The Effect of an App-Based Training on L2 Fluency and Foreign Language Anxiety

Foekje Reitsma

The Effect of an App-Based Training on L2 Fluency and Foreign Language Anxiety

Foekje Reitsma

The Effect of an App-Based Training on L2 Fluency and Foreign Language Anxiety

Von der Carl von Ossietzky Universität Oldenburg,

Fakultät III Sprach- und Kulturwissenschaften,

zur Erlangung des Grades einer

Doktorin der Philosophie (Dr. phil.)

genehmigte Dissertation

von

Foekje Martha Reitsma

Geboren am 3. August 1966 in Dokkum (Niederlande)

Vom Fachbereich Sprach- und Kulturwissenschaften der Carl von Ossietzky Universität Oldenburg als Dissertation angenommen.

Referentin: Prof. Dr. Esther Ruigendijk

Korreferentin: Prof. Dr. Katrin Kleinschmidt-Schinke

Tag der Disputation: 05.03.2025

Acknowledgements

I am grateful for the opportunity to do doctoral research and devote myself to a topic and question that fascinate me and in which theory and teaching practice, research and design, could be combined. This dissertation would not have been possible without the support, commitment, and pleasant cooperation of the people around me. For this, I would like to thank all of them.

First of all, I would like to thank Ralf Grüttemeier and Esther Ruigendijk. About fifteen years ago, Ralf Grüttemeier suggested to me that writing a dissertation might be something for me. At that time, I felt it was too early because our children were still too young. When he asked me again six years ago, I decided to take the step, and I am still glad every day that I did. Writing this dissertation has been enriching in many ways from beginning to end.

The past few years have been hard work alongside my job as a language teacher, lecturer in didactics, and teacher educator, but it also brought me a great deal of joy. I have had to update my knowledge and skills very much in various areas—such as state of the research, statistics, writing texts in English, and developing an app—due to the many years between finishing my studies at the University of Groningen in 1992 and starting my dissertation journey in 2019. I would therefore like to thank my supervisor, Esther Ruigendijk, for her pleasant and highly professional guidance throughout this process in acquiring these skills. Her supervision style suited me very well: supportive and scientifically encouraging, adjusting where needed, but also giving me a lot of freedom to work on my research theme. I would also like to thank my second referee, Katrin Kleinschmidt-Schinke, for the time she took to read and evaluate my work.

I am also very grateful to Marije Michel for her support and feedback in developing the app, as well as her input and feedback on the article about the effects of the app intervention. She pointed me to many interesting articles, and I appreciate her enthusiasm with which she approaches research. Many thanks also to the team at the Institut für Niederlandistik in Oldenburg and the members of the linguistics colloquium in Niederlandistik. I thank Benedicte for showing me the way in R. Ankelien, Andreas, Katrin, Marijke, Matthias and Roos thank you for carefully reading parts of my dissertation and for the pleasant time. I would also like to thank all the regular participants of the 'Doktoranden Kolloquium' for their questions and feedback on all the presentations I gave.

For conducting my second study, I developed an app together with Team MSML from Deventer (Netherlands): Marcel Lethink, Jasper Jorna, and Leontine Voerman. I want to thank you for your tremendous effort, enthusiasm, and expertise in creating the app. However, the app would not have come about without funding from Taalunie and Ems Dollard Regio. For this I would like to thank the Taalunie (Helmut van Berlo and Jo Sterckx) and EDR (Eske Kadijk) very much. I would also like to thank Julia Grimm for her practical assistance during the visits to the participating schools. Thanks to Victor Bernal for his advice and help with statistics, particularly in building multivariate analyses. I also want to thank the students of the Dutch Language Course, Nederlands 2b, Summer Semester 2020, for testing the app.

I would not have found all the interesting results without the participants. Many thanks to the participating schools, teachers, and students: Helmut Kuiter (TGG Leer), Soeren Heinsohn (Gymnasium Rhauderfehn), Kerstin Marwede (IGS Aurich), Nadine Diegel (IGS Kreyenbrück), and Tabea Zebrowski (IGS Flötenteich), who gave me the opportunity to recruit participants, let them work with the app, and record their speaking performances.

Outside the university, I would like to thank my dear parents, *heit* Jan and *mem* Hinke, family, and friends for their interest in my research, and their support. My final words are for my family, my dear children, Jan Siebrant, Jorien, and Nynke, who constantly remind me that there are far more important things in life than writing this dissertation. Thank you for your love, support, and patience during the times when there was still something to read or write. Finally, thank you, Brant, for your love, encouragement and support.

I would also like to thank everyone—my parents; Brant; Jan Siebrant and Penda, Jorien and Olaf, Nynke and Daniel; my sister Ina; my brother Koos; Wilt and Annelies; and Gertjan and Suzanna—who traveled to Oldenburg on March 5, 2025, to celebrate with me after my dissertation defense, making this day all the more special and memorable.

Zusammenfassung

Diese Dissertation befasst sich mit dem Konzept des flüssigen Sprechens (*fluency*) in einer Zweitsprache (L2) und deren Entwicklung im Unterricht. L2-Sprecher sprechen oft in einem langsameren Tempo mit mehr Pausen und Wiederholungen als Erstsprachler (Kahng, 2014; Riazantseva, 2001). Diese Unflüssigkeiten beim Sprechen in der L2 hängen bis zu einem gewissen Grad mit dem Sprachniveau (*proficiency*) zusammen. Sie kann jedoch auch durch kognitive, soziale und affektive Faktoren beeinflusst werden (Kormos, 2006; Lennon, 2000; Segalowitz, 2010). In dieser Dissertation konzentriere ich mich auf den affektiven Faktor Fremdsprachenangst (*Foreign Language Anxiety,* FLA). Es wird angenommen, dass FLA kognitive Prozesse (wie z. B. das Formulieren), die, während der Sprachverarbeitung Aufmerksamkeit erfordern, beeinträchtigen kann. Einige Studien deuten darauf hin, dass sich FLA negativ auf *fluency* auswirkt (MacIntyre & Gardner, 1994; Pérez Castillejo, 2019, 2021; Sanaei et al., 2015; Bielak, 2022).

In dieser Dissertation wurde eine psycholinguistische Perspektive auf *fluenc*y (erste Studie) mit einer pädagogischen Perspektive (zweite Studie) kombiniert. In der ersten Studie wurde untersucht, inwieweit gibt es eine Beziehung zwischen *speed, breakdown* und *repair fluency* Messungen und der FLA und *proficiency*, und wie interagiert dies mit der Aufgabenkomplexität? Einundsechzig deutsche Niederländischlerner der Sekundarstufe I füllten einen Fragebogen zur Fremdsprachenangst und einen Leistungstest aus und führten zwei Sprachproduktionsaufgaben (eine einfache und eine komplexe) durch. Korrelationsanalysen zeigten, dass FLA negativ mit der *mid-clause* Pausen bei der komplexen Aufgabe zusammenhing. *Proficiency* stand in einer einfachen Aufgabe in einem positiven Zusammenhang mit verschiedenen Messungen der Geschwindigkeit (*speed fluency*) und der Pausen (*breakdown fluency*). Lineare gemischte Modelle zeigten, dass *proficiency* ein signifikanter Prädiktor für zwei Messungen der *fluency* war, aber Angst war kein signifikanter Prädiktor für irgendeine Messung der *fluency*, was mit dem relativ niedrigen Angstniveau der Teilnehmer zusammenhängen könnte.

Die Ergebnisse der ersten Studie und frühere Untersuchungen zum Zusammenhang zwischen FLA und *fluency* wurden in der zweiten Studie verwendet. Die zweite Studie untersuchte *fluency* aus einer pädagogischen Perspektive, und zwar die Effekte von App-basiertem Training

auf die Entwicklung von L2 *fluency* und FLA. Neunundachtzig deutsche Niederländischlerner, aufgeteilt in drei Gruppen, nahmen an der Studie teil. Gruppe 1 nahm an einem App-basierten Fluency-Training teil, gefolgt von einem App-basierten Sprechtraining ohne Fokus auf *fluency*; Gruppe 2 begann mit einem App-basierten Sprechtraining und nahm dann an einem App-basierten Fluency-Training mit Fokus auf *fluenc*y teil; während die Kontrollgruppe am regulären Unterricht ohne Training mit der App teilnahm. Die Daten wurden über einen Zeitraum von fünf Monaten mit einem Pretest-Posttest-Delayed-Posttest-Design erhoben. Die mündlichen Daten wurden für verschiedene Messungen der *fluency* analysiert. Die Ergebnisse zeigten, dass beide experimentellen Gruppen signifikante Fortschritte bei der *speech rate*, *articulation rate*, *phonation time ratio* und der Anzahl der *end-clause* Pausen machten, wobei die Gruppe, die mit dem App-basierten *fluency*-Training begann, signifikant mehr Fortschritte bei der *articulation rate* machte. Die Kontrollgruppe machte nach fünf Monaten deutlich weniger Fortschritte im flüssigen Sprechen. Zusammenfassend lässt sich sagen, dass die Studie die Vorteile eines App-basierten Trainings für das Erlernen der niederländischen Sprache in Deutschland zeigt.

Abstract

This dissertation examines the concept of fluent speech (fluency) in a second language (L2) and its development in instructional contexts. L2 speakers often speak at a slower rate and produce more pauses and repetitions than native speakers (Kahng, 2014; Riazantseva, 2001). These disfluencies in L2 speech are to some extent related to proficiency level, but they may also be influenced by cognitive, social, and affective factors (Kormos, 2006; Lennon, 2000; Segalowitz, 2010). In this dissertation, the focus is on the affective factor of foreign language anxiety (FLA). It is assumed that FLA can interfere with cognitive processes that require attention during L2 speech processing, such as formulation. Several studies suggest that FLA has a negative impact on fluency (MacIntyre & Gardner, 1994; Pérez Castillejo, 2019, 2021; Sanaei et al., 2015; Bielak, 2022).

This dissertation combines a psycholinguistic perspective on fluency (Study 1) with a pedagogical perspective (Study 2). The first study investigated the extent to which measures of speed, breakdown, and repair fluency are related to foreign language anxiety and proficiency, and how these relationships interact with task complexity. Sixty-one German learners of Dutch in lower secondary education completed a language anxiety questionnaire and a proficiency test and performed two oral production tasks (one simple and one complex). Correlational analyses showed that FLA was negatively related to number of mid-clause pauses in the complex task. In the simple task, proficiency was positively related to several measures of speed fluency and breakdown fluency. Linear mixed-effects models revealed that proficiency was a significant predictor of two fluency measures, whereas anxiety was not a significant predictor of any fluency measure, which may be related to the relatively low anxiety levels of the participants.

The findings of the first study and previous research on the relationship between FLA and fluency informed the second study. This study examined fluency from a pedagogical perspective by investigating the effects of app-based training on the development of L2 fluency and foreign language anxiety. Eighty-nine German learners of Dutch, divided into three groups, participated in the study. Group 1 received app-based fluency training followed by app-based speaking lessons without an explicit focus on fluency; Group 2 began with app-based speaking lessons and subsequently received app-based fluency training with an explicit focus on fluency; the control group followed regular classroom instruction without app-based training.

Data were collected over a five-month period using a pretest–posttest–delayed posttest design. Oral data were analyzed using multiple fluency measures. The results showed that both experimental groups made significant gains in speech rate, articulation rate, phonation time ratio, and the number of end-clause pauses, with the group that started with app-based fluency training showing significantly greater gains in articulation rate. The control group showed substantially smaller gains in fluency over the five-month period. Overall, the study demonstrates the benefits of app-based training for learning Dutch as a foreign language in Germany.

Table of Contents

Preface

As a lecturer in Dutch as a foreign language and a teacher educator, I have always been particularly interested in speaking skills. I have observed that students not only enjoy practicing speaking in a foreign language, but that doing so also enables them to connect with others and engage with different cultures.

However, both beginning and experienced teachers indicate that teaching speaking skills is challenging due to time constraints and large classes, but especially in how to engage all students actively and involved in speaking exercises. Often, during a speaking task, only a few students are talking while others say only a few words and mostly listen. Another challenge is that it is important for motivation that students experience progress. For this, providing feedback is essential. However, giving individual feedback after a speaking task is not easy because of the transient nature of speaking, especially in classes with many students. Together, this makes it difficult to get a grip on students' learning process. Additionally, there are students in every class who feel anxious speaking in the foreign language and withdraw during speaking tasks. For a teacher, it is a challenge to get these students talking as well.

These experiences from teaching practice led me, five years ago when I had the opportunity to start a PhD project, to choose a project that is relevant for research but also practically relevant to teaching practice. This gave me the idea to investigate whether digital technologies could provide a possible solution to the challenges I described above. That led me to investigate in this dissertation whether using an app affects learners' fluency and speaking anxiety. With this dissertation I hope to contribute to research-based knowledge in the field of foreign language teaching and learning.

1. General Introduction

1.1 Introduction

This dissertation is about the concept of fluency in a second language (L2) and its development in a classroom setting. Fluency refers to ease or automaticity in the learner speech production and is manifested in flow, continuity, and smoothness of speech (Segalowitz, 2010; Skehan, 2014).

For many L2 learners, even after a substantial period of study, it remains difficult to speak fluently. L2 speakers speak often at a slower pace with more pauses and repairs than first language (L1) speakers (e.g., Kahng, 2014; Riazantseva, 2001). These disfluencies in L2 speech are, to some extent, related to proficiency level. Tavakoli et al. (2020) found, for example, that at lower levels of proficiency, speakers produce more pauses at mid-clause positions than speakers at higher levels of proficiency. However, these disfluencies may also be influenced by cognitive, social and affective factors (Kormos, 2006; Lennon, 2000; Segalowitz, 2010). In this dissertation, I focus on the affective factor foreign language anxiety (FLA). FLA is characterized as a type of anxiety that is situation-specific and linked to a language-learning setting (Horwitz et al., 1986). L2 learners who feel anxious perceive speaking in an L2 to be an uncomfortable experience and experience more difficulty demonstrating the skills and knowledge they possess (Horwitz et al., 1986). A body of research, including meta-analyses, has shown that FLA is negatively related to L2 performance, including speaking performance (Botes et al., 2020; Teimouri et al., 2019; Zhang, 2019). It is thought that FLA may interfere with cognitive processes (e.g., formulation) that require attention during speech processing. Some studies indicate that FLA negatively affects speech fluency (MacIntyre & Gardner, 1994; Pérez Castillejo, 2019, 2021; Sanaei et al., 2015; Bielak, 2022).

From a developmental perspective, it is believed that greater L2 fluency comes from having access to a wider range of lexical and grammatical knowledge, but primarily through a more automatized access to that knowledge, developed through practice (Kormos, 2006). This means that for language learning in classroom settings, teachers should create opportunities for students to produce language to develop speaking fluently (Derwing & Munro, 2013).

Dörnyei and Kormos (1998) state that learners struggling with speaking fluently could benefit from fluency strategies like the use of fillers (e.g., "well," "you know") to maintain the flow of speech and provide them with more processing time. There is some evidence in the literature (Tavakoli et al., 2016; Seifoori & Vahidi, 2012) that teacher-led instruction on such strategies with printed learning materials may improve fluency. Given the technological advancements in recent decades, strategy instruction might also be implemented using technology-enhanced language learning (TELL). TELL involves the use of digital devices and software applications in L2 settings (Chang & Hung, 2019). Studies have found that the use of TELL applications has a positive impact on L2 development in terms of improving listening (e.g., Tan et al., 2019), reading (Darling-Aduana & Heinrich, 2018), and writing proficiency (e.g., Chen et al., 2017).

The use of TELL applications seems to have several advantages. Hauck and Hurd (2005) argue that “anxiety levels are likely to be lowered if students can learn in a non-threatening environment which encourages them to try things out and have fun, which builds confidence and promotes respect for different learning styles, approaches and personality traits” (p.17). In addition, TELL can be a means for providing individualization in a large class, self-pacing and private space to make mistakes (Brown, 2007). Furthermore, Automatic Speech Recognition (ASR) can be integrated into TELL. ASR refers to technologies that allow for the identification of spoken words and conversion into text (Cardoso, 2022). Learners can speak into a microphone, and ASR can analyze their speech, offering immediate and individualized feedback and suggestions for improvement. ASR has shown promise in improving pronunciation, but its application in fluency studies remains underexplored (McCrocklin, 2016; Golonka et al., 2014). To date, there is only limited research evidence (Tecedor & Campos-Dintrans, 2019; Wang et al., 2018; Grimshaw & Cardoso, 2018) that the use of TELL applications can positively impact L2 fluency development. However, these studies do not focus on fluency instruction for improving L2 fluency. To address these gaps, this dissertation investigates the effects of an app-based intervention drawing on the mobile app *Reppen* (see Chapter 5). This app is specifically designed and developed to train L2 fluency (see Chapter 3). In the app, users individually complete the fluency training. They acquire an understanding of what constitutes fluency and how it can be enhanced using fluency strategies, which they subsequently apply in speaking tasks. After each task, the app analyses speech recordings using ASR and provides feedback on fluency-related aspects, such as speech rate and pausing.

The aim of this dissertation is to accumulate experimental evidence on the role of FLA on L2 fluency, as well as on the effect of an app-based intervention on L2 fluency and FLA. In this dissertation, FLA is examined from two perspectives: first, whether FLA leads to reduced fluency (Chapter 4), and second, whether an app-based training results in a reduction of FLA and more fluent speech (Chapter 5). Accordingly, this dissertation addresses the following research question:

To what extent is there a relationship between FLA and L2 fluency and to what extent is there an effect of an app-based intervention drawing on the mobile app Reppen on L2 fluency and FLA?

1.2 Outline of this dissertation

The outline of this thesis is as follows. In Chapter 2, the relevant background literature on the concepts of fluency and FLA, and the relationship between FLA and L2 fluency, is discussed. Additionally, it presents an overview of relevant literature on L2 fluency development in instructional settings including the use of TELL. In Chapter 3, I present the *Reppen* app, which I developed for the study described in Chapter 5. This app consists of L2 fluency training and speaking lessons. In Chapter 4, the relationship between FLA and L2 fluency is examined among German learners of Dutch in relation to task complexity. In Chapter 5, the effects of app-based training on L2 fluency development and FLA among German L2 learners of Dutch is investigated. In the general discussion in Chapter 6, I discuss the main findings of the studies, methodological challenges and choices in app-based research, suggest theoretical and pedagogical implications, and provide directions for future research.

Reading guide

The main chapters (Chapters 4 and 5) of this dissertation have been written as individual papers: each chapter can be read on its own. As a result, there will be some overlap in the method sections and literature overviews. An adapted version of Chapter 4 is accepted for publication Reitsma & Ruigendijk (2024) in *Dutch Journal of Applied Linguistics*, and an adapted version of Chapter 5 is submitted to a journal.

2. Psycholinguistic and pedagogical background

2.1 Introduction

To the question *what did you do this weekend*, a German learner of Dutch replies:

Eh … zaterdag heb ik heel veel geleerd omdat we donderdag een …. …. ehm.. .. . ... Arbeit schrijven .. En 's avonds was ik bij mijn vrienden en .. zondag eh .. heb ik ook weer geleerd ... En dan eh … ben ik met mijn Moped .. en een vriend eh …. gefahren naar Leer .. .. En daar .. zijn wir … zijn we eh … een beetje eh een beetje …. gefahren … … en 's avonds zijn we eh … eten gegaan … naar Bingum … En … daarna heb ik ook weer een beetje geleerd en dan … heb ik tv gekeken en ben slapen gegaan.

(Translated into English[1]: *Uh ... on Saturday I studied a lot because on Thursday we have to write a ... uh ... Arbeit... And in the evening I was with my friends and … on Sunday uh ... I studied again.. And then uh ... I have gefahren with my Moped .. and a friend eh.. to Leer... And there .. wir have … we have uh... gefahren around a bit... And in the evening we went out for dinner… in Bingum… and ... after that I studied a little bit and then... I watched TV and went to sleep)*

What stands out in this excerpt is the disfluent nature of spontaneous speech, as indicated by many silent pauses, filled pauses (*eh, ehm*), corrections (*zijn wir, zijn we*), and repetitions (*een beetje, een beetje*). In addition to a few errors in grammar (e.g. *'s avonds zijn we eten gegaan* instead of '*s avonds zijn we gaan eten*), the learner also uses some German words (*Arbeit, Moped, wir, gefahren*) to express themselves when talking Dutch.

Disfluencies, such as silent and filled pauses, repetitions, and self-corrections, are common in the spontaneous speech of both L2 and L1 speakers (Kahng, 2022). While speaking, a speaker needs to translate their thoughts into intelligible sounds quickly. This rapid

[1] The German learner of Dutch also uses German words. I have not translated these words into English but underlined them.

translation of thoughts into speech occurs in roughly three stages: conceptualizing what to say, formulating how to say it, and finally articulating the appropriate phonemes (Levelt, 1998, 1999). For L1 or highly proficient L2 speakers, two of the three speech processes, formulating and articulating, are largely automatized, allowing the speech processes to run incrementally, meaning that speakers conceptualize and formulate the next part of the message while articulating the current one. If the speaker encounters a problem during any stage of the speech production process, they may become disfluent. This may involve the speaker stopping speaking (and being temporarily silent), using a filled pause, slowing down the current articulation, or using a repetition. In general, L1 speakers are more likely to become disfluent when conceptualizing and/or linguistic formulation is relatively difficult. For instance, disfluencies are more likely to appear before major sentence constituents or at syntactic boundaries (Swerts, 1998). Within clauses, pauses are more likely to occur before open-class words (Maclay & Osgood 1959), low-frequent words (Kircher et al. 2004), or in situations where speakers have several lexical options to choose from (Hartsuiker & Notebaert, 2010). In addition, an L1 speaker also becomes disfluent when, while monitoring their own speech, they notice an error that requires correction.

