in teaching idioms. *Modern Journal of Language Teaching Methods, 5*(1), 202-210.

Martínez Agudo, J. D. (2017). What EFL student teachers think about their professional preparation: Evaluation of an English language teacher education programme in Spain. *Australian Journal of Teacher Education, 42*(8), 62-76. http://dx.doi.org/10.14221/ajte.2017v42n8.5

Mosley Wetzel, M., Hoffman, J. V., & Maloch, B. (2017). *Mentoring Preservice Teachers through Practice: A Framework for Coaching with CARE*. London: Routledge.

Peacock, M. (2009). The evaluation of foreign-language-teacher education programmes. *Language Teaching Research, 13*(3), 259-278.

Pennycook, A. (2014). *The Cultural Politics of English as an International Language*. London: Routledge.

Plano Clark, V. L., & Creswell, J. W. (2014). *Understanding Research: A Consumer's Guide*. London: Pearson.

Rivers, I. (2011). *Homophobic Bullying: Research and Theoretical Perspectives*. Oxford: Oxford University Press.

Safari, P., & Rashidi, N. (2015). Iranian EFL teachers' voices on the pedagogy of word and world. *Journal of Social Science Education, 15*(4), 39-52.

Sánchez-Auñón, E., & Férez-Mora, P. A. (2021). Film for EFL: Exploring the perceptions of a high school teacher. *Indonesian Journal of Applied Linguistics*, *11*(1), 49-59. https://doi.org/10.17509/ijal.v11i1.34634

Scacco, J. (2007). Beyond film: Exploring the content of movies. *English Teaching Forum, 45*(1), 10-15.

Shahani, S., Tahriri, A., & Divsar, H. (2014). EFL learners' views towards video materials and viewing techniques. *International SAMANM Journal of Business and Social Sciences, 2*(1), 42-60.

Stewart, D. M. (2006). Film English: Using films to teach English. *Electronic Journal of English Education, 24*, 1-17.

Sturm, J. L. (2012). Using film in the L2 classroom: A graduate course in film pedagogy. *Foreign Language Annals, 45*(2), 246-259. https://doi.org/10.1111/j.1944-9720.2012.01187.x

Tajedding, Z., & Teimournezhad, S. (2015). Exploring the hidden agenda in the representation of culture in international and localised ELT textbooks. *The Language Learning Journal*, *43*(2), 180-193. https://doi.org/10.1080/09571736.2013.869942

Tuncay, H. (2014). An integrated skills approach using feature movies in EFL at tertiary level. *Turkish Online Journal of Educational Technology, 13*(1), 56-63.

Vandrick, S. (2010). Social class privilege among ESOL writing students. In M. Cox, J. Jordan, C. Ortmeier-Hooper, & G. G. Schwartz (Eds.), *Reinventing Identities in Second Language Writing* (pp. 257-272). Urbana: National Council of Teachers of English.

Wang, Y. (2009). Using films in the multimedia English class. *English Language Teaching, 2*(1), 179-184.

Yue, J. (2019). Use of foreign films in cultivating intercultural communicative competence in ELT—A case study. *Theory and Practice in Language Studies*, *9*(2), 198-203. http://dx.doi.org/10.17507/tpls.0902.09

Appendix 1. Lesson plan

Objectives: Students will be able to describe actions; students will be able to participate in a debate on bullying; students will be able to reflect on bullying; students will be able to establish and play roles; students will be able to create a questionnaire; students will be able to describe places; students will be able to do a group presentation; students will be able to promote anti-bullying actions.

Linguistic contents: When clauses, modal verbs, question formation, adjectives of emotions, vocabulary to make descriptions, vocabulary to express and justify opinions, linkers, report-writing conventions.

Intercultural and socio-critical contents: Bullying, social isolation, critical thinking, North American education system, appreciation of beauty, fantasy, others' and one's own emotions.

Activities for "Bridge to Terabithia"

1. Students are prompted to make hypotheses about the plot of the film from its cover.

2. To illustrate the topic of the film, students are shown short clips from the first 20 minutes. They have to identify the disagreeable actions portrayed, why they are disagreeable and what is common to all the depicted situations. It is expected that students identify the clips that show bullying scenes occurring at a school in the United States. Then, the class has to engage in a debate to decide why bullying happens in general.