Producing speech in an L2 resembles the processes involved in speaking one's L1 because speech production in an L2 also involves conceptualization, formulation, and articulation (Levelt, 1989, 1999). But there also differences. One important difference between L1 and L2 speech is that in L1, much of the speech production is automatic and happens in parallel, whereas in L2 processing especially at lower proficiency levels, speech production is not yet automatic, and the different processes may not happen in parallel (Tavakoli & Hunter, 2018). The lack of automaticity and parallel processing makes the speech production process more intensive, resulting in slower speech, characterized by more frequent pauses, particularly in the middle of clauses (de Jong, 2016). Automaticity refers to the efficiency (not necessarily the speed) with which planning, encoding (selecting the words and syntactic structures needed), and articulating the message occur (Segalowitz, 2010; Zuniga, 2015).

Two possible sources are thought to be responsible for disfluencies in L2 speech. First, L2 speakers may experience difficulties because of incomplete linguistic knowledge of the L2 (e.g., a limited L2 vocabulary, unknown grammatical rules etc.). Second, L2 speakers may also have insufficient skills with which L2 knowledge is used (e.g., lexical access, speed of articulation, etc.). Both incomplete linguistic knowledge of the L2 and slow access to that knowledge may be causes why two of the three speech processes—formulation and articulation—are not as automatized in L2 speakers as in L1 speakers. Consequently, these processes require a lot of working memory capacity and often cannot occur in parallel with other speech processes, which can result in disfluencies, such as a slower speech rate.

Fluency can thus be influenced by incomplete linguistic knowledge of the L2 and slow access to that knowledge, but it may also be influenced by affective factors (Kormos, 2006; Lennon, 200; Segalowitz, 2010), such as foreign language anxiety (FLA). Studies into the relationship between FLA and L2 fluency show that high-anxiety L2 speakers tend to speak less in a given task (MacIntyre & Gardner 1994; Pérez Castillejo 2019; Sanaei et al. 2015), are perceived as less fluent (MacIntyre & Gardner 1994), pause more frequently within clauses (Pérez Castillejo 2019, 2021) and produce shorter stretches of meaningful syllables (Pérez Castillejo 2019, 2021; Sanaei et al. 2015; Bielak, 2022).

The present dissertation studies fluency from the perspective of the L2[2] speaker, investigating where dysfluencies come from (speech production processes). In this chapter, I discuss two views on fluency: fluency as a phenomenon that reveals underlying processes in L2 speaking (a psycholinguistic view) and fluency from a pedagogical perspective, focusing on how to support gains in fluency within the limited practice opportunities provided by language classes. This dissertation combines a psycholinguistic and pedagogical view to come to a better understanding of the relationship between FLA and L2 fluency and the development of L2 fluency. Below, both perspectives will be introduced, and it will be explained how the present dissertation combines both.

[2] In this dissertation, L2 refers to the acquisition of a language that occurs after the age of approximately four years, following the acquisition of the L1. L2 can also refer to a third, fourth language etc..

2.2 Psycholinguistic perspective on L2 fluency

2.2.1 Defining fluency

In this section, I will first discuss what it means to be fluent in an L1 and then explore what it means to be fluent in an L2. I will explain that 'fluent' can have both a 'broad' and a 'narrow' definition in an L2, before suggesting that Segalowitz's (2010) conceptualization of fluency may be useful in bringing together the different definitions of L2 fluency into one model.

Fillmore (1979) identified four dimensions in which L1 speakers can be considered fluent. The first dimension of fluency is “the ability to talk at length with few pauses” (p. 93). He explains that this means not having to pause to think of what to say and mentions professions such as radio presenters and sports reporters as examples of people who need this kind of fluency to do their jobs. The second dimension Fillmore describes is the ability to speak coherently using "semantically dense" sentences, with few unnecessary words or fillers (1979, p. 93). The third dimension involves “the ability to have appropriate things to say in a wide range of contexts” (p.93). This refers to the ability knowing what to say and how to speak in different places with different people. The final type of fluency considers the ability to be creative with language—displayed through jokes, metaphors, and the use of different styles. Fillmore’s (1979) dimensions focus not only on the form of speech but also on its content (e.g., relevance or coherence) (Bosker, 2014).

To distinguish between the form and content of speech, Lennon (1990) defined fluency in two ways: broad and narrow. Fluency in the *broad sense* is often used, for example, in statements such as “Daniela speaks Dutch fluently.” It relates to someone’s overall speaking proficiency (Chambers, 1997), which may refer to anything from error-free grammar to large vocabulary size or near-native pronunciation skills. In contrast, in much of the research literature, fluency is regarded as one of the components of speaking proficiency, along with accuracy of form and complexity of linguistic units (both syntactic and lexical) - brought together in the CAF (Complexity, Accuracy, Fluency) framework for exploring language ability (Skehan, 1998). In this context, fluency refers to Lennon’s (1990*) narrow’* definition of fluency. Fluency in this sense has been defined as an “impression on the listener’s part that the psycholinguistic processes of speech planning and speech production are functioning easily and smoothly" (Lennon, 1990, p. 391). This narrow sense is often found in oral exams (see also 2.3): in addition to grammar and vocabulary, the flow and smoothness of the speech is also assessed. This dissertation focuses on this narrow sense of fluency.

There are numerous definitions of fluency in the narrow sense. The above-cited definition by Lennon (1990) focuses primarily on how speech is interpreted by the listener. In a later publication Lennon (2000) introduced another definition on fluency as "the rapid, smooth, accurate, lucid, and efficient translation of thought or communicative intention into language under the temporal constraints of on-line processing" (Lennon, 2000, p. 26). This interpretation of fluency focusses on the quality of the speech. It suggests that fluency involves not only speed but also precision and clarity in speech. Housen and Kuiken (2009) focus on fluency as it relates to language learners. They define fluency as being "primarily related to learners' control over their linguistic L2 knowledge, as reflected in the speed and ease with which they access relevant L2 information to communicate meanings in real time" (p. 462). This definition underscores the importance of language learners' ability to access and use their L2 knowledge quickly and effectively during communication. Just like Lennon's definition (2000), fluency here is associated with cognitive speech production processes, such as linguistic control and lexical access. Skehan (2009) provides a different perspective by defining fluency as "the capacity to produce speech at normal rate and without interruption" (Skehan, 2009, p. 510). This definition shifts the focus toward the observable qualities of speech itself. Here, fluency is an acoustic phenomenon that can be measured (Bosker, 2014).

The multiplicity of definitions illustrates the complex and multidimensional nature of fluency. Nevertheless, it is possible to distinguish various patterns. Some studies emphasize the efficiency of the cognitive processes responsible for (dis)fluency, while others focus on the acoustic effects of these cognitive processes for the spoken utterance. Still, others focus on the impact that (dis)fluent speech can have on the listener. Segalowitz (2010) attempted to differentiate these various interpretations of fluency for L2 in one framework, which will be described in the next section.

2.2.2 An L2 fluency framework

Segalowitz's fluency framework (2010, 2016) brings together insights from various scientific disciplines, including behavioral and brain sciences and social sciences. Segalowitz argues that sociolinguistic factors (social context), psycholinguistic factors (neurocognitive processes involved in speech production), and psychological factors (such as motivation and anxiety) are interconnected within a dynamic system and contribute all to a speaker's level of fluency. This perspective views fluency as a complex, ever-changing phenomenon.

He describes a framework that distinguishes three subdomains of fluency: *cognitive fluency*, which refers to "the efficiency of operation of the underlying processes responsible for the production of utterances"; *utterance fluency*, which refers to "the features of utterances that reflect the speaker's cognitive fluency" and can be measured acoustically; and *perceived fluency*, "the inferences listeners make about speakers' cognitive fluency based on their perceptions of the utterance fluency" (Segalowitz, 2010, p. 165). Using the fluency framework of Segalowitz (2010), I will summarize the literature on (the relationships between) cognitive, utterance, and perceived fluency.

Cognitive fluency

Cognitive fluency refers to a speaker´s capacity to use the underlying cognitive processes responsible for fluent speech production (e.g. efficiency of lexical retrieval, grammatical/phonological encoding (Kahng, 2022). Segalowitz adopts the speech production model of Levelt (1989, 1999), who describes the process of L1 speech production and offers a "blueprint" for the native speaker. Levelt considers speech production to be the outcome of several components working in a modular way. These components are regarded as relatively autonomous and specialized in the functions they perform. Each component transmits its output to the next in a sequential order. There are two main types of components: knowledge and processing components. The knowledge components represent declarative knowledge, including speaker's knowledge of the world, as well as models of discourse and interaction situations. A separate language-specific knowledge component, the lexicon, is included in the model and is used in later stages of language production and monitoring. The processing components, on the other hand, comprise procedural knowledge that processes the information from the knowledge stores to produce speech.

The three main components of speech production are the conceptualizer, the formulator, and the articulator. The start of the speech process begins with the conceptualizer, where a communicative intention is turned into concepts. Message generation considers information about the conversational setting and discourse model, which includes the selection of a linguistic register. Conceptualization includes macroplanning and microplanning. Macro-planning is the ‘process by which the speaker decides what to say next’ (Levelt, 1999, p. 92), further refined in micro-planning to select appropriate semantic information or ‘perspective’ (p.94). Within the conceptualizer, the message is generated and monitored internally to see if the (preverbal) plan matches the intended message. Finally, the set of concepts is turned into the preverbal message, which is transferred to the formulator.

In the formulator stage, the preverbal message is encoded into a grammatical form, resulting in a surface structure of the to-be-produced utterance. The surface structure forms the input to morpho-phonological encoding (choosing the right words with the correct word forms) and phonetic encoding (building an appropriate phonetic gestural score). The output of the formulator is the input of the articulator, which transforms the phonetic plan, into actual speech, which is achieved through motor control of the muscles needed to produce sound. In addition, the speech plan generated from the formulator is also sent to the speech comprehension component for internal monitoring. This component controls for errors and can give feedback to the conceptualizer if corrections or alterations are necessary.

Processing in the formulator and the articulator is highly automatic, while planning processes in the conceptualizer and monitoring processes require attention of the speaker. The system uses parallel processing when producing speech. For example, once the conceptualizer transmits the pre-verbal message to the formulator, it can immediately work on the next information that will be generated, while the formulator and the other components are simultaneously processing information. In L1 or highly proficient L2 speech, these processes are automatic and operate in parallel, taking place virtually simultaneously, and demanding little cognitive effort. These characteristics of L1 speech production explain the fluency, in a speaker's performance.

Levelt’s model was developed to explain L1 speech production, but several researchers have extended this framework to L2 speech production (De Bot, 1992, Segalowitz, 2010). De Bot’s and Segalowitz’s models both assume that the same basic psycholinguistic mechanisms - conceptualization, formulation, and articulation - underlie L1 and L2 speech production. De Bot (1992) suggested that the first process in L2 speech production, the macroplanning stage of conceptualizing, is language general. It is assumed that encyclopedic and social knowledge is not language-specific organized and works the same way regardless of the language in which the utterance will be produced. However, De Bot (1992) argues that the next stage, microplanning, is language specific, because different conceptual features need to be specified depending on the language spoken.

Segalowitz's (2010) model is designed to show how L2 speech is vulnerable to disfluencies at various points in the speech production process due to the additional processing load imposed by directing attention and effort to processes—formulation and articulation—that occur automatically in L1. Segalowitz agrees with De Bot that macroplanning is language general and does not involve any additional L2-specific processing difficulties. In contrast, the later stages of microplanning, formulating, and articulating are predicted to lead to L2-specific disfluencies because of deficits in L2 linguistic knowledge and less automatized processing (see also 2.1).

Utterance fluency

Within Segalowitz's (2010, 2016) framework, utterance fluency refers to observable temporal features, such as speed of delivery, pauses, and hesitations, which reflect the speaker's cognitive fluency (Suzuki et al., 2021). Over the years, researchers have identified numerous measurements associated with fluency, such as speech rate, mean length of run, number of repetitions per minute, number of silent or filled pauses per minute, mean length of pauses (see Table 1.1 from Segalowitz, 2010, p. 6). There is also a large diversity in the way researchers calculate specific measurements. To counter the multiplicity and diversity, measurements of utterance fluency are divided into three dimensions: speed, breakdown and repair fluency (Skehan, 2003, 2009; Tavakoli & Skehan, 2005). *Speed fluency* reflects the speed of delivery. *Breakdown fluency* refers to pausing behavior and is usually measured in terms of the frequency, duration, type, and location of pauses (Suzuki et al., 2021; Tavakoli & Wright, 2020). Finally, *repair fluency* includes for example self-corrections and reformulations.

The measures of utterance fluency that are most frequently used in research are presented in Table 1 (based on De Jong (2018, p. 240, Table 1), Derwing (2019, p. 247, Table 14.1, and Kormos (2006, p. 163, Table 8.2).

Table 1

Frequently used Measures of Utterance Fluency (Kahng, 2022)

Measure	Formula
Speech rate	Number of syllables / total time
Articulation rate	Number of syllables / (total time – silent pausing time)
Pruned syllables per second	(Number of syllables – number of filled pauses and repairs) / total time in seconds
Mean length of run	Mean number of syllables between silent pauses
Phonation time ratio	Speaking time / total time
Mean length of silent pauses	Pausing time / number of silent pauses
Number of silent pauses (per minute)	Number of silent pauses / total time
Number of filled pauses (per minute)	Number of filled pauses / total time
Number of repetitions (per minute)	Number of repetitions / total time
Number of repairs (per minute)	Number of repairs and restarts / total time

Compared to L1 speech, L2 speech tends to be slower and have more pauses and repairs (e.g., Kahng, 2014; Riazantseva, 2001). There is also a difference in the distribution of pauses. When compared to L1 speakers', L2 speakers use more pauses within clauses (Kahng, 2014; Tavakoli, 2011) and within Analysis of Speech (AS) units (De Jong, 2016; Skehan & Foster, 2007). Foster et al. (2000, p. 365) define an AS unit as "a single speaker's utterance consisting of an independent clause, or sub-clausal unit, together with any subordinate clause(s) associated with either. L2 speakers become more fluent as their proficiency in L2 increases (e.g., Freed, 2000; Towell et al., 1996). Studies on the relationship between utterance fluency and proficiency suggest that pure speed measures (e.g., articulation rate, mean syllable duration) and the number of silent pauses, especially those within clauses or AS units seem to be reliable indicators of speaking proficiency (Kahng, 2022). However, there is overlap between a person's fluency in the L2 and L1 in terms of speaking style and preferences (e.g. speed of speech, tendency to pause), as shown by De Jong et al. (2015). This means that when we measure aspects of L2 fluency, we partly measure personal speaking style (e.g. speed of speech, tendency to pause), in addition to L2-specific fluency that is related to proficiency.

In this dissertation, I use various fluency measurements to investigate the relationship between FLA and L2 fluency (Chapter 4) and the effect of app-based intervention on L2 fluency and FLA (Chapter 5). The fluency measures examined in this study are as follows: speech rate, articulation rate, mean length of run, phonation time ratio, number of silent and filled pauses, number of pauses, number of mid-clause and end-clause pauses, number of reformulations, mean length of pauses, mean length of mid- and end-clause pauses (more details will follow in Chapter 4 and 5).

Perceived fluency

Perceived fluency refers to the impression that a listener has of a speaker's fluency (Lennon, 2000; Segalowitz, 2010). Perceived fluency has been measured in various ways, such as by asking teachers (Freed, 2000; Kormos & Denes, 2004; Riggenbach, 1991; Rossiter, 2009) or native (Bosker et al., 2013; Derwing et al., 2009; Freed, 2000) or non-native (Rossiter, 2009) speakers to rate speech samples.

Some studies have allowed raters to make judgments on their own definitions (Freed, 2000; Kormos & Denes, 2004), which may lead to inconsistencies in their ratings (Rossiter, 2009). Others have provided definitions based on temporal aspects of fluency, such as speech rate, pauses, and self-corrections (Bosker et al., 2013). Previous research findings suggest that even when raters are instructed to focus on temporal features of fluency, their perceptions of fluency tend to be influenced by nontemporal features as well, such as grammatical errors (Kormos & Dénes, 2004; Rossiter, 2009; Suzuki & Kormos, 2020). Apparently, listeners make inferences about how efficiently the speaker conveys the intended message by paying selective attention to temporal features of fluency that they believe reflect the speaker's efficiency in mobilizing L2 knowledge for speech production (i.e., cognitive fluency; Segalowitz, 2010). L2 fluency research has extensively investigated which temporal features of fluency can explain listeners' perceptions of fluency. Saito et al. (2018) found that in assessing L2 speech, L1 raters used speed (articulation rate) as the primary indicator of fluency and pausing (mid-and end-clause) as a secondary indicator but did not pay attention to the incidence of repairs (repetitions and self-corrections). Thus, if speech is spoken at a speed listener are used to, and pauses are used as oral punctuation marks, listeners seem tolerant of repairs and do not consider them as disfluencies.

So far, the discussion of Segalowitz's (2010) fluency framework has focused on the three aspects of L2 fluency – cognitive fluency, utterance fluency, and perceived fluency. In the fluency framework, these aspects are situated within a broader theoretical perspective that can provide a basis for understanding the challenges that fluency poses to individual learners.

2.2.3 Affective factor foreign language anxiety and L2 speech production

According to Segalowitz's (2010) fluency framework, L2 speech performance and production ability can be subject to cognitive (i.e., age of acquisition, cognitive fluency, learning styles, and learning strategies), affective (i.e., motivation to communicate, anxiety), and social (i.e., learning context, social contact, and cultural interest) factors. Research shows that for L2 learners with similar L2 learning histories, variability in fluency is huge. For example, while some learners may speak English at near-native rates (approximately 240 syllables per minute), many others are much slower, with rates ranging from 110 to 230 syllables per minute (Mora & Valls-Ferrer, 2012). Thus, some speak fluently without apparent difficulties, while others pause frequently. Research on individual differences aims to understand and describe the sources of this inter-learner variability and identify the explaining factors (Mora, 2022). In this dissertation, I focus on the affective factor anxiety.

Foreign language anxiety

In L2 research, the construct anxiety has been referred to as Foreign Language Anxiety (FLA). (Baran-Lucarz, 2022). Early research defined FLA in general terms as either a relatively stable personality trait across various situations, or as a state, an emotional response to a situation at a particular time, e.g., during a high-stakes exam (Cattell et al., 1961). Later, based on observations of L2 learners in instructional settings, Horwitz et al. (1986) developed the construct of FLA and hinted at the complexity of the concept: "a distinct complex of self-perceptions, beliefs, feelings, and behaviors related to classroom learning arising from the uniqueness of the language learning process" (p. 128). Thereby, it is emphasized that FLA is not the same as general anxiety but rather a type of situation-specific anxiety that occurs in the context of language learning and/or language use (Horwitz et al., 1986; MacIntyre & Gardner, 1989, 1991; Horwitz, 2017). Horwitz et al. (1986) developed also the Foreign Language Classroom Anxiety Scale (FLCAS), which is the most widely adopted means of both capturing and measuring situation-specific anxiety among L2 learners (Teimouri et al., 2019). FLA is seen as a persistent negative emotion resulting from recurring experiences of L2 learning and use, and is negatively associated with L2 language achievement, as demonstrated in meta-analyses by Botes et al. (2020), Teimouri et al. (2019), and Zhang (2019).

Recently, a dynamic approach has been argued for in which FLA is understood as an emotion that is more transient and changeable on a moment-to-moment basis due to several intervening variables (Szyszka & Lintunen, 2023). MacIntyre (2017) emphasized that FLA "is continuously interacting with a number of other learner, situational, and other factors including linguistic abilities, physiological reactions, self-related appraisals, pragmatics, interpersonal relationships, specific topics being discussed, type of setting in which people are interacting, and so on" (MacIntyre, 2017, p. 16). Here, FLA is seen as a more dynamic and temporary condition (MacIntyre, 2017).

In this dissertation, FLA is operationalized as a situation-specific emotion triggered by L2 learning and use (see Chapter 4) to compare our results with previous research (e.g., Pérez Castillejo, 2019, 2021; Sanaei et al., 2015; Bielak, 2022). In addition, it has been considered as a more transient, task-specific emotion (see Chapter 5) to investigate whether training in app-based fluency strategies influences anxiety.

Below, literature on the relationship between FLA and L2 fluency, as well as between FLA, L2 fluency, and task complexity, will be summarized. The section begins with a discussion of the construct of working memory, attention and Attention Control Theory, which predicts that anxiety may interfere with cognitive processes that require attention during L2 speech processing.

Attention, proficiency, FLA and L2 fluency

During speech production, working memory and attention are distinct but closely interconnected cognitive functions. *Working memory* temporarily stores and processes information needed for planning the message, retrieving lexical items, and constructing syntactic structures (Baddeley, 2003; Levelt, 1989). *Attention* is a cognitive function that determines how limited cognitive resources are allocated to different aspects of speaking, such as meaning, linguistic form, or articulation (Skehan, 1998). Attention is essential for the effective use of working memory, as it selects which information is prioritized and processed.

According to an attention-based perspective (De Bot 1992; Kormos, 2006; Segalowitz, 2010) producing L2 speech requires the complex coordination of attentional resources. The allocation of these resources varies depending on the cognitive demands of the task (see below) and the degree of automaticity of linguistic processing. The degree of automatization of linguistic processing is strongly related to proficiency (Kormos, 2006; Segalowitz, 2010). More proficient L2 speakers typically have more automatized access to linguistic knowledge and therefore require fewer working memory resources for formulation and articulation, leaving more attentional capacity available for conceptualization. In contrast, less-proficient speakers must divide their attention between conceptualization, formulation, articulation, and monitoring, which increases cognitive load and the likelihood of disfluencies (Lambert et al., 2017).

Any circumstances that inhibit or limit *attention control*, that is, which make the management of attentional resources less efficient, may interfere with L2 speech processing and, in turn, affect L2 production (Kormos 2000, 2006; Zuniga 2015). Some attention-related aspects of L2 speech production include message planning, self-repairs, and word choice among possible alternatives (Segalowitz 2010). These are processes that require attention to use the L2 effectively. According to Attentional Control Theory (Eysenck, 2010; Eysenck et al., 2007), which builds on Processing Efficiency Theory (Eysenck & Calvo, 1992), anxiety may negatively affect performance by interfering with cognitive processes, such as lexical retrieval, that require attention control. Attentional Control Theory (Eysenck et al., 2007) explains the mechanisms of how anxiety – worry – reduces cognitive efficiency by shifting attentional resources from task-relevant to task-irrelevant thoughts. Worrisome thoughts consume the limited attentional resources of working memory, which are therefore less available for concurrent task processing. Worry is characterized by concerns about evaluation and failure, and anticipation of negative outcomes (e.g., Borkovec, 1994). According to Attentional Control

Theory, the negative effects of anxiety on performance are large in complex tasks because these tasks require high demands on the processing and storage capacity of working memory. Based on Processing Efficiency Theory (Eysenck & Calvo, 1992; Calvo & Eysenck, 1996), Kormos (2015) made further specific predictions in that high levels of anxiety may be related to more effort in retrieving words from the mental lexicon and encoding syntactic structures in the L2. In addition, efficient switching between the different processes of conceptualization, formulation, and monitoring (Levelt (1989, 1999); See 2.2.2) might also be more demanding when anxiety is high. In turn, higher levels of anxiety are hypothesized to be related to lower levels of fluency during speaking performance (Bielak, 2022; Kormos, 2006, 2015; MacIntyre, 2017; MacIntyre & Gardner, 1994; Segalowitz, 2010).

Studies that have focused on the relationship between FLA and L2 show strong negative correlations between FLA and speed fluency measures (i.e. more anxious participants had shorter runs between pauses and shorter phonation time) during a non-exam situation (Sanaei et al. (2015). Pérez Castillejo's (2019) examined how FLA and proficiency relate to L2 fluency during a final oral exam among learners with low-intermediate proficiency level. The results indicated medium and strong negative relationships between FLA and those fluency measures that reflect speech formulation/encoding, such as shorter phonation time, shorter runs, and more frequent pauses within clauses among more anxious participants. However, the negative relationship between FLA and the length of pauses between clauses, mainly linked to message conceptualization, was weaker. Regression analyses indicated that FLA was a stronger predictor of fluency than proficiency. Pérez Castillejo (2021) partly replicated Pérez Castillejo's (2019) results, especially those regarding the relationship between FLA and the fluency of speech formulation/encoding (i.e. mean length of run and number of mid-clause pauses). However, the results regarding the relationship between FLA and conceptualization fluency (the length of pauses) were not replicated. From this, the author concluded that formulation/encoding is more vulnerable to anxiety than conceptualization. Furthermore, Pérez Castillejo (2021) found that proficiency was a much stronger predictor of L2 fluency than FLA when fluency was measured in a task followed by another similar task, that is, in a condition of prior L2 processing.

So far, research has primarily focused on FLA in relation to L2 fluency in less cognitively demanding tasks (Sanaei et al., 2015; Pérez Castillejo, 2019, 2021), but we have limited insights into the relationship between FLA and L2 fluency in complex tasks. This section continues by discussing the concept of task complexity and the literature that has examined the relationship between FLA, L2 fluency, and task complexity.

Task complexity, FLA and L2 fluency

Regarding the role of *task complexity*, that is, the cognitive demands of a task (Michel, 2017), two theories are prevalent: Robinson's Cognition Hypothesis (Robinson 2001, 2005) and Skehan and Foster's (2001) Limited Attentional Capacity model. According to Robinson's Cognition Hypothesis, cognitively demanding tasks direct learners' attention towards language form. The need to meet the cognitive demands of a complex task results in learners using a more extensive and varied lexis and producing more complex linguistic structures. For example, a task with many elements, as opposed to just a few elements, is expected to evoke more specific lexis and more complex syntactic structures because all the different elements must be distinguished and compared. Thus, the increase in cognitive demands focuses the learner's attention on linguistic aspects. However, improved accuracy and linguistic complexity might come at the expense of fluency due to a higher processing load in complex tasks. In contrast, the Limited Attentional Capacity model predicts that cognitively demanding tasks may lead to trade-off effects, particularly between complexity and accuracy, due to competition for limited attentional resources (Skehan, 2009). According to this view, 'attentional limitations for the L2 learner and -user are such that different areas of performance compete with one another for the resources that are available' (Skehan & Foster, 2001, p. 205). Skehan and Foster (2001) argue that an increase in cognitive task demands puts pressure on the attentional system. Therefore, L2 learners must prioritize between linguistic complexity, accuracy, and fluency. Attending to one aspect of performance will lead to less attention to other aspects. Consequently, a task that leads to linguistically more complex L2 performance may negatively affect performance not only with respect to grammatical accuracy, but also with respect to fluency (de Jong et al., 2012).

The effect of task complexity on speech processing may vary depending on the proficiency and anxiety level of the speaker. Since the automaticity of linguistic encoding is related to proficiency (Kormos, 2006; Segalowitz, 2010), more proficient L2 speakers may have more attentional resources available to conceptualize the message than less-proficient speakers, who must divide their attention between conceptualization, formulation, and monitoring (Lambert et al., 2017).

Additionally, the effect of task complexity on speech processing might vary depending on the anxiety level of the speaker completing the task. According to Eysenck et al.'s (2007) Attentional Control Theory, the adverse effects of anxiety on performance are larger on tasks that impose substantial demands on the processing and storage capacity of working memory. Robinson's (2001) Cognition Hypothesis also states that individual learner differences, such as anxiety, have a greater influence on spoken task performance if tasks are more cognitively complex. To date, only Bielak (2022) investigated the relationship between FLA and fluency in a complex task among advanced L2 learners. Participants completed a less cognitively demanding and a more cognitively demanding task. Numerous negative correlations between FLA and fluency measures, which reflect formulation (i.e., more anxious participants had shorter runs and lower articulation rate), and conceptualization (more end-pauses), were found in the first (i.e., less complex) task, and only one negative correlation (more end-pauses) was found during the second (i.e., complex) task. Regression analyses showed that proficiency was a stronger predictor of fluency than FLA. The author concluded that this could be attributed to the advanced proficiency level of the participants and, consequently, their more automatized L2 use.

In summary, the aforementioned studies have demonstrated a link between FLA and L2 fluency. These studies indicate that the attention-limiting effects of anxiety can impose constraints on efficiency in cognitive fluency, particularly during the stages of formulation and encoding. However, further research is needed, especially concerning complex tasks and different levels of proficiency. The relationship between FLA and L2 fluency in low-intermediate L2 learners has not been investigated in relation to task complexity. Lower participant proficiency may enhance the effects of FLA due to less automatization (i.e., efficiency with which planning, selecting words and syntactic structures, and articulating message occur in L2 use (Segalowitz, 2010). This dissertation aims to fill this research gap by investigating which aspects of L2 fluency (i.e., speed, breakdown and repair fluency) are related to FLA and how this interacts with task complexity (See Chapter 4). To operationalize task complexity, we followed De Jong et al. (2012). In their opinion, complex tasks contain more elements than simple tasks; complex tasks concern a more general topic as opposed to simple tasks, which concern topics of personal life; and complex tasks involve more abstract notions as opposed to simple tasks, which involve mostly concrete notions.

In conclusion, from a psycholinguistic perspective, the literature review above shows that L2 fluency can be hindered by gaps in lexical and grammatical knowledge, as well as a lack of automatized access to that knowledge. In addition, FLA may contribute to disfluency by limiting attention control during speech processing, mostly on encoding/formulation. From a developmental perspective, greater L2 fluency is expected to develop both by having access to a broader range of lexical and grammatical knowledge, but primarily through a more automatized access to that knowledge, developed through practice (Kormos, 2006). The following sections focus on L2 fluency development from a pedagogical perspective.

2.3 L2 fluency from a pedagogical perspective

This section will discuss the relevance of paying attention to fluency in the classroom. Subsequently, research will be presented showing that fluency nevertheless receives little attention in language classes.

Paying attention to L2 fluency in the classroom is important for several reasons. Firstly, fluency has been a criterion since the early days of L2 speaking assessment (Fulcher, 2003). Fluency is included in rating scales of speaking exams (e.g., International English Language Testing System (IELTS), Internet-Based Test of English as a Foreign Language (TOEFL iBT) and language benchmarks for L2 communicative ability (e.g., Common European Framework of Reference for Languages (CEFR), Council of Europe, 2001). Looking more closely at the rubrics of the different speaking exams, it is noticeable that the role of fluency varies widely within the English-speaking exams. Fluency is regarded as a separate construct in Pearson Test of English Academic (PTEA); as a construct associated with pronunciation on the one hand, and with complexity and accuracy on the other (TOEFL iBT); and as a construct inseparable from coherence (IELTS). In addition, the descriptions in the scales related to fluency allow for subjective interpretation of fluency. For instance, terms like “well-paced” (TOEFL iBT) and “unnaturally lengthy” hesitations (American Council on the Teaching of Foreign Languages, ACTFL) can be interpreted in different ways (de Jong, 2018). These findings suggest that while there is consensus about the importance of fluency as a key component of L2 proficiency, the fluency descriptors in speaking exams are not carefully operationalized in several international language tests (de Jong, 2016, 2018; Tavakoli et al., 2017).

Secondly, production (or output) plays an essential role in fluency development. Swain's Output Hypothesis (1985) emphasizes that L2 learners need ample opportunities to produce output not only to test their hypotheses about the L2, learn from errors, and discover where there are gaps in their knowledge, but also to develop fluency. Without producing output, language learners cannot practice language, nor can they automatize their language skills. This means that in classroom settings, L2 learners should have ample opportunities to produce language to develop fluency. Thirdly, listeners can find it tiring and annoying to interact with highly dysfluent speakers (Derwing, 2017; Varonis & Gass, 1982). The consequence of this is that potential conversation partners may avoid communicating with dysfluent learners (Derwing, Rossiter & Munro, 2002), resulting in learners missing the opportunity to get a lot of practice.

Although speaking practice is necessary for fluency development (Segalowitz, 2010), it appears that various factors hinder practice in classroom settings, such as crowded classrooms, limited time, the need to develop other language skills, and a lack of knowledge on the part of the teacher about activities that improve oral fluency (Derwing, 2017; Rossiter et al., 2010; Tavakoli & Hunter, 2018). Research conducted by Rossiter et al. (2010) and Tavakoli and Hunter (2018) indicates that teachers associate fluency with the broad concept of global L2 proficiency according to Lennon (1990; see also 2.2.1).

Rossiter et al. (2010) examined textbooks and teacher manuals to determine whether and to what extent fluency activities were included in these books. They reported five types of activities that were designed to promote fluency: consciousness-raising activities, repetition activities, activities that promoted use of formulaic sequences and free production activities (also known as ‘general speaking activities’). Rossiter et al. (2010) reported that free production activities were the most popular in these textbooks, followed by formulaic sequences and repetition activities. Rossiter et al. concluded that the textbooks they examined were unbalanced in terms of fluency enhancing activities and did not pay adequate attention to promoting fluency in instructional settings. Tavakoli and Hunter (2018) investigated what teachers do in the classroom with regard to fluency development. They found that teachers were more likely to turn to "free communication activities' to improve students' fluency. Tavakoli and Hunter, in line with Rossiter et al. (2010: 345), conclude that "fluency, in its focused and narrow sense, might very well be neglected in the L2 classroom." Teachers view fluency as something that doesn't necessarily require separate instruction, but rather emerges as evidence of successful learning and practice of L2 grammatical and lexical structures.

Although fluency might be neglected in the L2 classroom in its focused sense, in contrast, research (e.g., N. de Jong & Perfetti, 2011; Lambert et al., 2017; Hunter, 2017; Tavakoli & Skehan, 2005; Wray, 2000, 2008; Wood, 2010, 2016) shows that classroom language techniques, such as task repetition, planning time, and teaching formulaic sequences, can support the development of fluency in a focused way (see section 2.3.1). This implies that there is a gap between research results and pedagogical practice (e.g., Chamber, 1997; Gatbonton & Segalowitz, 1988; Tavakoli & Hunter, 2018; Rossiter et al., 2010; Hunter, 2017), and that findings and developments in language research do not necessarily lead to immediate changes in language teaching materials and pedagogical practice. According to Tavakoli and Wright (2020), the reason for this lies in L2 teacher education programs, where little attention is paid to fluency. This results in teachers in training have little knowledge of fluency or what activities could be used to promote it. Tavakoli and Hunter (2018) suggest that the gap between teacher educators and researchers could be narrowed if they have a shared understanding of the concept of fluency. If teachers adopt a narrower definition of fluency, they could promote fluency in the classroom in a more focused way, rather than the current approach of developing speaking skills and perhaps expect fluency to develop as learners become proficient speakers of their L2.

2.3.1 Empirical evidence of L2 fluency development

This section reviews empirical research on fluency in instructional settings, focusing on task repetition, planning time, and teaching formulaic sequences. Most of these studies used cross-sectional data by focusing on the performance of learners at a single point in time.

Task repetition involves performing a speaking task more than once (Foster, 2020). Bygate (2001) argues that when L2 learners perform a task for the first time, their speech production system needs to execute all the relevant processing steps (Levelt, 1989, 1999) under time pressure. It can be argued that during the first performance of a task, learners must distribute their attentional resources among conceptualization, formulation, and monitoring. They also have to deal with performance breakdowns due to incomplete lexical or syntactic knowledge representation or slow access to these representations (Dörnyei & Kormos, 1998). Task repetition allows L2 learners to rely on previously conceptualized task content and to activate recently used linguistic structures for conveying their messages. This, in turn, may reduce the attentional demands on learners to simultaneously conceptualize, formulate, and monitor their messages (Lambert, 2017). Multiple studies have found that task repetition has a significant impact on fluency (e.g. Ahmadian, 2011; Bygate, 2001; Lynch & Maclean, 2000; 2001; Sample & Michel, 2014; Wang, 2014; Lambert et al, 2017). However, few studies have taken a detailed approach to investigate the various components of fluency that may be affected by task repetition. Lambert et al. (2017) investigated the effect of performing the same communicative task six times on several fluency measures. They observed that the number of end-clause pauses decreased during the first two performances only, while the number of mid-clause pauses gradually decreased up to the fifth performance. Self-repairs decreased in the fifth and sixth performances. The researchers concluded that a reduction in end-clause pauses suggests a lower processing load of the conceptualization process, decreased mid-clause pausing indicates a more efficient linguistic encoding, and the reduced frequency of self-repairs suggests that participants' accuracy and efficiency in linguistic encoding had improved to the extent that fewer reformulations and self-corrections were necessary.

Planning time, spending a few minutes preparing for a speaking task rather than starting it right away, has also been shown to promote a more fluent performance (Foster & Skehan, 1996; Skehan & Foster, 1999). When L2 learners are asked to speak spontaneously, they must think about what they want to say and how to say it simultaneously. This task is even more difficult for L2 learners since they do not have automatic processes of lemma selection and phonological encoding. However, more planning time can increase the attentional resources available to L2 speakers, reducing the pressure to generate content (conceptualization) and the pressure on encoding processes (formulation). Research suggests that providing learners with planning time before performing a task can improve fluency, in terms of speed of speech (Tavakoli & Skehan, 2005), mean length of run (Skehan & Foster, 2005), and reduced mid-clause pauses (Foster & Skehan, 1996).

Several researchers (e.g. Wray, 2000, 2008; Wood, 2010, 2016) have argued that explicit teaching of *formulaic sequences* (e.g. *Hoe gaat het met jou? How are you?),* also called chunks, contributes to the development of L2 fluency. Wray (2000: 465) defined formulaic sequences as "A sequence, continuous or discontinuous of words or other meaning elements, which is, or appears to be, prefabricated: that is stored and retrieved whole from the memory at the time of use, rather than being subject to generation or analysis by the language grammar''. Formulaic sequences have processing advantages since processing them is less demanding than processing a sequence of non-formulaic words. They are stored in the long-term memory as single units and in the mental lexicon as ready-to use items and can be retrieved and processed as inseparable wholes or with a minimal amount of pressure from the encoding processes (Kormos, 2006; Richards & Schmidt, 2002; Wray 2002, 2008). Thus, formulaic sequences can lead to more fluent L2 production because retrieving and articulating them requires little working memory capacity, allowing L2 learners to focus on conceptualization and formulating the rest of the message.

Although the evidence from the above-mentioned studies, mostly based on cross-sectional data, supports the hypothesis that task repetition, planning time and instruction of formulaic sequences may improve L2 fluency, the main limitation of this type of research is that it does not provide insight into the role of pedagogy in improving fluency, which is, however, of interest to teachers and learners. However, a few studies have examined whether pedagogic intervention can help enhance learner fluency over a limited period of time by teaching fluency strategies. The next section reviews the literature on this.

2.3.2 Strategy instruction of fluency strategies

In L2 speech production, two of the three speech processes, formulation and articulation, are less automatized than in L1 due to incomplete linguistic knowledge and slower access to that knowledge (e.g., Kormos, 2006; see also 2.2.2). Therefore, these processes require a lot of working memory capacity and often cannot run in parallel with other speech processes. This can lead to a loss of fluency. To increase fluency, it is necessary to enhance the linguistic knowledge of L2 speakers and make that knowledge more accessible (Welie, 2022).

When the lexical knowledge of an L2 learner is insufficient to convey the message, they can employ strategies such as circumlocutions. When L2 speakers have slow access to linguistic knowledge to express their message fluently, they can use strategies such as fillers (e.g., "well"), formulaic sequences or chunks (e.g., "let me think") to compensate for a loss of fluency. In this dissertation, *fluency strategies* are regarded as strategies that L2 learners can use to compensate for insufficient lexical knowledge (e.g., by using circumlocution) and to gain time (e.g., by using fillers and formulaic sequences) when they are unable to access their linguistic knowledge quickly enough to convey their message. The benefits of using fluency strategies are that L2 learners can avoid silent pauses, gain more time, and keep the communication channel open during difficulty (Dörnyei & Kormos, 1998). Fluency strategies like fillers and formulaic sequences need to be fully automatized so that their encoding does not require attention, and their use frees the speakers' attentional resources (Schmidt, 1992).

From a pedagogical perspective, the question now arises how teachers can promote the use of fluency strategies. More specifically, what kind of teaching and learning activities can enhance learners' control over their fluency? According to Chamot and Harris (2019), strategy instruction enables learners to gain control over their learning process and foster their development as effective and independent learners. This control over the learning process can be exercised through *metacognition* which refers to "an awareness of and reflections about one's knowledge, experiences, emotions and learning" (Haukas 2018). In the context of this dissertation, metacognition refers to awareness and reflection on the knowledge learners have about fluency and fluency strategies and how to use that knowledge to improve their fluency skills.

The impact of strategy instruction on the development of fluency has been investigated by few researchers (Seifoori & Vahidi, 2012; Tavakoli et al., 2016). Seifoori and Vahidi (2012) and Tavakoli et al (2016) combined fluency strategy training with awareness-raising activities. Awareness raising (i.e., guiding students to "notice" (Schmidt, 2001) specific aspects of language) has also been reported to contribute to the development of independent learning and enhance L2 development (Fotos, 2012; Pickering, 2001). An example of an awareness-raising activity is analyzing fluent and non-fluent L2 speech. Seifoori and Vahidi (2012) found that fluency strategy training combined with awareness raising activities made the speech of intermediate English language learners in a classroom setting more fluent. The strategies were focused on avoiding repetitions and false starts, and the participants learned to minimize the use of unfilled pauses and learned to use fillers as a strategy to maintain the fluency of their speech. The learners who received the training used less filled mid-clause and end-clause pauses and less unfilled mid-clause pauses than the untrained group. The results suggest that fluency strategies could be beneficial for developing fluency.

Tavakoli et al. (2016) conducted a study on two groups of L2 learners - a control group and an experimental group. Both groups performed a monologic task at the beginning and end of the four-week intervention. During the intervention, the experimental group received fluency training twice a week, which included awareness-raising activities (e.g., listening to the speech of a non-native speaker and identify those features of fluency that had an impact on the learners' perceptions as listeners (e.g., long pauses, hesitations, repetitions), strategies for improving fluency (e.g., using lexical fillers and avoiding repetitions and hesitations) and speaking practice. The researchers found that the experimental group demonstrated higher levels of fluency in the post-test than the control group, as measured by speech rate, articulation rate, length of run, and phonation time ratio. Thus, these few studies suggest that the combination of awareness-raising activities and strategy training might be an effective way to enhance L2 fluency. This raises the question of whether strategy instruction combined with awareness raising activities could also be beneficial for anxious L2 learners.