3. The teacher provides a brief summary of the main characters and main events for the first half an hour of the film.

4. Drawing on the bullying scenes presented before, students have to establish the roles of the different people involved in the scene (bully, victim or bystander) and define their emotional states through the technique "flow of feelings."

5. The teacher informs the class about a key to prevent bullying, the three Rs: refuse to be a bully bystander, recognise bullying, report bullying.

6. The teacher asks students in groups to come up with a list of typical actions that can be related to bullying and they have gone through or have seen in class and/or school. In a whole-class activity, they all decide which items must be included in a questionnaire (one enquiring about students as victims and the other enquiring about students as bullies). The items of the questionnaires have to be formulated as questions.

7. The teacher gathers all the answers anonymously and computes data. They provide the data to the students, who, in groups, have to write a report about the bullying climate in class.

8. Excerpts from the fantastic events in the forest are projected. In pairs, students are asked to answer the following questions: "What is Terabithia?" "Why do Lesley and Jess go there?" "Do Lesley and Jess have a safe relationship with fantasy?" "Can you connect Terabithia with yourself?" "Have you ever created a fantasy world where to take refuge when things are not going well in life?"

9. As a final task, students are asked to prepare group presentations to inform the head of the school in *Bridge to Terabithia* of a plan to stop bullying in the school, so that Lesley and Jess can put their own private bridge to Terabithia to an end.

Appendix 2. Questionnaire

QUESTIONNAIRE ON THE PERCEPTIONS OF STUDENTS IN THE MDTT (ENGLISH) CONCERNING THE USE OF CINEMA AS A TOOL FOR THE TEACHING OF ENGLISH AS A FOREIGN LANGUAGE*

This questionnaire is aimed at exploring your views on the use of cinema as a tool for the teaching of English as a foreign language. There is not any right or wrong answer. Remember to answer **all the questions** and to select the corresponding options quicky, based on your immediate reactions to the statements. Indicate your level of agreement with each statement according to the following scale:

Stronly disagree	Disagree	Agree	Strongly agree
1	2	3	4

This is an **anonymous** questionnaire, and all the information provided will be handled **confidentially**. If you are interested in knowing the results obtained, please, contact Professor… from the Department…, University of…

I IDENTIFICATION DATA

IDENTIFIER: _____ (six numbers from your birth date, e.g., 010100 if you were born on the 1st January 2000) to be able to match your questionnaire with subsequent data sets.

QUALIFICATION: Degree [ ] Master's degree [ ] in _____

YEAR: 1º [ ] 2º [ ] 3º [ ] 4º [ ]

SEX: Male [ ] Female [ ] Other [ ] **AGE**: _____

PRIOR TEACHING EXPERIENCE (not as a private teacher): Yes [ ] No [ ]

IF SO, PLEASE, SPECIFY: Educational level: _____(e.g., Primary Education).

Time:_____ months / years (select).

ANSWER THE QUESTIONNAIRE TAKING INTO ACCOUNT HOW YOU THINK CINEMA CAN CONTRIBUTE TO THE TEACHING OF ENGLISH AS A FOREIGN LANGUAGE

(1) Strongly disagree (2) Disagree (3) Agree (4) Strongly agree	
Cinema is a useful resource for the teaching of English as a foreign language because	
1. it helps improve students' writing skills	1 2 3 4
2. it helps improve students' speaking skills	1 2 3 4
3. it helps improve students' reading skills	1 2 3 4
4. it helps improve students' listening skills	1 2 3 4
5. it helps improve students' lexical knowledge	1 2 3 4
6. it helps improve students' grammatical knowledge	1 2 3 4
7. it helps improve students' pronunciation	1 2 3 4
8. it helps improve students' knowledge of the L2 culture	1 2 3 4
9. it helps improve students' knowledge of their own (L1) culture	1 2 3 4
10. it allows students to express their own opinions and feelings	1 2 3 4
11. it allows students to understand other people's emotions and feelings	1 2 3 4
12. it allows for the promotion of social justice	1 2 3 4
13. it allows for the development of students' higher order thinking skills	1 2 3 4
14. it elicits aesthetic pleasure	1 2 3 4
15. it is authentic material	1 2 3 4
16. it tends to deal with nontrivial topics	1 2 3 4
17. it is audio-visual material	1 2 3 4

PLEASE, CHECK YOU HAVE ANSWERED ALL THE QUESTIONS

———————————

*Adapted from Férez Mora et al. (2020).

Contents:

Universidade Vigo