2.3.3 Using technology to improve L2 fluency and reduce FLA

The empirical evidence (Seifoori & Vahidi, 2012; Tavakoli et al., 2016) introduced above provides support for teacher-led strategy instruction with printed learning materials to develop L2 fluency. With the technological advancements of the last two decades, strategy instruction might also be carried out using technology-enhanced language learning (TELL). TELL involves the use of technological devices and software applications in L2 settings (Chang & Hung, 2019). TELL has some advantages: (a) learners seem to find it less stressful and easier to express themselves in the L2 in digital environments than in the face-to-face context of the classroom (Cote & Gaffney, 2018; Satar & Ozdener, 2008), because they can practice without being observed or listened to by the whole class or teacher (Gonzalez-Lloret & Ortega, 2014); (b.) learners can practice at their own pace and have equal opportunities to participate because the amount of time they spend producing a message is not limited by the presence or influence of other members of the class. This seems to promote learner autonomy (Gonzalez-Lloret & Ortega, 2014); (c.) with the help of Automatic Speech Recognition (ASR), users can receive immediate individualized speech analyses along with instructional feedback on their performance (McClain, 2016). In contrast, in-class speaking instruction often adopts a one-size-fits-all approach, focusing on common challenges.

Despite the growing interest in TELL, only a few empirical studies have examined the development of fluency using digital technology. Research has shown that online voice recording and video conferencing (Tecedor & Campos-Dintrans, 2019), MOOC-based flipped classrooms (Wang et al., 2018), and mobile games (Grimshaw & Cardoso, 2018) can have a positive impact on L2 fluency development. However, these studies do not provide insight into how fluency instruction can influence L2 fluency development. In addition, the emphasis was on peer-to-peer conversation activities and not on individual practice through self-study. Finally, although positive results have been achieved with ASR in the development of pronunciation (McCrocklin, 2016; Golonka et al., 2014), to the best of my knowledge, this has not yet been applied in fluency studies. To address these gaps, this dissertation investigates the effects of an app-based intervention drawing on the mobile app *Reppen* (see Chapter 5). This app is specifically designed and developed for the purpose of training L2 fluency (see Chapter 3).

2.4 Combining psycholinguistic and pedagogical perspectives on L2 fluency

The aim of this dissertation is to accumulate experimental evidence on the role of FLA on L2 fluency, as well as the effect of an app-based intervention on L2 fluency and FLA. In this dissertation, a psycholinguistic perspective on fluency (Chapter 4) is combined with a pedagogical perspective (Chapter 5). Chapter 4 examines the relationship between FLA and L2 fluency in relation to task complexity. Studies have found a link between FLA and L2 fluency, and it is assumed that FLA may negatively affect L2 speech processing by interfering with cognitive processes (particularly formulation/encoding) that require attention control. The relationship between FLA and L2 fluency in low-intermediate L2 learners has not been investigated in relation to cognitively demanding tasks. Studying the effect of task complexity is relevant for teachers who use these tasks in classroom practice and for researchers to better understand the role of FLA on L2 fluency. This study addresses the following research question:

RQ 1: To what extent is there a relationship between speed, breakdown and repair fluency measures and FLA and proficiency, and how does this interact with task complexity?

Based on previous research (e.g., MacIntyre & Gardner, 1994; Pérez Castillejo, 2019, 2021; Sanaei et al., 2015; Bielak, 2022), we hypothesize that a cognitively demanding task negatively affects speed fluency of anxious learners. Regarding breakdown fluency, we expect that complex tasks elicit more pausing behavior. Considering repair fluency, we hypothesize that a cognitively demanding task will lead to fewer self-corrections and reformulations by anxious learners. The idea underlying this hypothesis is that at lower proficiency levels, when L2 knowledge is not automatized, learners' attentional capacity can be too taxed by other processes, particularly formulation, so that there is no room left to make corrections.

In the next step, the results of this study (Chapter 4), together with previous research on the relationship between L2 fluency and FLA, as well as findings from studies on L2 fluency training that did not involve the use of an app (Tavakoli et al., 2016; Seifoori & Vahidi, 2012), will be used to develop an app that teaches fluency strategies. This app-based training is hypothesized to improve fluency by equipping learners with strategies to compensate for potential losses in fluency during speaking tasks. In addition, the app provides feedback on fluency through ASR. This app, called *Reppen*, and its development are described in Chapter 3. Chapter 5, adopting the pedagogical perspective, investigates the effect of an app-based intervention on L2 fluency and FLA. Embedded in the use of the *Reppen* app, the study in Chapter 5 addresses three research questions:

RQ 2: To what extent does an app-based intervention in general affect L2 fluency?
RQ 3: To what extent does specific app-based training of fluency strategies affect L2 fluency?
RQ 4: To what extent does an app-based intervention in general affect FLA?

Based on the literature, we expect an app-based intervention in general improves fluency (Swain, 1985). Within this intervention, we expect fluency training to improve fluency the most (Tavakoli et al., 2016; Seifoori & Vahidi, 2012). Finally, we expect that individual app-based practice will lead to a reduction in FLA (Côte & Gaffney, 2018).

The results of this study can support teachers and learners through a research-based app that can be easily implemented in their educational environments. More specifically, the findings might be valuable to teachers and learners looking for tools to improve fluency and support anxious learners, where the *Reppen* app might serve as a supplement to formal L2 instruction and as a medium for independent L2 study in online and face-to-face settings.

Finally, the general discussion in Chapter 6 discusses the main findings of the studies, methodological challenges and choices in app-based research, and suggests theoretical and pedagogical implications as well as directions for future research.

3. The Reppen app

This chapter introduces the *Reppen* app (see Figure 1), specially developed for the purpose of training L2 fluency. The app, developed and designed by me in collaboration with computer programmers from MSML (https://www.msml.nl/), a software company in the Netherlands, is intended for students learning Dutch as L2 with at least a basic proficiency level of B1 according to the Common European Framework of Reference (CEFR; Council of Europe, 2011). We focus on learners at B1 level, as fluency training may be less effective for learners with lower proficiency levels, whose production processes tend to be slower and more frequently interrupted due to limited lexical and/or grammatical knowledge or slow access to that knowledge (Tavakoli et al., 2016; Kormos, 2006).

Figure 1 Frontend of the Reppen app

Fluency training with an app can have several benefits. Firstly, research into L2 speaking indicates that learners express themselves more easily in digital environments than in the face-to-face context of the classroom (Côte & Gaffney, 2018; Satar & Özdener, 2008), because they can practice without being observed or listened to by the whole class or teacher (Gonzalez-Lloret & Ortega, 2014). Secondly, learners experience more autonomy, take on a more active role, and are more motivated to perform tasks (Gonzalez-Lloret & Ortega, 2014). Another benefit is that Automatic Speech Recognition (ASR) can be integrated relatively easily in digital environments and can offer immediate and individualized feedback (McClain, 2016). In contrast, in-class speaking instruction often adopts a one-size-fits-all approach, focusing on common challenges.

The *Reppen* app consists of two instructions: fluency training and task-based speaking lessons. The first instruction includes explicit fluency training, where students learn fluency strategies and practice these specific strategies in speaking tasks. Through these strategies, learners are better able to compensate for potential breakdowns in fluency. The second instruction, task-based speaking lessons, does not explicitly focus on fluency. Here, participants learn language functions and practice them in speaking tasks. Both instructions were chosen to investigate the effects of fluency training and task-based speaking lessons on L2 fluency (see Chapter 5). Based on previous findings, it is expected that both types of instruction can improve learners' fluency (Swain, 1985; Tavakoli et al., 2016; Seifoori & Vahidi, 2012), but explicit fluency training is expected to lead to greater improvement because, by using fluency strategies, participants are better equipped to compensate for any loss in fluency.

Below, the goals and instructional models of both parts of the app are explained, lessons and features of the app are described, as well as participants' and teachers' experiences with the app.

3.1 Fluency training and Task-based speaking lessons

This section explains the rationale behind fluency training and task-based speaking lessons, their objectives, and the instructional models used.

Fluency training

The aim of the fluency training is to raise awareness of different aspects of fluency, teach L2 fluency strategies, and apply these strategies in speaking tasks. Research shows that strategy instruction of fluency strategies can improve fluency development (Rossiter, 2003; Seifoori & Vahidi, 2012; Tavakoli et al., 2016) (see Chapter 2.3.2).

This dissertation considers *fluency strategies* as techniques to compensate for insufficient lexical knowledge and to gain time when there is no quick access to the linguistic knowledge needed to convey a message. In the fluency training, attention is given to the following *fluency strategies*: lexical fillers, formulaic sequences, and circumlocution. The use of lexical fillers (e.g., *so*, *well*, and *you know*) (Guillot, 1999; Nattinger & DeCarrico, 1992; Tyler, 1992; Tavakoli et al., 2016) contributes to the naturalness of speech and reduces time spent in silence while processing (Peltonen, 2017). The use of formulaic sequences (e.g., "*how do you say that again*"; "*let me think*") can be beneficial (Boers et al., 2006; Wood, 2010, 2016; Wray, 2000, 2002) because they are stored in the mental lexicon as ready-to use items, and can be retrieved and processed as inseparable wholes or with minimal amount of pressure from the encoding processes (Kormos, 2006; Richards & Schmidt, 2002; Wray 2002, 2008), allowing L2 learners to focus on conceptualization and formulating the rest of the message. Furthermore, the use of circumlocution can aid in overcoming lexis-related problems (Dornyei & Kormos, 1998) and allow the learner to gain time and compensate for missing vocabulary items.

The app-based fluency training is guided by the Cognitive Academic Language Learning Approach (CALLA) model of strategy instruction (Chamot, 2009). The CALLA model is learner-centered and guides students to become independent learners who can evaluate and reflect on their own learning (Rubin et al., 2007). This instructional model for L2 learners has been implemented because of its sequential phases - preparation, presentation, practice, evaluation and expansion - in which students can learn and practice strategies within contextualized activities. The CALLA model integrates content (here: the concept of fluency and fluency strategies) with language activities (here: meaning-focused speaking tasks) and explicit instruction in strategies (here: fluency strategies). These three components are taught in an integrated manner. In other words, content, language, and strategies are taught in five instructional stages: preparation, presentation, practice, evaluation and expansion. The strategy instruction is initially very explicit, and then the support from the instructor (in this case, the app) gradually decreases, so that students begin to take more responsibility for their own learning.

The first stage, preparation, is to raise learners' metacognitive awareness of fluency strategies they are already using, in Dutch or their native languages. This awareness is heightened in the presentation stage when the teacher (or in our case, the app) presents new strategies and models how to use strategies on speaking tasks. During the practice stage, students are given ample opportunities to use the new strategies in 16 authentic meaningful speaking tasks and receive feedback on fluency through the app. In the fourth stage, students reflect on what they have achieved and on their learning process, thereby increasing their metacognitive awareness. Activities at this stage are for example checklists and open-ended questionnaires. The last stage of CALLA model is expansion. The main purpose of this stage is to encourage students to apply what they have learned to new tasks, fostering their development as autonomous and self-regulated learners (Chamot et al., 1999). Section 3.1.1 provides a more detailed description of the lesson content during the five phases.

Task Based speaking lessons

The second instruction of the *Reppen* app consists of task-based speaking lessons where the goal is to practice speaking skills, but unlike fluency training, there is no focus on fluency. This training uses the design of task-based instruction (TBI), also known as task-based language teaching (TBLT). We chose TBI rather than the communicative approach for several reasons. While the communicative approach, like TBI, emphasizes interaction and meaning, TBI provides a framework for learning activities, where the focus is on completing real tasks (Ellis, 2003; Willis & Willis, 2007).

In the communicative approach, learners are typically guided step by step towards spoken production. Lessons often begin with the explanation and practice of new vocabulary and grammar, and grammatical structures are frequently treated explicitly and in considerable detail. Such an extensive focus on form may hinder fluency during subsequent speaking tasks, which we aim to avoid. Moreover, while in the communicative approach, lessons usually progress from highly structured speaking activities to more open-ended speaking tasks. In contrast, TBI moves more quickly towards the actual performance of the speaking task. This quicker transition is advantageous because it enables learners to become aware of gaps in their linguistic knowledge and skills at an earlier stage (Welie, 2022).

TBI is a teaching method that focuses on the use of authentic language to complete meaningful tasks in the target language (Ellis, 2003; Willis, 1996). This means that in TBI, tasks are oriented towards meaning, communicative in nature, and emphasize the content of the message. Additionally, the tasks are goal-oriented and authentic, incorporating real, contextualized language with applications outside of the activity itself (Ellis, 2003; Willis, 1996; Willis & Willis, 2007).

In the app-based task-based speaking lessons, participants receive instructions about different language functions such as '*beschrijven van activiteiten'* (describing activities*), 'beschrijven van grafieken' (*describing graphs*), 'mening geven'* (giving an opinion), '*iemand advies geven'* (giving advice), and '*bespreken van voor- en nadelen'* (discussing advantages and disadvantages).

Learners are considered to have succeeded on a task if they accomplish the task and achieve something with the language, rather than by mastering a particular linguistic piece.

This concept of doing something with the language, as opposed to merely knowing something about the language, is an essential principle of TBI (Gonzalez Lloret, 2017), suggesting that a language can be learned by actively engaging learners in its practical use.

In the instructional models, (Ellis, 2003; Willis, 1996) for creating a task-based lesson, three phases are used, which are also applied in the *Reppen* app: pre-task, during task and post-task. The pre-task stage prepares learners for task performance by introducing the language function and activating necessary vocabulary, grammatical constructs, or phrases. The second stage, during task, learners perform the main speaking tasks. The speaking tasks were developed based on Hulstijn's (2012) study, which assessed the suitability of tasks for CEFR levels B1 and B2 according to criteria such as formality, discourse type (descriptive and persuasive), and topic complexity (simple and complex) (see Table 1). The third stage, post-task, involves reflecting on the task performance and students receive feedback on an aspect not directly related to fluency, namely lexical diversity.

In summary, the instructional models CALLA and TBI are different in their goals, but their frameworks are comparable. The CALLA model is a more comprehensive version of the TBI framework. Its five phases correspond with the *pre-task, during task*, and *post-task phases* of TBI. The *preparation* and *presentation* phases in the CALLA model and the *pre-task* in TBI introduce the learning goals (strategies and language functions) and the new content (phrases, vocabulary or strategies). The *practice* phase in CALLA and *during task* in TBI involve students' practice of the target language function and strategy through meaning-focused speaking tasks. The *evaluation* phase in CALLA and *post-task* in TBI involve an assessment of students' learning progress in terms of fluency and lexical diversity, respectively. Finally, in CALLA model, *expansion* encourages learners to apply their new knowledge into other contexts.

3.1.1 Fluency training and task-based speaking lessons

The following section provides a detailed description of the lesson content of the fluency training and task-based speaking lessons. Table 1 provides an overview of the lesson plans during the intervention (see Chapter 5).

Table 1

Overview of the Lessons During the Intervention

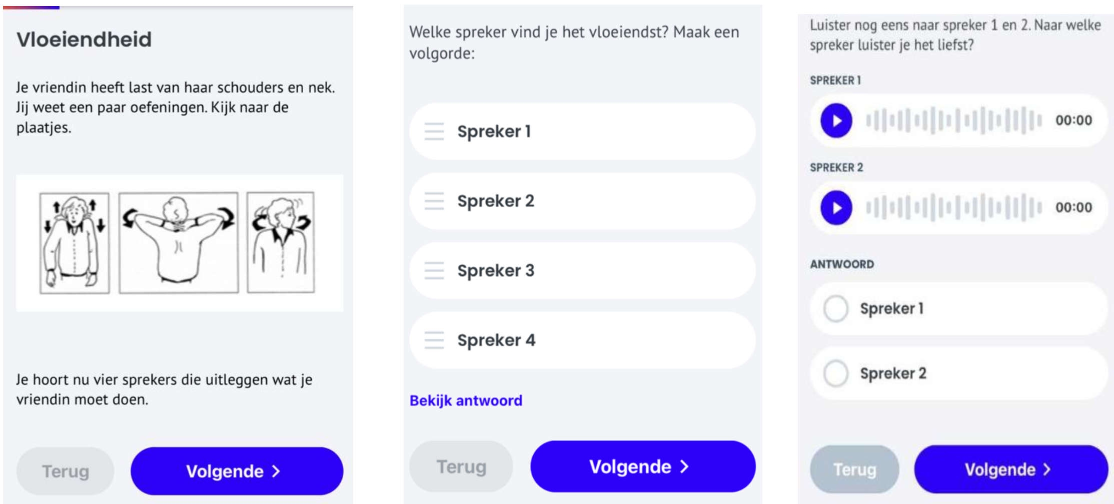
Weeks	Lessons	Descriptions
	Fluency training	
1	Preparation	Awareness-raising of the importance of fluency and identify those features of fluency that have an impact on the listener (e.g., long pauses, hesitations, repetitions).
2	Exploring fluency strategies	Demonstrating why fluency strategies are useful and when they can be used.
2-4	Practice fluency strategies	Participants practice fluency strategies in speaking tasks
4	Evaluation	Participants evaluate the effectiveness of strategies used and reflect on overall performance
4	Expansion	Encouraging use of strategies
	Task-based speaking lessons	
5	Describing activities (simple and descriptive)	For each task-based lesson*:* *Pre-task*: Present the situation, language function and the necessary vocabulary and structures. *During-task*: Participants perform the speaking task *Post-task*: the app provides feedback
6	Describing graphs (complex and descriptive)	
7	To give an opinion (simple and convincing)	
8	To give advice Discussing advantages and disadvantages (complex and convincing)	

Below follows a more detailed description of the intervention lessons, starting with the fluency training and then the task-based speaking lessons.

a. *Preparation*. The purpose is to raise learners' awareness of the importance of fluency and identify those features of fluency that have an impact on the listener (e.g. long pauses, hesitations, repetitions). The fluency training starts with an activity called *'Welke spreker is het vloeiendst?'* (*Which speaker is the most fluent?)* in which participants listen to both fluent and dysfluent L2 speakers of Dutch who describe pictures. Participants are then asked to rank the speakers in terms of fluency. Then participants study two transcripts of the recordings and identify where the speakers are less fluent. Screenshots of these exercises are shown in Figure 2. Participants are also instructed to think about the impact of dysfluent speech on the listener and what can be done to avoid dysfluency.

Figure 2

Screenshots illustrating the Exercise 'Which Speaker is the Most Fluent?'



An exercise in which four speakers explain three shoulder exercises (left). Which speaker is the most fluent (middle) and to which speaker do you prefer to listen (right). (Full English translation see Appendix A).

b. *Exploring the target strategies*. The aim is to raise awareness of fluency strategies and explore these strategies. The training starts with an introduction to fluency strategies focusing on why these strategies are useful and when they can be used. Next, participants make a recording in their L1, transcribe it, and are asked which strategies they use in their native language to gain time (e.g., long silent pauses; filled pauses: *uhm, uh*; expressions like *Ich denke …, Was ich meine, ist…*).

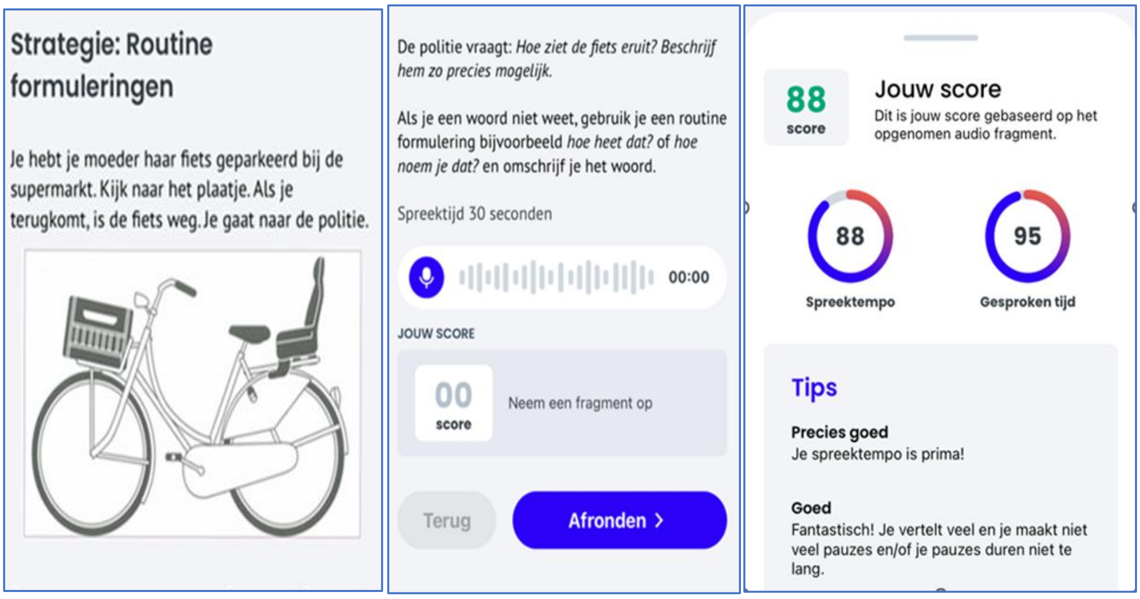
Afterwards, the participants listen to a recording in Dutch, and analyze the native speakers' use of strategies for buying time, without sounding dysfluent. The aim is to raise participants' awareness of 'markers of fluency', such as filled pauses, fillers and formulaic sequences to indicate that the speaker still holds the floor. Then the strategies, circumlocution (*omschrijven van het onbekende woord* - describing an unknown word), formulaic sequences like *'hoe zeg je dat ook alweer?*' (how do you say that again?), and lexical fillers like *'nou ja'* (well), '*kijk*' (look), *'weet je'* (you know), are introduced.

This is followed by a so-called 'tracking or shadowing exercise' with the purpose of raising awareness of fluency strategies and pausing and to improve automatization. In this exercise, participants listen to a short audio recording, identify pauses, and fluency strategies such as lexical fillers and formulaic sequences in the transcript. Then, they listen carefully to the recording two or three times, following the transcript and paying attention to pauses and fluency strategies throughout. Finally, they read it aloud several times with or slightly after the speaker, paying particular attention to the target features and make a recording; the app provides feedback on speech rate.

c. *Practice.* Students practice the target strategies in meaning-focused speaking tasks. There are a total of 16 speaking tasks (see Figure 3 for an example). The app provides feedback on speech rate, phonation time ratio, and recommends using the target strategies if necessary (for example, formulaic sequences like "how do you say that again?" or lexical fillers like "well", "look", "you know") to increase fluency.

Figure 3

Screenshots of the Reppen App



Speaking task in which a bicycle needs to be described, practicing fluency strategies of formulaic sequences and circumlocution (left and middle). The recording must last at least 30 seconds before it can be analyzed (middle). Scores for speech rate and phonation time with feedback tips (right). (Full English translation is available in Appendix B).

d. *Evaluation and expansion.* The learners evaluate the effectiveness of the strategies and reflect on the overall performance. Post-task evaluation helps learners understand which strategies were or were not helpful for them, for example, in situations when a word doesn't come to mind. This knowledge can further assist them in making informed decisions when choosing strategies for future speaking tasks. The main purpose of expansion is to encourage participants to try using the strategies in other speaking activities.

Design features of the fluency training are speech recording, playback, and automatic fluency analyses (i.e., speech rate, total phonation time and pause time). (Figure 4 illustrates some of the app's features: speech-recording, speech analysis, feedback). Speech rate and phonation time ratio were selected based on previous research. Studies on L2 fluency training without an app showed that experimental groups typically achieved higher speech rates and phonation time ratios after training (Tavakoli et al., 2016), and research examining the relationship between FLA and fluency found a negative association between FLA and phonation time ratio (e.g., Pérez Castillejo, 2019). Pérez Castillejo (2019, 2021) also reported a relationship between FLA and the number of mid-clause pauses. The software used, Azure Cognitive Services (for more information, see: https://azure.microsoft.com/en-us/services/cognitive-services/speech-services/#overview), allowed speech rate and phonation time ratio to be analyzed. However, it did not permit analysis of pause location, meaning that mid-clause pauses could not be examined.

In the analyses conducted by the app, speech rate is calculated as number of words divided by the duration of the speech recording and multiplied by 60 (Huang & Graf 2020). Based on the score, the participant receives feedback. The feedback is formulated at three levels (insufficient, moderate, good) that are related to the levels A2, B1, and B2 of the CEFR. The scores and corresponding levels are based on the study of Huang and Graf (2021) and Götz (2013). Regarding phonation time (i.e., the proportion of time spent speaking compared to the total amount of time spent on the task), we based the scores on the study of Tavakoli et al. (2020). In addition to the scores on speech rate and phonation time ratio, students receive a motivational comment like *Fantastisch!* (Fantastic!), *Goed op weg*! (Doing well!), *Probeer het!* (Try it!) and, if necessary, advice how to use a target strategy.

In the second part of the app, task-based speaking lessons, participants receive instructions about different language functions such as '*beschrijven van activiteiten'* (describing activities*), 'beschrijven van grafieken' (describing graphs), 'mening geven'* (giving an opinion), '*iemand advies geven'* (giving advice), and '*bespreken van voor- en nadelen*' (discussing advantages and disadvantages). During the first step, the pre-task, learners are prepared for task performance. The app presents the language function and situation along with the necessary vocabulary and structures. During the during-task phase, participants perform speaking tasks, and in the third step, post-task, the app provides feedback on lexical diversity. Like fluency training, the participant completes a total of 16 speaking tasks. Figure 4 illustrates an example of a pre-task, a speaking task (during-task) and post-task for describing a diagram.

Figure 4

Example of a pre-task, during task and post-task

Pre-task

Figuren beschrijven

Als je een grafiek of diagram beschrijft kun je de volgende woorden en zinnetjes gebruiken:

Het diagram/de grafiek laat zien ...
Het diagram/de grafiek toont dat ...

Meer worden		Minder worden	
toenemen (is toegenomen)	*de toename*	*afnemen (is afgenomen)*	*de afname*
stijgen (is gestegen)	*de stijging*	*dalen (is gedaald)*	*de daling*

Niet meer worden of minder worden
gelijk blijven/constant blijven

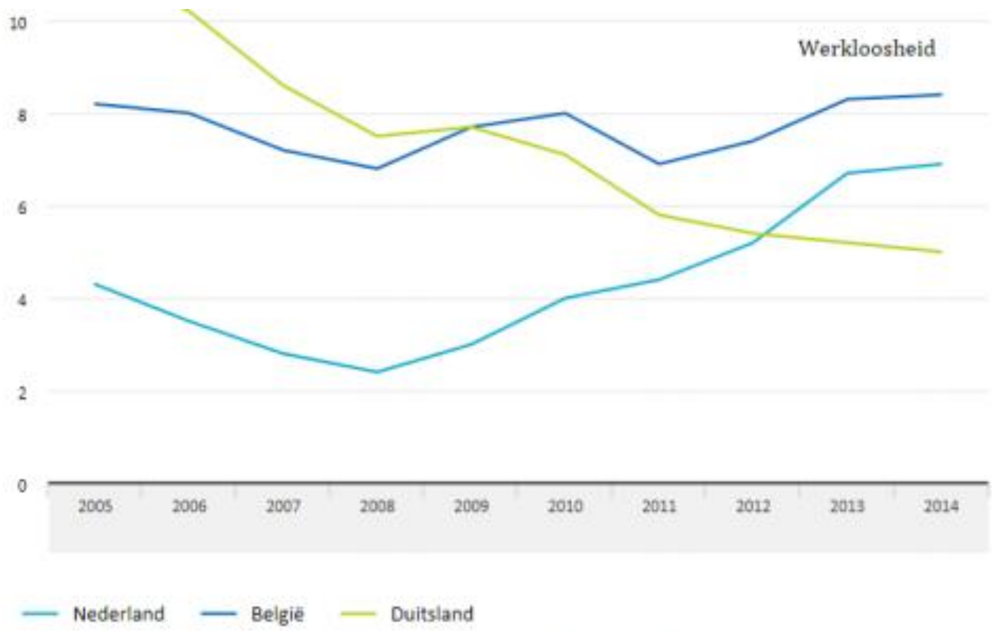
(Translated into English:
Describing Figures

When describing a graph or diagram, you can use the following words and phrases:

The diagram/chart shows ...
The diagram/chart indicates that ...

Increasing		Decreasing	
to increase (has increased)	*the increase*	*to decrease (has decreased)*	*the decrease*
to rise (has risen)	*the rise*	*to fall (has fallen)*	*the fall*

Not increasing or decreasing
to remain the same/constant



Grafiek 6: Werkeloosheid in Nederland, Belgie en Duitsland sinds 2005.

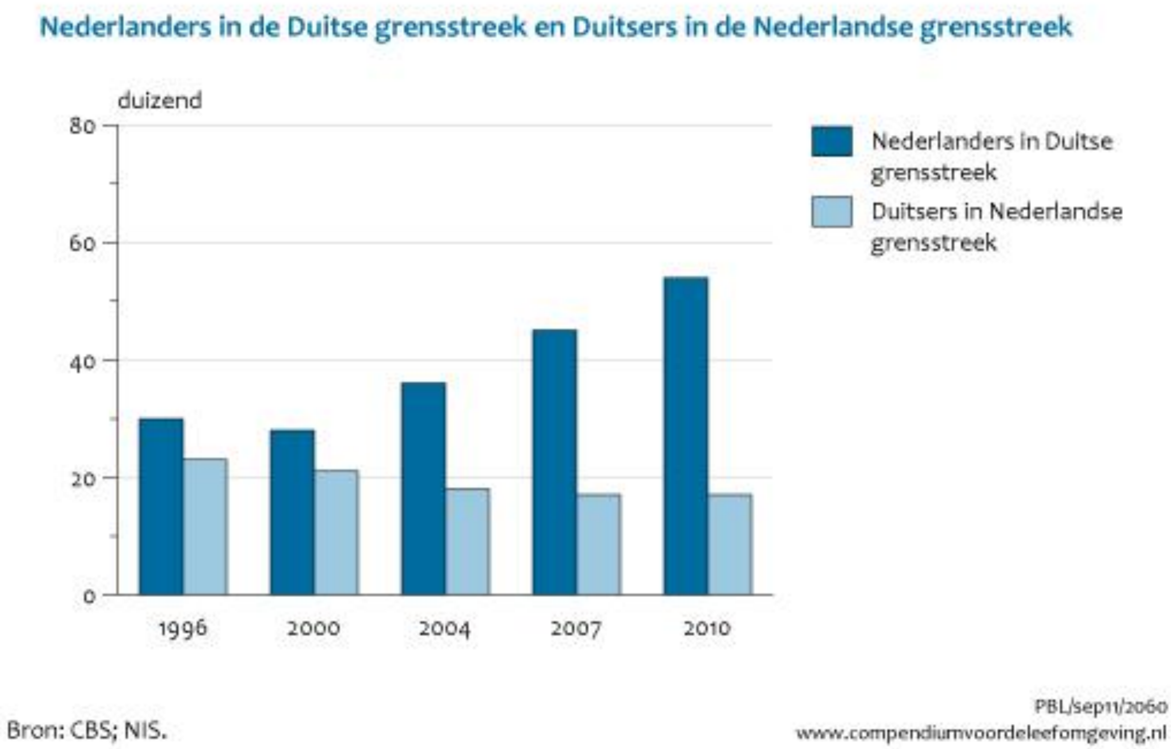
Voorbeeld: Beschrijving grafiek

Deze grafiek toont de werkloosheid in Nederland, België en Frankrijk vanaf 2005 tot 2014. In de grafiek wordt de werkloosheid in Nederland, België en Duitsland met elkaar vergeleken.
In de grafiek is te zien dat in Nederland de werkloosheid van 2005 tot 2008 daalde en na 2008 weer toenam. Maar in Duitsland daalde de werkloosheid juist fors in de periode van 2006 tot 2014. Terwijl in dezelfde periode de werkloosheid in België min of meer gelijk is gebleven.

(Example: Description of Graph

This graph depicts the unemployment rates in the Netherlands, Belgium, and France from 2005 to 2014. The graph compares the unemployment rates among the Netherlands, Belgium, and Germany. In the graph, it can be observed that the unemployment rate in the Netherlands decreased from 2005 to 2008 and then increased after 2008. However, in Germany, the unemployment rate declined significantly from 2006 to 2014. Meanwhile, during the same period, the unemployment rate in Belgium remained more or less constant).

1. Kijk naar het diagram.
(Look at the diagram)



Vul de ontbrekende woorden in.

Het diagram.......(a)........... het aantal Nederlanders dat in de Duitse grenstreek woont en het aantal Duitsers dat in de Nederlandse grensstreek woont tussen 1996 en 2010. Uit het diagram blijkt dat in deze periode het aantal Nederlanders dat in de Duitse grensstreek woont sterk is ....(b).....................,terwijl het aantal Duitsers dat in de Nederlandse grensstreek woont licht is ....(c)..................... Welke verklaringen zijn er voor deze ....(d)..........van het aantal Nederlanders in de Duitse grensstreek?

(juiste oplossing:
a. toont/ laat zien/gaat over
b. gestegen/toegenomen
c. gedaald/afgenomen
d. toename/stijging)

(Fill in the missing words.
The diagram........(a)........... the number of Dutch people living in the German border region and the number of Germans living in the Dutch border region between 1996 and 2010. From the diagram, it is evident that during this period, the number of Dutch people living in the German border region has ……….(b)…….significantly, while the number of Germans living in the Dutch border region has slightly ……(c)……….. What explanations are there for this ....(d)........of the number of Dutch people in the German border region?)

(Correct solution:
a. shows/illustrates
b. risen/increased
c. decreased/dropped
d. increase/rise)

Als je een grafiek of diagram bespreekt, kun je de volgende indeling gebruiken:

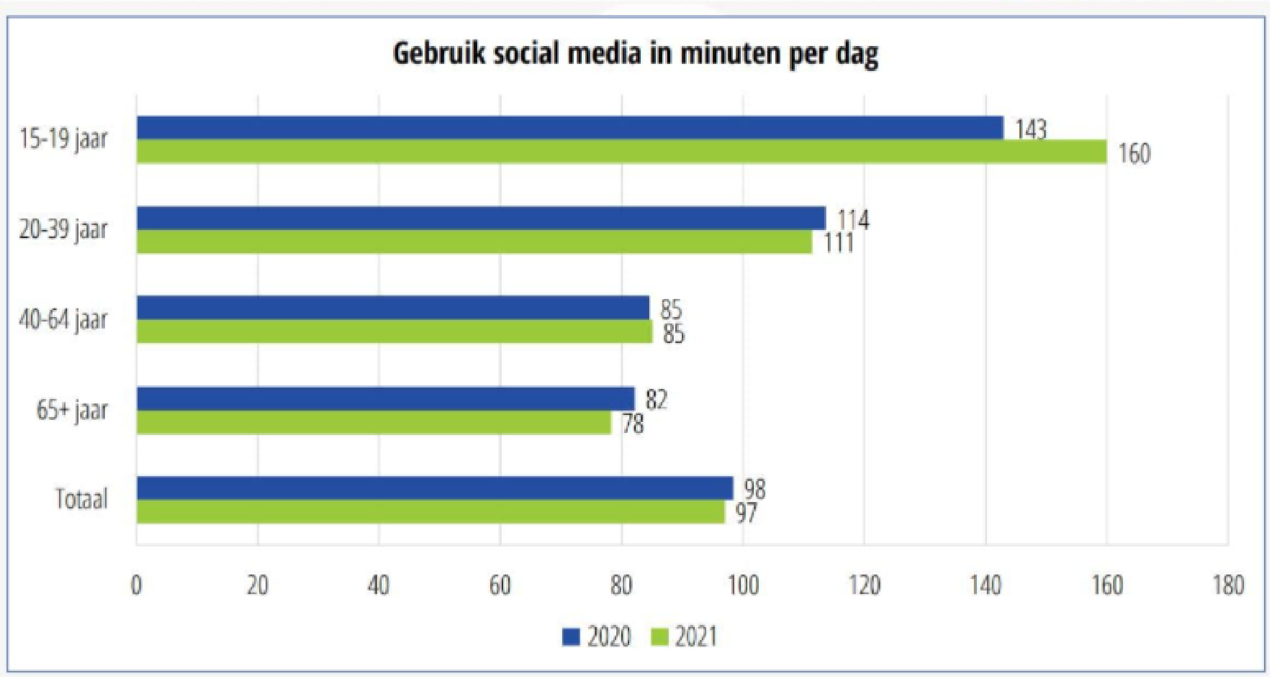
*Inleiding: Waar gaat de grafiek/het diagram over?
*Midden: Wat valt op?
*Slot: Wat kun je concluderen?

(When discussing a graph or diagram, you can use the following structure:

*Introduction: What does the graph/diagram depict?
*Middle: What stands out?
*Conclusion: What can be concluded?)

During- task

2. Je presenteert in de klas het volgende diagram.
(*You present the following diagram in class*)



https://www.frankwatching.com/archive/2021/01/26/social-media-onderzoek-2021/

Bereid je voor door eerst trefwoorden op te schrijven. Gebruik de volgende indeling. (*Prepare by first writing down keywords. Use the following format*)

Inleiding: .... (*Introduction*) Kern: .... (*Main*) Slot: .... (*Conclusion*)

Voorbereidingstijd: 4 minuten (*Preparation time: 4 minutes)*
Spreektijd: 1 minuut (*Speaking time: 1 minute)*

Post-task

Feedback		Advies
Variatie in woordkeuze	klein	Probeer woorden te gebruiken als *toenemen, stijgen, afgenomen, dalen, gelijk blijven* (zie uitleg *Figuren beschrijven*)
	voldoende	Het gaat al goed, maar varieer nog meer in in woorden die het diagram beschrijven (zie uitleg *Figuren beschrijven*).
	groot	Fantastisch, een uitstekende variatie!

(Feedback		Advice
Variation in word choice	small	Try to use words like *increase, decrease, decline, remain constant* (see explanation describing figures).
	sufficient	You're doing well but try to vary even more in words describing the diagram (see explanation describing figures).
	large	Fantastic, an excellent variety!)

Regarding feedback in task-based speaking lessons. The software we used in the app, Azure Cognitive Services, enables speech-to-text conversion. Within the capabilities of the software, we were limited to providing feedback on either pronunciation or lexical diversity. I am aware that feedback on both lexical diversity and pronunciation can indirectly improve fluency. We chose to give feedback on lexical diversity because focusing on the pronunciation of sounds, rhythm, and intonation might make pronunciation smoother and might affect fluency more than lexical diversity.

The software analyzes lexical diversity by counting the frequency of each word in a speech sample, excluding the most frequent words in Dutch (Tiberius & Schoonheim, 2014), such as *de, het, een* from the analysis. The more variety in word choice the higher the score. The features of the app in this part include speech recording, playback, automatic analyses of lexical diversity, feedback, and advice on how to improve lexical diversity.

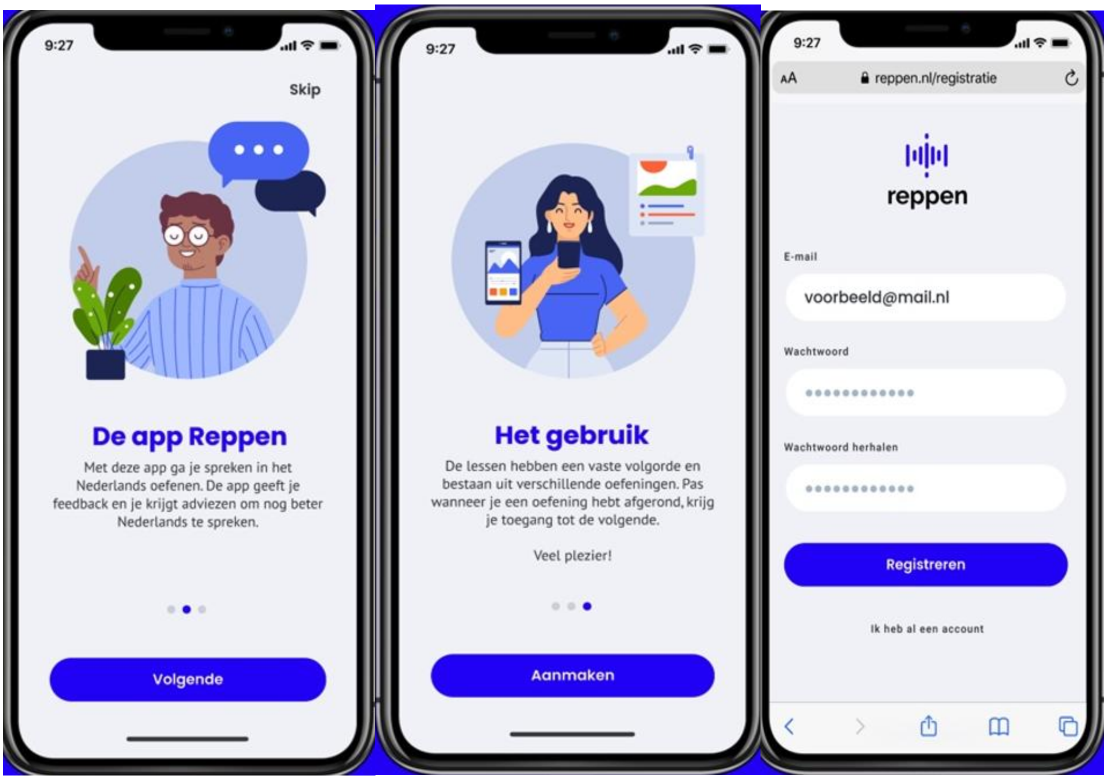
Before the intervention started, the fluency training and task-based speaking lessons were presented to two teachers of Dutch (as a foreign language) who were asked to review whether the instructions, explanation of the strategies, and the speaking tasks were clear and adequate for the participants´ Dutch level. They confirmed the instructions and explanations of the strategies were clear and the speaking tasks were appropriate. Furthermore, the app was piloted by eight German students of Dutch, who were not involved in the study, to test functionality and content of the app, after which minor adjustments were made regarding task instructions and speech analyses.

3.2 Implementation of the Reppen app

The *Reppen* app can be installed on mobile devices running iOS, Android, and Windows operating systems. To start the app, participants need to enter their username and are randomly assigned to one of the two treatment groups (see Chapter 5). The first time participants worked with the app, I provided a brief instruction on how to log in (see Figure 5), explained the app's structure (two parts, each lasting four weeks), and highlighted the app's features. It was emphasized that it not possible to skip instructions or exercises. Hence, a fixed sequence of instructions and speaking tasks must be completed. In addition, participants were instructed to work individually with the app during their regular Dutch classes, dedicating 15 to 20 minutes to it twice a week, under the supervision of their regular teacher.

Figure 5

Introduction Screens of the Reppen App



The first screens of the app where information about the app's use is displayed (left and center) and the screen where participants can log in (right).

3.3 Participants' and teachers' experiences with the Reppen app

After the intervention, 53 of the 60 participants in the treatment groups completed a short debriefing questionnaire about their experiences using the *Reppen* app comprised of three parts: (1) two checkbox questions targeting participants' feelings in working through the app; (2) five statements about the perceived effectiveness of the fluency training that participants rated on a scale from 1 to 5; and (3) three open-ended questions, which focused on experiences working with the *Reppen* app and suggestions for modifications (see Appendix C).

Regarding part 1, participants' feelings in working through the app. The responses revealed that most of the learners rated the app training as interesting (45 of 53), helpful (29 of 53), and 27 of the 53 learners indicated that they found the app training informative (see Figure 6).

Figure 6

Responses to the Question on How Working with the App Was Perceived

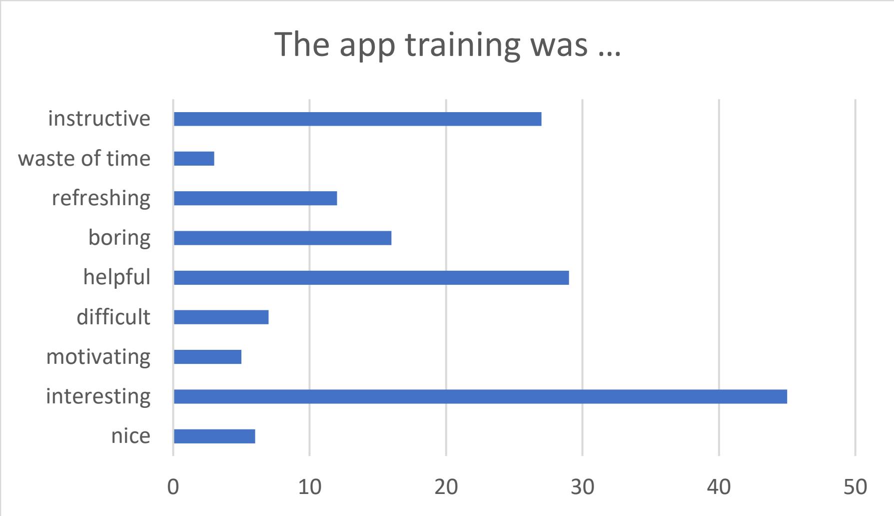


Figure 7 shows that most of the participants (43 of 53) reported that using the app increased their awareness of how disfluency impacts the listener. In addition, participants (25 of 53) reported that they had gained more knowledge about the components of fluency; part of the users felt more confident speaking Dutch (15 of 53), experienced less anxiety (18 of 53), and spoke Dutch more fluently (16 of 53).

Figure 7

Responses to the Question of What Participants Learned

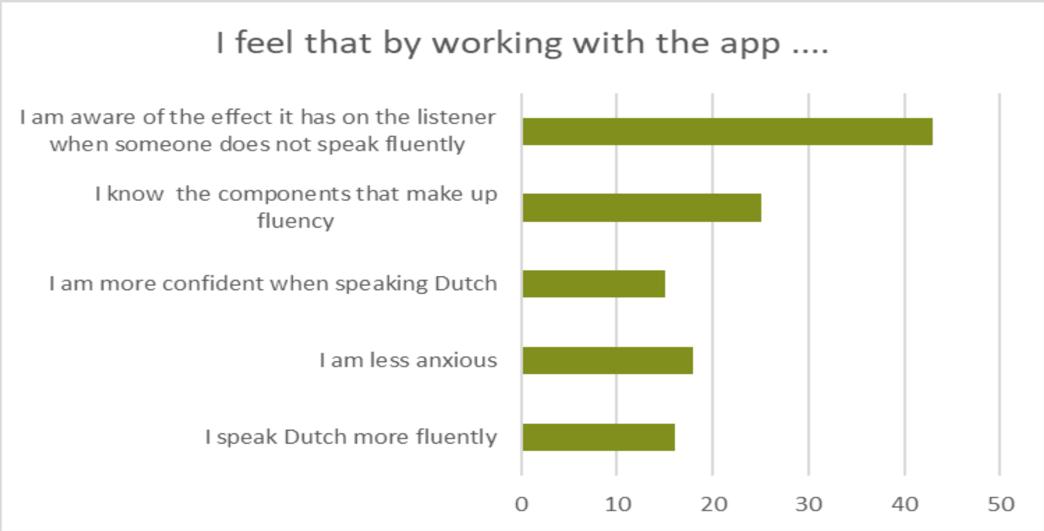


Table 2 summarizes the mean scores for the five statements targeting feedback on the perceived usefulness of the fluency training that made up the second part of the questionnaire. The results showed that all three strategies were rated equally highly. The participants were able to apply the tips and strategies they learned in the speaking tasks. Using the feedback in the speaking exercises was rated the lowest.

Table 2

Mean Scores on Debriefing Questionnaire Responses

Question No.	**Statement to be rated by participants**	**Mean response**	**Range**
1.	I was able to use the feedback provided by the app in the subsequent speaking exercises	2.74	1–5
2.	I was able to use the tips provided by the app in the subsequent speaking exercises	2.23	1–4
3.	I was able to use the *“describe the unknown word”* strategy	2.15	1–4
4.	I was able to use the *routine formulations* strategy *(e.g., “How do you say…?”)*	2.23	2–4
5.	I was able to use the *lexical fillers* strategy *(e.g., well, you know).*	2.00	1–4

Note: Statements were rated on scale from 1 (*strongly agree*) to 5 (*strongly disagree*).

The third part of the questionnaire consisted of three open-ended questions. In response to the question, “What did you like about the app training?”, participants mentioned that they found it useful to train fluency because they had not done it before. They also noted that the app allowed them to speak a lot of Dutch in a row, something that does not often happen in regular lessons. Many found the direct feedback (i.e., scores and tips) after the speaking exercises and the option to listen to the recordings to be beneficial. In response to the question, “What did you not like about the app training?”, participants wrote that the speech analyses could be improved. Some students mentioned that they occasionally had the impression that the scores did not match the recording. Responses to the question, “Is there anything you would change about the app?”, varied widely. Participants gave suggestions for technological improvements and more variety in the speaking tasks and mentioned that tips and feedback could be more comprehensive.

Summarized, the participants were generally positive about the fluency training with the app; they had gained more knowledge about fluency and felt they could use the strategies in speaking tasks. However, participants indicated that they got the impression that the speech analyses were not always completely precise. Additionally, a part of the users felt more confident speaking Dutch and experienced less anxiety.

The teachers who participated in the app-based intervention were also asked to answer some questions in writing about their impressions regarding the use of the *Reppen* app (see Appendix D). Two of the three teachers responded.

To the question, “What are your experiences using the *Reppen* app in your Dutch lessons?”, teachers indicated that they really enjoyed using the app and that students were always motivated to engage with it. However, sometimes there were technical problems that prevented students from going to the next exercise or from saving exercises. Both teachers wrote that they felt that speaking became somewhat more fluent after the fluency training.

In response to the question, “Do you think anxious students benefit from practicing speaking skills via the app?”, they replied that they indeed thought anxious students benefit from working with the app because it allows them to practice alone without fellow students listening. But, as one teacher pointed out, they still need to dare to become active during lessons or outside school in the Netherlands/with Dutch conversation partners. The teachers had no suggestions for improving the app and indicated that they would like to use *Reppen* in the future, because it provides an opportunity to practice speaking in a different way and to build self-confidence. These responses seem to confirm what has been emphasized in the literature regarding the benefits of Technology-Enhanced Language Learning (TELL), namely that in a digital environment, it is less stressful for anxious learners to speak in the L2 (Cote & Gaffney, 2018; Satar & Özdener, 2008), learners can practice at their own pace without others listening in and are (more) motivated to perform tasks (Gonzalez-Lloret & Ortega, 2014).

4. The role of foreign language anxiety and task complexity on fluency[3]

Abstract

This study investigated which aspects of fluency are related to foreign language anxiety and proficiency, and how this interacts with task complexity during a non-exam situation. Sixty-one low-intermediate German learners of Dutch completed a foreign language anxiety questionnaire, a proficiency test and two speech production tasks. Correlational analyses showed that anxiety was negatively related to number of mid-clause pauses in a complex task. Proficiency was positively related to numerous speed and breakdown fluency measures in a simple task. Mixed-effects models demonstrated that proficiency predicted two fluency measures. Task type positively predicted speed fluency. Anxiety was not a significant predictor of any fluency measure, which may be related to participants' relatively low anxiety level. This finding suggests that anxiety may not have a strong influence during speaking tasks that are not part of formal assessments.

4.1 Introduction

Highly proficient speakers produce language at a fast rate and with little effort. In contrast, learners of a second or foreign language (L2) are less fluent and produce speech at a slower speech rate with more pauses and repairs. Fluency refers to the ease or automaticity with which a learner produces speech and is manifested in flow, continuity, and smoothness of speech (Segalowitz, 2010; Skehan, 2014). These disfluencies (i.e., pausing, slowing down or repairing) are occasionally assessed as a sign of lower proficiency in L2 research or in the classroom, but they may also be related to affective factors such as foreign language anxiety (FLA) (Segalowitz, 2010; Tavakoli & Wright, 2020; Kormos, 2015). Horwitz et al. (1986) define FLA as "a distinct complex of self-perceptions, beliefs, feelings, and behaviors related to classroom language learning arising from the uniqueness of the language learning process" (p.

[3] This article is accepted for publication in *Dutch Journal of Applied Linguistics*: Reitsma, F. & Ruigendijk, E. (2024). The role of foreign language anxiety and task complexity on fluency in German learners of Dutch.

128). According to Horwitz et al. (1986), "anxious students feel a deep self-consciousness when asked to risk revealing themselves by speaking the foreign language in the presence of other people" (p. 129). Against this background, studies indicate that FLA negatively affects speech fluency (MacIntyre & Gardner, 1994; Pérez Castillejo, 2019, 2021; Sanaei et al., 2015; Bielak, 2022), by interfering with the cognitive processes that require attention control during speech processing (Eysenck et al. 2007).

So far, especially L2 fluency in less cognitively demanding tasks has been investigated in relation to FLA (Hewitt & Stephenson, 2012; MacIntyre & Gardner, 1994; Pérez Castillejo, 2019, 2021; Phillips, 1992). Bielak (2022) investigated the link between FLA and fluency in a more demanding task among advanced L2 learners and found that proficiency was the strongest predictor of fluency, followed by much weaker predictive power of anxiety. To our knowledge, the relationship between FLA and fluency in low-intermediate L2 learners has not been investigated in relation to task complexity[4]. Lower participant proficiency (as compared to advanced L2 participants in Bielak, 2022) may cause FLA effects to be stronger due to less automatization (i.e., efficiency with which planning, selecting words and syntactic structures, and articulating message occur (Segalowitz, 2010) in L2 use. Studying the effect of task complexity is relevant for teachers who use these tasks in classroom practice and for researchers to better understand the role of FLA on fluency. The present study aims to fill this research gap by investigating which aspects of L2 oral fluency (i.e., speed, breakdown and repair fluency, more details will follow below) relate to FLA and how this interacts with task complexity.

4.2 Literature review

4.2.2 The speech production process and L2 fluency

When teaching and evaluating L2 speech production, both teachers and researchers consider fluency, as it reflects an L2 learner's ability to formulate a message in their L2, within strict

[4] But note that Robinson (2007) focus on the interaction between input, processing and output anxiety (MacIntyre & Gardner, 1994) and language production (use of complex speech structures).

time constraints. To understand how L2 fluency is achieved, it is necessary to look at the speech production process. Levelt (1989) or Levelt et al. (1999) distinguishes three main processes in his L1 speech production model: conceptualizing, formulating, and articulating. In L1 (or highly proficient L2) speech, formulation (including lexical retrieval and grammatical encoding) and articulation are mainly automatic and allow for parallel processing with other modules (i.e., occurring virtually simultaneously and requiring little cognitive effort). Conceptualization (utterance planning) and monitoring processes require attention and may rely on serial processing, one after the other (Kormos, 2006). These L1 speech production features explain "speed and fluidity, that is, a speaker's fluency" (Tavakoli & Wright, 2020, p.9).

For less proficient L2 speakers, not only conceptualization and monitoring, but also lexical retrieval, grammatical encoding and articulation might require conscious attention (de Bot, 1992; Kormos, 2006). The process will then occur less automatically and hence more slowly and effortfully. When L2 learners have access to a larger repertoire of lexical and grammatical knowledge and progress to a higher proficiency level, their lexical access and grammatical encoding becomes more automatic and they therefore produce output of higher fluency. The fluency of L2 performance thus reflects the efficient functioning of speech production processes, namely conceptualization, formulation, articulation and monitoring (Kormos, 2006). This aspect of fluency is referred to as cognitive fluency (Segalowitz, 2010).

The degree of cognitive fluency is reflected in speech in the form of so called utterance fluency (Segalowitz, 2010), which can be further divided into speed fluency, or the number of words or (pruned[5]) syllables that are produced; breakdown fluency, or the number and length of pauses and repair fluency, which reflects hesitations and repairs (Lambert et al., 2017; Saito et al., 2018; Pérez Castillejo, 2019). Speed fluency reflects the degree of automatization of linguistic knowledge and the ease of accessing it (Pérez Castillejo, 2019, 2021; Bielak, 2022), which refer to the formulation/encoding (Levelt, 1989) aspect of cognitive fluency. Regarding breakdown fluency, pauses within clauses, which can reflect processing difficulties (Kahng, 2014), may be indicative of formulation/encoding difficulties/processes. Pauses between clauses may be indicative of content planning and conceptualization (Götz, 2013). Repair fluency is assumed to reflect the effectiveness of the monitoring system (Kormos, 1999).

[5] Pruned syllables means syllables excluding those that are repeated, reformulated or replaced.

In summary, while L1 speech processing is more automatized and stable, L2 speech processing, especially at lower proficiency levels, usually requires more control or conscious attention. Besides the degree of automatization, fluency can be affected by FLA, which will be discussed in the next section.

4.2.3 FLA, L2 fluency and task complexity

Over time, attempts have been made to classify the FLA construct as trait anxiety (i.e., a permanent likelihood of experiencing anxiety regardless of the situation); state anxiety (which may arise in response to a situation at a particular time); or situation-specific anxiety (i.e., a personal predisposition to experience anxiety in particular situations (Baran-Lucarz, 2022). Since a pioneering article by Horwitz et al. (1986), FLA has been treated as a situation-specific type of anxiety that is linked to a language-learning setting.

One explanation for the idea that FLA may negatively influence the speech production process (i.e., more specifically fluency) comes from Processing Efficiency Theory (Eysenck & Calvo, 1992) and Attentional Control Theory (Eysenck, 2010; Eysenck et al., 2007). They suggest that anxiety impairs the efficiency of cognitive processing during L2 speech production by posing additional demands on working memory and attentional processes (see also Chapter 2.2.3).

The literature that empirically investigates the claim that anxiety negatively affects fluency will now be further discussed. Sanaei et al. (2015) found strong negative correlations between FLA and speed fluency (i.e. more anxious participants had shorter runs between pauses and shorter phonation time) during a non-exam situation. Pérez Castillejo's (2019) examined how FLA and proficiency relate to L2 fluency during a final oral exam in learners with low-intermediate proficiency level. The results indicated medium and strong negative relationships between FLA and those fluency measures that reflect speech formulation/encoding (i.e. more anxious participants had a smaller phonation time, shorter runs between pauses and paused more frequently within clauses). The negative relationship between FLA and the length of pauses between clauses, which is mainly related to message conceptualization, was weaker. Regression analyses showed that FLA was a stronger predictor of fluency than proficiency. Whereas Pérez Castillejo (2021) replicated Pérez Castillejo's (2019) results, regarding the relationship between FLA and the fluency of speech formulation/encoding (i.e., mean length of run and number of mid-clause pauses), their results regarding the relationship between FLA and conceptualization fluency (the length of pauses) were not replicated, indicating that there is more evidence that formulation/encoding is affected by anxiety than conceptualization is affected. Pérez Castillejo (2021) also found that repair fluency was very weakly related to FLA and proficiency, probably due to the low proficiency level of participants. Proficiency was however a much stronger predictor of L2 fluency than FLA when fluency was measured in a task following another similar task, that is, in a condition of prior L2 processing (understood as L2 use earlier in discourse).

In addition to anxiety, cognitive fluency may also be influenced by *task complexity*. According to Robinson's Cognition Hypothesis (2001), for instance, cognitively demanding tasks focus L2 speakers' attention on language form, to increase grammatical accuracy and linguistic complexity of their L2 production at the cost of fluency (Michel et al., 2007; Robinson, 2001). The Limited Attentional Capacity Model (Skehan & Foster, 2001) on the other hand suggests that complex tasks may lead to trade-off effects, particularly between complexity and accuracy. Skehan and Foster (2001) argue that an increase in cognitive task demands puts pressure on the attentional system. Therefore, L2 learners must prioritize between linguistic complexity, accuracy, and fluency. Attending to one aspect of performance will lead to less attention to other aspects. Thus, a task that leads to linguistically more complex L2 performance may negatively affect performance in terms of both grammatical accuracy and fluency. Both Robinson (2001) and Skehan and Foster (2001) thus predict that task complexity would negatively impact fluency. De Jong et al. (2012) examined how task complexity affected L2 speaking performance. Results showed that complex tasks led to more pausing and repair behavior than simple tasks, but no significant difference in speed fluency was found.

The effect of task complexity may vary depending on the proficiency and anxiety level of the speaker. Since the automaticity of linguistic encoding is related to proficiency (Kormos, 2006; Segalowitz, 2010), more proficient L2 speakers may have more attentional resources available to conceptualize the message than less-proficient speakers, who need to divide their attention between conceptualization, formulation and monitoring (Lambert et al., 2017). In addition, the effect of task complexity might vary depending on the anxiety level of the speaker. Eysenck and Calvo (1992), Eysenck et al., (2007) and Robinson's (2001) Cognition

Hypothesis suggest that the effect of anxiety on task performance is stronger when the task is more cognitively demanding. To date, only Bielak (2022) investigated the relationship between FLA and L2 fluency in advanced L2 learners. Participants completed both a group task and a monologue (i.e., complex task). Numerous negative correlations between FLA and fluency measures, reflecting formulation (i.e., more anxious participants had shorter runs and lower articulation rate), and conceptualization (more end-pauses), were found in the group task, and only one negative correlation (more end-pauses) was found during the complex task. Regression analyses indicated that proficiency was a stronger predictor of fluency than FLA. This could be attributed to participants' advanced proficiency level and consequently more automatized L2 use. A limitation of this study was that FLA was measured twice, after the group and complex tasks, but fluency was measured only once, during the complex task; therefore, it is not certain whether the correlations between the first anxiety measurement and fluency are reliable.

To summarize, the aforementioned studies have found a link between FLA and L2 fluency, and in the literature, it is assumed that FLA negatively affects attentional processes, which in turn affects fluency in L2, mostly on formulation and encoding of L2 processes. However, the relationship between FLA and L2 fluency in low-intermediate L2 learners has not been investigated in relation to task complexity. Based on previous research, we hypothesize that a cognitively more demanding task negatively affects speed fluency of anxious learners. Regarding breakdown fluency, we expect that complex tasks elicit more pausing behavior. Considering repair fluency, we hypothesize that a cognitively demanding task will lead to fewer self-corrections and reformulations by anxious learners. The idea underlying this hypothesis is that at lower proficiency levels, when L2 knowledge is not automatized, learners' attentional capacity can be too taxed by other processes, particularly formulation, so that there is no room left to make corrections.

Our research question is therefore as follows: "To what extent is there a relationship between speed, breakdown and repair fluency measures and FLA and proficiency, and how does this interact with task complexity?".

4.3 Method

In order to answer our research question, we performed two speech production tasks with a group of German learners of Dutch to examine their fluency and related this to their level of anxiety. Participants, materials, as well as a detailed description of the analyses are presented in the following section.

4.3.1 Participants

The participants were 61 German learners of Dutch (20 male and 41 female) from three secondary schools, all of whom were in their fifth year of Dutch. The L2 learners reported German as their L1. Six participants had two L1s: three German/*Niederdeutsch*, one German/Russian, one German/Kurdish and one German/Ukrainian. Their ages ranged from 14 to 16 with an average age of 15.2 years. All students participated in this study voluntarily with signed informed consent.

4.3.2 Anxiety measure

The original FLCAS questionnaire (Horwitz et al., 1986) consists of 33 items with a 5-point Likert scale concerning learners' feelings and behavior in the foreign language classroom, and in conversation with native speakers. This instrument has been validated and widely used in language learning research.

Mak (2011) found that the FLCAS tests five subconstructs of anxiety: (a) speech anxiety and fear of negative evaluation, (b) uncomfortableness when speaking with native speakers, (c) negative attitudes toward the L2 class, (d) negative self-evaluation, and (e) fear of failing the class or the consequences of personal failure. Considering our research question only three of these subconstructs with in total 21 items were of relevance and used in our study: the first two (a) and (b), as well as (d). The "negative attitudes toward the Dutch class" subconstruct was not selected because the questionnaire was completed in the first weeks of the school year and hence, participants' responses were less likely to be influenced by their attitudes toward the class. The "fear of failing the class or the consequences of personal failure" subconstruct also was not relevant because the speech production tasks did not constitute an exam and not related to the class in that sense.

A native German speaker with a Master's degree in English translated the original English version into German. A professional translator subsequently back-translated the questionnaire and confirmed that the English and German versions were completely comparable. The questionnaire was then piloted with participants not involved in the study to test whether the items were clearly formulated. Feedback showed that one item had a translation error which was subsequently corrected. Participants completed the questionnaire in approximately five minutes. The reliability of the questionnaire as measured by Cronbach's alpha was 0.89, which indicates a high level of internal consistency.

4.3.3 Proficiency measure

The participants completed the Dutch version of the LexTALE (Lemhöfer & Broersma, 2012), a five-minute, 60-item vocabulary test, in which they must indicate whether or not a string of letters represents an existing word in the language. The test measured participants' general proficiency in Dutch. A benefit of the LexTALE is that the ranges of scores are associated with the proficiency levels as described in the Common European Framework of Reference (CEFR) (Council of Europe, 2001); see Lemhöfer and Broersma (2012) for the correspondence between LexTALE and CEFR).[6] According to Lemhöfer and Broersma (2012), LexTALE scores of 80–100% represent CEFR level C1–C2, 60–80% level B2, and below 59% level B1 and lower. Cronbach's alpha assessed the LexTALE's reliability and was 0.71, which indicates an acceptable level of internal consistency.

4.3.4 Speech production tasks

Speech data was collected using two tasks that differed in complexity. To operationalize complexity, we followed De Jong et al. (2012). In their opinion, complex tasks contain more

[6] Lextale is a test for advanced learners, it can distinguish between C1-C2, B2 and B1 or lower and is validated only for English so far, and so are the corresponding CEF levels. We therefore use the scores with care, since it may not be the same for Dutch.

elements than simple tasks; complex tasks concern a more general topic as opposed to simple tasks, which concern topics of personal life; and complex tasks involve more abstract notions as opposed to simple tasks, which involve mostly concrete notions. In our simple task, the participants discussed their weekend activities. This is a concrete topic of personal life, and the task consists of one element: “What did you do last weekend?”. In our complex task, participants were asked to persuade the class of their opinions and arguments about smartphones in the classroom. This is a more general topic; it is more abstract and consists of two elements (providing their opinion and arguments; see Appendix E). The regular class teachers confirmed that the tasks were appropriate for the participants’ Dutch level. Moreover, 14 German students of Dutch piloted the tasks. Their feedback showed that the instructions were clear and that the more cognitively demanding task was perceived as more complex than the simple task.

4.3.5 Procedure

The participants completed the FLCAS questionnaire, as well as the demographic and language background questionnaire (see Appendix F) and the proficiency test some weeks before the speech production tasks. We measured anxiety independently of the speaking tasks, but some weeks before to be better able to compare our data to similar studies (Sanaei et al., 2015; Pérez Castillejo, 2019, 2021). It was operationalized as a measure of general FLA experience in the language classroom rather than as an emotional state triggered during a speaking task. The data from the oral tasks were collected on regular school days where the participants met individually with the researcher whom they had already met during the first session (completion of questionnaires and proficiency test). The participants were given the simple and complex tasks on paper and, after checking for comprehension, were instructed to answer without time to prepare. They were asked to speak for two minutes per task and told that their performances were audio recorded. In both tasks, participants spoke only to the researcher, not to a classmate. The role of the researcher during the performance was limited to nodding to show understanding. The tasks had not been practiced before and were always taken in the same sequence; first the simple task and then the complex task. All speech samples were recorded on an Olympus LS 10 digital voice recorder.

4.3.6 Data coding and analysis

In the simple task, participants spoke on average for 58.00 seconds (SD = 25.41) from first to last syllable, and in the complex task for 54.33 seconds (SD = 22.39). All speech recordings of the two tasks of 59 participants (roughly two hours of speech in total) were transcribed in detail, including information regarding pauses and repairs. Two of 61 participants were absent when the speech data was collected. Three datasets were discarded due to technical problems. One recording lasted five seconds: we excluded it from the analyses, as fluency measures are unstable for short speech samples. This resulted in 56 datasets from the simple task and 58 from the complex task. These 114 recordings were transcribed into Analysis of Speech (AS) units following the procedures of Foster et al. (2000). The AS unit is defined as "a single speaker's utterance consisting of an independent clause, or sub-clausal unit, together with any subordinate clause(s) associated with either" (Foster et al., 2000, p. 365).

As a first step, the "textgrid-to-silences" function in PRAAT (Boersma & Weenink, 2015) was used to automatically mark silent pauses. Since the background noise was occasionally disturbing automatic detection of silent pauses, the identified pauses were manually checked and adjusted as necessary. Only silent pauses (i.e., greater than or equal to 0.25 seconds) and filled pauses were marked as pauses. Short "micropauses" (Riggenbach, 1991) that are irrelevant for measures of L2 fluency were excluded (De Jong & Bosker, 2013). All filled pauses were manually marked. For each pause it was determined whether it was mid-clause or end clause through careful listening to the recordings and examination of the transcript, which had been marked with clause boundaries. Mid-clause pauses were filled and silent pauses occurring in the middle of clauses. Pauses between clauses and when a conjunction occurred between clauses were considered to be end-clause pauses. For each speech recording, syllables were counted manually. Syllables were counted for the purpose of calculating speech rate, articulation rate and mean length of run (i.e., stretches of speech uninterrupted by pauses > 0.25 seconds). The number of self-corrections and reformulations during each performance was also counted manually.

Regarding fluency measures, it has been argued (Bosker et al., 2013; Huensch & Tracy-Ventura, 2016; Hunter, 2017; Skehan, 2015, Tavakoli et al., 2020) that composite and pure fluency measures need to be distinguished. Speech rate, mean length of run and phonation time ratio are described as composite measures because they combine two or more of the aspects of fluency. A measure such as mean length of run combines speed and breakdown fluency. In contrast, pure measures relate to one of the three aspects of fluency: speed, breakdown, or repair. For example, articulation rate relates to speed fluency. Pure measures have been argued to tell more about the underlying processes involved in speech production (Huensch and Tracy-Ventura, 2017; i.e., linking specific fluency measures to stages of speech production; Lambert et al., 2017).

The present study used several different (composite and pure) fluency measures (see Table 1). Sanaei et al. (2015), Pérez Castillejo (2019, 2021) and Bielak (2022) included a limited number of fluency measures. To capture a more complete picture of how anxiety relates to speed, breakdown and repair phenomena in a simple and complex task, we used a broad range of fluency measures (see Table 1 for details on their operationalization). For comparison purposes fluency measures from FLA and fluency studies were used, namely: speech rate, articulation rate; phonation time ratio; mean length of run (Sanaei et al., 2015); number of mid-clause and end-clause pauses; mean length of mid-clause; mean length of end-clause pauses (Pérez Castillejo, 2019; Bielak, 2022) and number of reformulations and self-corrections (Pérez Castillejo, 2021). In addition, we used other commonly used speed, breakdown and repair fluency measures (i.e., number of silent and filled pauses per minute, number of pauses per minute, and mean length of pauses) based on earlier fluency studies (See Kahng, 2022).

Table 1

Fluency Measures and Their Operationalizations

Measure	**Operationalization**
Speed fluency	
Speech rate per minute	The total number of syllables divided by total sample time[7] and multiplied by 60.

[7] Sample time refers to total duration of the recording

Mean length of run	The total number of syllables divided by the total number of runs.
Articulation rate per minute	The total number of syllables divided by the total speaking time (excluding all pauses > 0,25 sec) and multiplied by 60.
Phonation time ratio	Total speaking time divided by total sample time and multiplied by 100.
Breakdown fluency	
Number of silent pauses per minute	The total number of silent pauses divided by the total speaking time and multiplied by 60.
Number of filled pauses (*uhm, eh*) per minute	The total number of filled pauses divided by the total speaking time and multiplied by 60.
Number of pauses per minute	The total number of pauses divided by the total speaking time and multiplied by 60.
Number of mid-clause pauses per minute	The total number of mid-clause pauses divided by the total sample time and multiplied by 60.
Number of end-clause pauses per minute	The total number of end-clause pauses divided by the total sample time and multiplied by 60.
Mean length of pauses	The total length of pause time divided by number of pauses.
Mean length of mid-clause pauses	The total length of mid-pause divided by number of mid-clause pauses.
Mean length of end-clause pauses	The total length of end-pause divided by number of end-clause pauses.
Repair fluency	

Number of reformulations and overt self-corrections per minute	The total number of reformulations plus total number of overt self-corrections divided by total sample time and multiplied by 60.

Data for six participants (i.e., 12 recordings; 10.53% of the data) was re-examined for syllable count by an independently trained teacher/researcher. Additionally, the rater analyzed and coded 10% of the data again for pause type, pause length, pause location, self-corrections and reformulations. Interrater reliability was 98% agreement on syllable count, 98% on pause type, 90% on pause length, 92% on pause location and 92% on self-corrections and reformulations, demonstrating high interrater reliability.

4.3.7 Statistical analysis

As a first step, we checked the data (fluency measures, anxiety, and proficiency) for normality. The results of Shapiro-Wilk test indicated that the fluency measures: mean length of run, number of filled pauses, mean length of pauses, mean length of mid and end-clause pauses, and reformulations were not normally distributed ($p < 0.05$).

To explore the research question, a correlation analysis was run with the L2 learners' data on speed, pause phenomena, repair, anxiety and proficiency scores. Due to non-normality of some fluency measures, we decided to use Spearman rank-order correlation test. Based on this exploration, we decided to run regression analyses. Prior to the regression analyses, the assumptions of LMM were checked, including linearity, absence of collinearity, homoscedasticity, and normality and no violations were noted. Then a mixed-effects analysis was conducted, using the lme4 package (Bates et al., 2015) in R Version 4.0.2 (R Core Team, 2020).

First, we tested interactions, for every fluency measure that correlated with anxiety and/or proficiency (i.e. speech rate, frequency of mid-clause pauses, number of pauses, number of silent pauses and phonation time ratio) and center the predictors involved in interactions (i.e., Anxiety, Proficiency and Task type). However, for all five fluency measures the models turned out to be non-significant and did not improve the model fit. We also added random slopes with interactions, but we received a warning that the data was not large enough to estimate random effects.

Finally, for each fluency variable, we built a model with Anxiety, Proficiency and Task type as fixed effects and Subject as a random effect and then added School as a fixed effect. The fixed-effect factor, Task type, was added to control for differences between performing the simple and complex tasks. School was added as a fixed effect to control for possible differences between the three schools that participated in the study. The variable Subject was added as a random effect to control for differences between participants. The models were subsequently compared; in essence, we tested whether to include the fixed effect School or not, and if the first model is preferable compared to second one. Model comparisons were assessed using the ANOVA (Model 1, Model 2) function in R.

4.4 Results

In this section, the results are summarized. Participants' FLCAS scores ranged from 23 to 72 or (when dividing the number of questions by the score) from 1.10 to 3.43. According to Horwitz (2013), averages below 3 indicate a low anxiety level, averages of approximately 3 a moderate level of anxiety and averages in the range near 4 to 5 as fairly anxious. The participants' mean score (2.23) can thus be considered to be at a low anxiety level. The scores on the proficiency test (LexTALE) showed that participants' scores ranged from 38.75% to 72.50%, with a mean score of 55.50% (SD = 6.06). This indicates that their Dutch proficiency level was situated between B1 (or lower) and B2 on the CEFR.

Table 2 contains descriptive statistics for all fluency measures in the simple and complex speech production task. The results indicate that the participants spoke faster in the complex (mean = 131.28; SD = 27.06) than in the simple task (mean = 123.85; SD = 22.33). They articulated on average 204.67 (simple task) versus 211.23 (complex task) syllables, spoke for approximately 63% (simple task) and 65% (complex task) of the total performance time (i.e., phonation time ratio) and paused after uttering three to ten syllables (i.e., mean length of run). In both tasks, participants used more silent than filled pauses. Pauses were more frequent within than between clauses, but mid-clause pauses were shorter than end-clause pauses (simple task: 0.61 versus 0.79; complex task: 0.62 versus 0.72). On average, participants made more reformulations and self-corrections per minute in the simple task than in the complex task (simple task: mean = 1.72, SD = 1.82; complex task: mean = 1.67, SD = 1.63). Overall, the complex task resulted in faster speech and articulation rate, longer runs, more phonation time, more and longer mid-clause pauses, and less and shorter end-clause pauses than the simple task.

Table 2

Descriptive statistics (M = mean; SD= Standard Deviation, Median, Min-Max = range) for the fluency measures in the simple and complex task

	Simple task (N=56)		Complex task (N=58)	
Measures	**M (SD) Median**	**Min - Max**	**M (SD) Median**	**Min – Max**
Speed fluency				
Speech rate per minute	123.85 (22.33) 123.82	77.08 - 168.50	131.28 (27.06) 131.83	48.37 - 184.83
Mean length of run	4.89 (1.55) 4.55	2.95 - 9.31	5.00 (1.33) 4.74	3.05 - 9.88
Articulation rate	204.67 (22.98) 206.77	151.53 - 249.16	211.23 (22.48) 212.07	159.58 - 254.32
Breakdown fluency				
Phonation time ratio	63.01 (9.04) 62.53	46.07 - 81.50	64.85 (9.25) 66.49	25.83 - 82.11
Number of silent pauses per minute	40.86 (12.97) 40.26	3.32 - 68.32	39.40 (9.92) 38.21	16.56 - 56.86
Number of filled pauses per minute	12.57 (9.44) 10.56	0.00 - 47.92	12.60 (9.00) 11.43	0.00 - 35.33
Number of pauses per minute	53.43 (17.80) 54.21	22.56 - 101.60	52.00 (13.30) 52.44	21.39 - 82.71
Number of mid-clause pauses per minute	17.57 (6.05) 15.71	4.77 - 32.06	18.76 (5.89) 18.66	6.79 - 30.91
Number of end-clause pauses per minute	14.74 (4.25) 14.04	6.84 - 29.89	14.08 (4.08) 14.36	5.86 - 24.42
Mean length of pauses (sec)	0.69 (0.13)	0.48 - 1.10	0.67 (0.26)	0.37 - 2.15

	0.67		0.62	
Mean length mid-clause	0.61	0.37 - 1.04	0.62	0.28 - 1.72
pauses (sec)	(0.15)		(0.25)	
	0.59		0.55	
Mean length of end	0.79	0.47 - 1.80	0.72	0.41 - 2.36
pauses (sec)	(0.22)		(0.31)	
	0.71		0.66	
Repair fluency				
Number of	1.72	0.00 - 6.58	1.67	0.00 - 4.68
reformulations and	(1.82)		(1.63)	
corrections per minute	1.35		1.00	

Table 3

Correlations between Anxiety (FLCAS), Proficiency (LexTALE) and Fluency Measures

	Anxiety		Proficiency	
	Simple task	Complex task	Simple task	Complex task
Measure				
Speed fluency				
Speech rate per minute	-.111	-.069	**.325****	.092
Mean length of run	-.110	-.122	.250	.072
Articulation rate	-.060	-.035	.015	-.002
Breakdown fluency				
Phonation time ratio	-.124	-.040	**.329****	.087
Number of silent pauses per minute	.137	.082	**-.266****	-.084
Number of filled pauses per minute	.124	.123	-.237	-.175
Number of pauses per minute	.166	.144	**-.320****	-.181
Number of mid-clause pauses per minute	.117	**.317****	**-.292****	-.253
Number of end-clause pauses per minute	.159	-.167	-.045	.056
Mean length of pauses	-.074	-.086	-.072	.083
Mean length of mid-clause pauses	-.021	.024	-.132	.039
Mean length end-clause pauses	-.079	-.130	-.012	.127
Repair fluency				
Number of reformulations and self-corrections per minute	.024	.139	-.164	-.103

*$p < 0.01$, **$p < 0.05$

Table 3 displays correlation coefficients between anxiety and the fluency measures, which basically show low values. We found only one significant correlation between anxiety and fluency: number of mid-clause pauses in the complex task (r = .317, $p < 0.05$). The negative correlation between anxiety and proficiency (LexTALE scores; r = -.128, p = 0.32) was not statistically significant. Higher proficiency (LexTALE scores) was significantly correlated in the simple task with higher speech rate (simple task: ***r*** = .325, $p < 0.05$); higher phonation-time ratio (PTR; ***r*** = .329, $p < 0.05$); fewer silent pauses (***r*** = -.266, $p < 0.05$), fewer pauses (***r*** = -.320, $p < 0.05$); and fewer mid-clause pauses (***r*** = -.292, $p < 0.05$).

Plonsky and Oswald (2014) recommend the following benchmarks for the interpretation of effect size in correlation coefficients: close to 0.25 small, 0.40 medium, and 0.60 large. The effect size of the statistically significant correlations (*r* = -.266 to *r* =.329) in the current study are thus considered small.

Regression analyses were run to further test the relative contributions of the factors Anxiety, Proficiency and Task type to the variance in the five fluency measures that (at least weakly) correlated with anxiety or proficiency (i.e., speech rate, phonation time ratio, number of silent pauses, number of pauses and mid-clause pauses).

Tables 4, 5 and 6 present a summary of the models on factors affecting speech rate, number of mid-clause pauses and number of pauses that gave a significant effect (for a summary of the models and model comparisons for which we did not find a statistical effect, see Appendices G and H).

Table 4

Results of Mixed-effects Models on Factors Affecting Speech Rate

Fluency measure	**Fixed effects: Factor**	***Estimate***	***SE***	***p***	**Random effects: Factor**	**Variance**
Speech rate	Intercept	92.61	30.69	**0.001****	Participant	344.5
	Profi-ciency	0.72	0.48	0.137		
	Anxiety	-0.16	0.26	0.537		
	Task type	6.26	3.11	**0.05***		

Table 4 presents the resulting model on factors affecting speech rate. Adding the fixed factor School did not significantly improve the model ($\chi^2(2) = 2.33$, $p = 0.31$). Task type emerged as the strongest predictor of speech rate, with the largest coefficient. The more complex the task, the higher the speech rate. Proficiency and Anxiety did not significantly predict speech rate.

Table 5

Results of Mixed-effects Models on Factors Affecting Number of Mid-clause Pauses

Fluency measure	**Fixed effects: Factor**	***Estimate***	***SE***	***p***	**Random effects: Factor**	**Variance**
Number of mid-clause pauses	Intercept	25.15	6.89	**<0.001**	Participant	15.20
	Proficiency	-0.23	0.11	**0.038**		
	Anxiety	0.10	0.06	0.084		
	Task type	1.41	0.79	0.081		

Table 5 shows the resulting model on factors affecting number of mid-clause pauses. The results from the model comparison indicate that adding of the factor School does not improve the model ($\chi^2(2) = 1.71$, $p = 0.42$). Proficiency was significantly and negatively related to number of mid-clause pauses. The more proficient the speaker, the less mid-clauses pauses they use. Anxiety and Task type were not found to be significant predictors.

Table 6

Results of Mixed-effects Models on Factors Affecting Number of Pauses

Fluency measure	**Fixed effects: Factor**	***Estimate***	***SE***	***p***	**Random effects: Factor**	**Variance**
Number of pauses	Intercept	76.45	18.58	**<0.001**	Participant	117.6
	Profi-ciency	-0.58	0.29	**0.048**		
	Anxiety	0.18	0.15	0.242		
	Task type	-0.82	2.03	0.689		

Table 6 presents the results of mixed-effects models on factors affecting number of pauses. The results from the model comparison show that adding the fixed factor School does not significantly improve the model ($\chi^2(2) = 0.44$, $p = 0.80$). Proficiency was negatively linked to number of pauses; the more proficient the speaker, the less pauses the speaker used.

To sum up, modeling demonstrated that Task type was a significant predictor of speech rate. Proficiency emerged as predictor of number of mid-clause pauses and number of pauses. Anxiety was not a significant predictor of any fluency measures. Furthermore, in all model comparisons, the addition of the fixed-effect School did not significantly improve the models.

4.5 Discussion

This study set out to answer the research question to what extent there is a relationship between speed, breakdown, and repair fluency measures, FLA, and proficiency, and how this interacts with task complexity. Correlational analyses showed that FLA was negatively related to one break-down fluency measure, while proficiency was positively related to numerous speed and breakdown fluency measures. Regression analyses indicated that proficiency was a stronger predictor than FLA. Below, we will first discuss the results of the correlation analysis, followed by the regression analysis, limitations, and possible directions for further research.

4.5.1 The relationship between FLA, proficiency and L2 fluency

The correlation analysis results that served as a first exploration of the data showed that more anxious speakers used significantly more mid-clause pauses in the complex task, but with a small effect size. The results generally support the theoretical claims of the Cognition Hypothesis (Robinson, 2001) and Processing Efficiency Theory (Eysenck & Calvo, 1992) regarding the negative role of anxiety in L2 speech processing, as well as previous research (Pérez Castillejo 2019, 2021). In contrast, Bielak (2022) found no significant correlations between FLA and mid-clause pauses in a cognitively demanding task among advanced L2 learners. Mid-clause pauses reflect cognitive efficiency during formulation. Especially lexical retrieval and syntactic encoding (Kahng, 2014) may require attention. The participants in the present study had generally lower proficiency, which suggests that their ability to formulate may require greater attention control (Kormos, 2006). Thus, any interference that limits attention control during L2 speech production, such as FLA, may impact pausing behavior within clauses (Pérez Castillejo, 2021).

There were positive correlations between proficiency and fluency (i.e. speech rate and phonation time ratio) and negative correlations (i.e. number of silent and mid-clause pauses and number of pauses) in the simple task. These pausing and speed phenomena are associated with message formulation and encoding during speech processing, and they indicate the efficiency of access to L2 knowledge (Kahng, 2014). In contrast, fluency measures that may be associated with message conceptualization such as number of end pauses were not related to proficiency in a statistically significant way. For L2 speakers, especially at lower proficiency levels, formulation and encoding may require more attention control than conceptualization (De Bot, 1992; Kormos, 2006; Segalowitz, 2010). This could explain why proficiency relates differently to the fluency measures associated with each of these processes. Note that the task we used, LexTALE has been validated for English, but not yet for Dutch, so especially with respect to assigning scores to the European Reference Framework, we cannot form any strong conclusions. The scores on the task can be used to correlate individual variation in proficiency though, which we have done here.

Furthermore, repair fluency was not significantly related to FLA or proficiency. This is in line with the findings of Pérez Castillejo (2021). Self-repairs may reflect that attentional resources are being used for monitoring (Kormos 2000), but at lower proficiency levels, when L2 knowledge is not automatized, attention may be needed for other processes (particularly formulation), and hence FLA or proficiency may not significantly affect monitoring.

4.5.2 Predictors of L2 fluency

We additionally investigated the contributions of the predictors (i.e., FLA, proficiency and task type) on the fluency measures. First, proficiency was a significant predictor of two fluency measures, whereas FLA was not a significant predictor of any fluency measure. In Pérez Castillejo's study (2019) and its replication (2021) on the relationship between FLA, proficiency and fluency during a final oral exam, the roles seem reversed: FLA was a significant predictor of the fluency measures analyzed and proficiency was not. Explanations for the contradictory results could be that in the present study, the tasks were not performed during an exam situation, and the participants' anxiety levels were lower compared to the levels measured in previous research (Bielak, 2022; Pérez Castillejo, 2019, 2021; Saito et al., 2018).

Secondly, we found a significant positive effect of the predictor task type on speech rate. The more complex the task, the higher the speech rate. As a reminder, the complex task resulted in a faster speech rate and articulation rate, longer runs, more phonation time, fewer and shorter end-clause pauses than the simple task. It is somewhat surprising that a cognitively demanding task results in a higher speech rate, as the Cognition Hypothesis (Robinson, 2001) and Limited Attentional Capacity Model (Skehan & Foster, 2001) predict that these tasks may negatively affect fluency. One possible explanation is that the complex task may not have been sufficiently complex, thus not resulting in more formulation problems. Another explanation is that participants had already used the L2 in the simple task and benefited from it while performing the complex task. The earlier activation of linguistic material facilitates later access to and use of related linguistic knowledge, which can lead to improved performance (McDonough & Trofimovich, 2008). This is also what Pérez Castillejo (2021) and Bielak (2022) found in their research.

While this study is one of the first to address a broad range of fluency measures and address the influence of FLA, proficiency and task complexity, it also has some limitations. First, the participants produced relatively short speaking samples. The participants were instructed to speak for two minutes for each task, but on average they did not speak for more than one minute. Another limitation is that the tasks were always taken in the same sequence. This should have been varied to avoid the potential effect of repetitiveness (better performance) in the second task. Future research might compare whether it makes a difference to vary the sequence of the simple and complex tasks. Finally, in the absence of L1 data from the participants, it is difficult to claim that the results obtained here are only due to the speakers' L2 fluency behavior and not affected by individual differences in their L1 speaking style (De Jong et al., 2015).

As far as pedagogical implications are concerned, the results of this study suggest that FLA may not have a strong influence during speaking tasks that are not part of formal tests. This could encourage teachers to base grades for speaking proficiency on classroom observations as well, rather than solely on formal tests (Bielak, 2022).

4.6 Conclusion

In conclusion, the current study examined the extent to which a relationship exists between speed, breakdown and repair fluency measures and FLA and proficiency, as well as how this interacts with task type. Based on previous research (e.g., Eysenck & Calvo (1992); Sanaei et al. (2015); Pérez Castillejo (2019, 2021); Robinson's Cognition Hypothesis (2001); Bielak (2022), we expected a cognitively demanding task to negatively relate to speed fluency and require more pausing behavior and fewer self-corrections and reformulations by anxious speakers. The expectations about the relationship between FLA and fluency were partially confirmed for breakdown fluency. The results indicated that FLA was a negative but not a significant predictor of number of mid-clause pauses, which may be related to participants´ relatively low anxiety level. In contrast, proficiency related to numerous speed and breakdown fluency measures and was found to be a significant predictor of two breakdown fluency measures: number of mid-clause pauses and number of pauses. Unexpectedly, task type positively predicted speed fluency, suggesting that a cognitively demanding task led to more rather than less fluent performance. This finding might suggest that the earlier activation of the L2 during the simple task could facilitate later access and use of L2 during the complex task.

5. The effect of an app-based training on L2 fluency and foreign language anxiety[8]

Abstract

Research shows that digital technology can support second language (L2) fluency development. However, most studies do not take a pedagogic perspective on improving fluency. This study investigated the effects of app-based training on L2 fluency and foreign language anxiety among 89 German high school learners of Dutch. Group 1 engaged in app-based fluency training followed by app-based speaking lessons that did not focus on fluency; Group 2 completed these modules in the reverse order. The control group attended regular classes without using the app. Results indicated that app-based training in general led to significant progress in speech rate, articulation rate, phonation time ratio, and a reduction in the number of end-clause pauses. Specific app-based fluency training resulted in more progress on articulation rate compared to app-based speaking lessons. The control group developed much slower on fluency. In conclusion, this study highlights the benefits of integrating app-based training into classroom learning.

Keywords: *fluency, technology-enhanced language learning, mobile-assisted language learning, fluency training, foreign language anxiety*

5.1 Introduction

Within communicative approaches to L2 learning and teaching, speaking fluently has received growing interest (Hanzawa, 2021). Fluency refers to the ease or automaticity with which a learner produces speech and is manifested in flow, continuity, and smoothness of speech (Segalowitz, 2010). Fluency is related both to the learners' linguistic knowledge and their

[8] This chapter has been submitted to a journal and is currently under review: Reitsma, F., Michel, M., & Ruigendijk, E. *The Effect of an App-Based Training on L2 Fluency and Foreign Language Anxiety: Drawing on the Reppen App.*

ability to efficiently retrieve and use that knowledge in real time (Housen & Kuiken, 2009). Crucially, the ability to speak fluently develops only through frequent and sustained practice in speech production. However, in some classroom settings, opportunities for such practice are limited. Moreover, teachers face several challenges in promoting the development of fluency: ensuring that all learners have the opportunity to produce spoken output, supporting those who experience speaking anxiety, and providing individual feedback on speaking performance.

Regarding the first challenge - ensuring that all learners have the opportunity to produce spoken output - Swain's (1995) Output Hypothesis suggests that producing output is essential for fluency development. However, not all students actively participate in classroom speaking activities. In group work, for example, a few students often dominate the conversation while others remain passive. Moreover, learners often perform well in structured tasks involving memorized phrases, but struggle with fluency during open-ended speaking tasks (Corda et al., 2012). This suggests a need for instructional support. Dörnyei and Kormos (1998) argue that learners could benefit from strategies that help maintain the flow of speech. Instruction in the use of such strategies, for instance the use of lexical fillers, may support learners in this regard (Tavakoli et al., 2016). The second challenge concerns learners who feel anxious when speaking the L2. This type of situation-specific anxiety, known as foreign language anxiety (FLA), is closely linked to the language-learning context (Horwitz et al., 1986). FLA can lead to avoidance behaviors and minimal participation, which further limits opportunities for speaking practice. Research has shown that FLA can negatively impact learners' fluency (Pérez Castillejo, 2019; Rood & de Jong, 2023). Therefore, helping anxious learners build confidence and supporting them in developing their fluency is essential. Finally, the third challenge involves the provision of feedback. Although feedback is a crucial component of the learning process (Hattie & Timperley, 2007), it is difficult for teachers to provide individualized feedback on spoken language during real-time classroom settings due to the transient nature of speaking and large class sizes.

To address these challenges, the present study explores the potential of technology-enhanced language learning (TELL). TELL refers to the use of digital tools and applications in language learning contexts (Chang & Hung, 2019). Previous research suggests that learners express themselves more easily in digital environments than in face-to-face classroom settings

(Côte & Gaffney, 2018), experience greater autonomy, and are more motivated to engage in speaking tasks (Gonzalez-Lloret & Ortega, 2014). A promising development within TELL is the use of Automatic Speech Recognition (ASR), which allows learners to receive immediate, individualized feedback on their spoken input (Cardoso, 2022). While a small number of studies have examined TELL applications aimed at improving L2 fluency and have reported positive results (e.g., Tecedor & Campos-Dintrans, 2018; Jiang et al., 2023), they do not adopt a pedagogical perspective that includes explicit fluency strategies, nor do they use ASR to provide feedback on fluency.

To address this gap, the present study - conducted in German high schools, which enhances the ecological validity of the study - investigates whether an intervention delivered through a mobile app can support learners in developing L2 fluency and reduce FLA. We address the following research questions: (1) how app-based intervention in general affects L2 fluency, (2) whether an app-based intervention incorporating fluency strategies affects L2 fluency, and (3) how app-based intervention in general affects FLA[9].

5.2 Literature review

5.2.1 L2 fluency

To understand how L2 fluency is achieved, it is important to consider the underlying speech production mechanism. According to widely accepted models, speech production involves the stages of conceptualization, formulation, articulation and monitoring (Levelt, 1989). In L1 (or highly proficient L2) speech, formulation and articulation processes are largely automatic and allow for parallel processing, that is, they occur incrementally and require little cognitive effort. Conceptualization (utterance planning) and monitoring processes need attention and, accordingly, rely on serial processing, one after the other. For less proficient L2 speakers,

[9] This study was conceptualised and implemented before the release of ChatGPT and other generative AI (genAI) tools. While the latter has introduced ample opportunities for TELL, there remain many open questions. The current study can inform future (also genAI) contexts, given its controlled context of an app developed specifically for the purpose of this study.

formulation (i.e., lexical retrieval, grammatical encoding) and articulation (i.e., forming speech sounds) demand conscious attention as well. Therefore, they are less automatic, relying on consecutive serial processing, which is likely slower and more effortful (Kormos, 2006). L2 fluency reflects the efficiency of speech production processes, known as cognitive fluency, often referred to as utterance fluency which is subdivided into speed, breakdown and repair fluency (Segalowitz, 2010). Speed fluency refers to how many words or syllables are produced in a specific amount of time (e.g., per minute or second). It reflects the degree of automatization of linguistic knowledge and the ease of accessing it (Pérez Castillejo, 2019). As such, it relates to formulation and encoding (Levelt, 1989). Breakdown fluency is typically measured by the number and length of pauses. Pauses within clauses, which can reflect processing difficulties (Kahng 2014), may be indicative of formulation and encoding constraints. Pauses between clauses are typically associated with content planning and conceptualization (Götz, 2013). Finally, repair fluency refers to hesitations and self-repairs and is assumed to reflect the effectiveness of the monitoring system (Kormos, 1999).

5.2.2 L2 fluency development and strategy training

Swain's (1985) Output Hypothesis stresses that L2 learners need ample opportunities to produce output to develop fluency. Speaking enables learners to notice gaps in their linguistic knowledge that prevent them from expressing exactly what they intend to say. This awareness creates a natural drive to close those gaps, for example, by looking up words or experimenting with new language forms in conversation. This process not only supports language development but also promotes fluency. As learners' L2 proficiency increases, greater fluency is expected, as they gain access to a broader range of lexical and grammatical knowledge, which they can retrieve and use more quickly and automatically through practice (Kormos, 2006).

However, not all learners achieve fluency easily. Dörnyei and Kormos (1998) argue that learners struggling with speaking fluently could benefit from *strategies* to maintain the flow of speech and provide more processing time. Accordingly, fluency can be enhanced by using, for example, lexical fillers (e.g., *so, well, you know;* e.g., Tavakoli et al, 2016). This contributes to more natural speech and reduces time spent in silence while processing. Additionally, employing formulaic sequences or high-frequency multi-word units (e.g., *by the* *way; in other words*) adds to fluency by reducing attention and processing effort during retrieval and articulation (e.g., Wood, 2010). This allows L2 learners to focus more on conceptualization and formulation of the rest of the utterance. Finally, using circumlocutions and paraphrases can be ways to overcome lexical-search issues (Dörnyei & Kormos, 1998) as they allow learners to buy some time to compensate for a missing vocabulary item.

It seems that instruction on such strategies (lexical fillers, formulaic sequences, and circumlocution) could help learners. Indeed, Chamot and Harris (2019), state that strategy training enables learners to gain more control over their learning process and foster their development as effective and independent learners. Establishing control can be exercised through metacognitive training, that is by increasing learners' knowledge about fluency and fluency strategies and how to use that knowledge to improve their fluency (Anderson, 2003 - see below).

To the best of our knowledge, only two studies have looked at the impact of strategy instruction on the development of fluency, both with designs employing one experimental group and one control group, with the experimental group receiving classroom-based fluency strategy training and teacher-provided feedback. Seifoori and Vahidi (2012) found that after fluency strategy training combined with awareness-raising activities - such as discussing the importance of reducing long pauses -learners used fewer filled mid- and end-clause pauses, and fewer unfilled mid-clause pauses than the control group. Tavakoli et al. (2016) also combined strategy training with awareness-raising activities over a period of four weeks. They found that training led to increased fluency in terms of speech and articulation rate, mean length of run and fewer repairs, but did not affect pausing behavior. These few studies suggest that strategy instruction might be an effective way to enhance fluency. However, they did not examine whether the use of fluency strategies had an effect on learners' fluency in subsequent speaking tasks. Including a delayed post-test could have provided an opportunity to investigate whether any observed effects were sustained over time.

5.2.3 Using technology to improve fluency and reduce FLA

Although previous studies (Seifoori & Vahidi, 2012; Tavakoli et al., 2016) have demonstrated the effectiveness of teacher-led fluency strategy instruction using printed learning materials, the rapid advancement of digital technology in recent decades has opened up new possibilities for delivering such instruction through TELL. TELL offers several advantages over traditional classroom-based instruction. First, it allows learners to practice individually and at their own pace, thereby providing *all* learners with opportunities for oral practice in the L2 (Jiang et al., 2023). Second, with the integration of ASR technology, learners can receive immediate, individualized speech analyses and instructional feedback on their performance (McClain, 2016). This stands in contrast to classroom-based speaking instruction, which often follows a one-size-fits-all approach, focusing on general issues relevant to the entire group. A few studies have shown that using TELL positively impacts L2 fluency development in the context of online voice recording and video conferencing (Tecedor & Campos-Dintrans, 2018), flipped classrooms (Jiang et al., 2023) and mobile games (Grimshaw & Cardoso, 2018). However, none of these studies investigated the effects of individual speaking practice with an app, nor did they examine the use of ASR technology to provide immediate, individualized feedback on fluency. We are not aware of earlier research focusing on app-based fluency strategy instruction. The current study aims to fill this gap.

In addition to its potential for promoting fluency, TELL may help reduce FLA. Interacting with a native speaker or a language teacher can provoke anxiety among L2 learners, often leading to a loss of confidence and a reduced willingness to speak (Horwitz, 2010). Research on the relationship between FLA and L2 fluency shows that highly anxious speakers tend to produce less speech (Pérez Castillejo, 2019) and are perceived as less fluent (MacIntyre & Gardner, 1994). In contrast, learners generally report feeling less anxious and more comfortable expressing themselves in digital learning environments compared to traditional face-to-face classroom settings. For example, Côté and Gaffney (2018) found that written computer-mediated communication reduced anxiety levels among L2 learners. Although written communication allows learners more time to carefully plan and edit their output, it does not provide opportunities for oral communication. Grimshaw and Cardoso (2018) found that game-based speaking tasks also helped reduce anxiety in L2 learners. However, whether TELL can reduce FLA when learners practice speaking skills individually with an app, in a setting where no one is listening, remains, to our knowledge, unexplored. To address the research gaps identified above, the current study investigates the effects of an app-based L2 fluency training program.

This study addresses three research questions:

RQ 1: To what extent does app-based intervention in general affect L2 fluency?

RQ 2: To what extent does specific app-based training of fluency strategies affect L2 fluency?

RQ 3: To what extent does app-based intervention in general feedback affect FLA?

Based on the literature, we expect that the app-based intervention (RQ1) will improve fluency by providing opportunities for output for all learners (Swain, 1985; Grimshaw & Cardoso, 2018). Additionally, feedback may encourage learners to reflect on their performance and make adjustments in subsequent attempts. Regarding RQ2, we expect that app-based fluency strategy training lead to greater fluency improvements than speaking lessons without focus on fluency (Tavakoli et al., 2016; Seifoori & Vahidi, 2012; Swain, 1985). While both types of instruction are expected to improve participants' fluency, explicit fluency training is hypothesized to result in greater gains, as it equips learners with concrete strategies to compensate for any loss in fluency during speaking tasks (see 2.4). Finally, we expect (RQ 3) that individual app-based practice will lead to a reduction in FLA (Côte & Gaffney, 2018; Grimshaw & Cardoso, 2018). To investigate the research questions, we developed an app called *Reppen*, which is used for app-based intervention in this study.

5.2.4 The Reppen app

The app *Reppen*, designed for the purpose of training L2 fluency, consists of two parts: fluency training and speaking lessons (see supplementary data, Appendix I, for screenshots from the app). The *fluency training* is guided by the Cognitive Academic Language Learning Approach (CALLA; Chamot, 2009). This instructional model was chosen due to its strong theoretical support based on research into language learning strategies. CALLA has been widely applied to all four language skills—listening, speaking, reading, and writing—with empirical evidence supporting its effectiveness (Chamot, 2005).

The CALLA model involves five steps of instruction: preparation, presentation, practice, evaluation, and expansion, with the goal of supporting students in becoming independent learners who can assess and reflect on their own learning. In the preparation stage of the fluency training, learners' awareness of the importance of fluency is raised by identifying features of (dys)fluent speech that might impact the listener (e.g., long pauses and hesitations), and they are instructed on how to avoid such dysfluencies. Second, in the presentation phase, they learn about fluency strategies: circumlocution (i.e., describe the unknown word); formulaic sequences, such as *How do you say that again;* and lexical fillers such as *well*. During the practice stage, learners apply the target strategies on speaking tasks, and the app automatically analyzes fluency and provides feedback. In stage four, evaluation, participants evaluate the effectiveness of strategies used and reflect on overall performance. Finally, the expansion stage encourages participants to use the target strategies in other speaking tasks. Design features of the fluency training app include speech recording, playback, and automatic fluency analyses. Regarding the automatic fluency analysis, the app provides immediate feedback on two fluency measures: speech rate and phonation time ratio.

The second module of the app is *task-based speaking lessons*. Task-based instruction focuses on meaning and includes three phases (Ellis, 2003). The pre-task phase prepares learners for task performance: the app provides vocabulary and language functions (e.g. describing activities; describing graphs). Subsequently, learners perform the main speaking task. The post-task phase engages them in reflection activities. The app provides feedback on aspects not directly related to fluency such as lexical diversity. Its features include speech recording, playback, automatic analyses of lexical diversity, feedback, and advice on how to improve lexical diversity. This aspect of the training is discussed elsewhere since we focus on fluency in this study.

5.3 Methods

5.3.1 Participants

Seven classes of German learners of Dutch (29 boys; 60 girls; M_{age} = 16.4, SD_{age} = 1.38) from five secondary schools in northwest Germany participated in this study. Schools were selected based on willingness to cooperate and the availability of students with at least an intermediate level of Dutch (i.e., B1 of the Common European Framework of Reference (CEFR); Council of Europe, 2011). Participants had learned Dutch in high school for 2 to 7 years (M_{Dutch} = 4.50, SD_{Dutch} = 1.48), where they had attended 3 to 5 hours of communicative Dutch language classes each week, with a focus on the four skills (i.e., reading, listening, writing, and speaking) and three form aspects (i.e., grammar, vocabulary, and pronunciation). All reported German as their L1 with three having an additional L1 (i.e., Swiss-German, Persian, and Albanian). All students participated voluntarily and signed for informed consent (themselves and, for minors, their primary caregiver).

5.3.2 Instruments

Background questionnaire

Before the intervention, participants completed a background questionnaire asking for information related to their gender, age, L1 and Dutch learning experience.

Proficiency measure

Participants' global target language proficiency was tested by the Dutch version of the LexTALE (Lemhöfer & Broersma, 2012), a web-based five-minute lexical decision task. LexTALE scores and the corresponding CEFR levels are only validated for English. We use them with care as a proxy for proficiency in the statistical analyses.

Anxiety measure

The FLCAS (Horwitz et al., 1986) is a standardized 33-item on a 5-point Likert scale. It has been validated and widely used in language learning research (Pérez Castillejo, 2019). A native speaker of German with a Master's degree in English translated the original English version into German. The current study operationalizes FLA as an emotional state that arises during an L2 speech task. All participants therefore completed the FLCAS after the speech production tasks at the beginning (T0) and end of the study (T3).

Speech production tasks

Four speech production tasks were designed to elicit spontaneous speech. Four complex, informal, and argumentative tasks (see supplementary data, Appendix J). These tasks were similar in structure but addressed different topics to allow for comparison. Task complexity was manipulated along the number of elements (Robinson, 2001) and personal concrete vs. abstract topics (de Jong et al., 2012). Participants were asked to role-play their performance as if they were in an authentic situation. Task versions were counterbalanced across participants and the four testing moments (T0 to T3).

5.3.3 Procedure

In a crossed treatment design (see Figure 1), five classes were involved, with half of the participants in each class randomly assigned to either treatment condition. Group 1 (N=31) or Fluency training+ Speaking lessons group (FT+SL group) started with the app-based fluency training lasting four weeks, which was followed by another four weeks of speaking lessons without a focus on fluency. Group 2 (N=33) or Speaking lessons + Fluency training (SL+FT group) engaged first in the speaking lessons and then app-based training focusing on fluency. The control group (N=25) attended regular classes without app training and participated in pre-/(delayed) post-tests only. Table 1 presents descriptive statistics for age, years of Dutch instruction, proficiency, and anxiety scores, disaggregated by group.

Table 1

Descriptive Statistics and One-Way ANOVA Results for Participant Characteristics by Group

Variable	**Group 1** N=31 Mean SD Range	**Group 2** N=33 Mean SD Range	**Control Group** N=25 Mean SD Range	**F**	***p* value**
Age (years)	16.4 (1.28) 15–19	16.3 (1.28) 15-19	16.7 (1.62) 14–20	0.75	0.47
Years of Dutch Instruction	5.23 (1.17) 4-7	5.12 (1.16) 4-7	4.04 (1.02) 3-5	9.08	<.001
Proficiency pretest score	57.90 (8.08) 43.75 -75	57.92 (8.04) 40 -72.5	52.35 (7.80) 35-70	4.35	0.02
Anxiety pretest score	2.19 (0.52) 1.36-3.66	2.22 (0.42) 1.36-3.97	2.36 (0.55) 1.36 -3.42	0.42	0.66

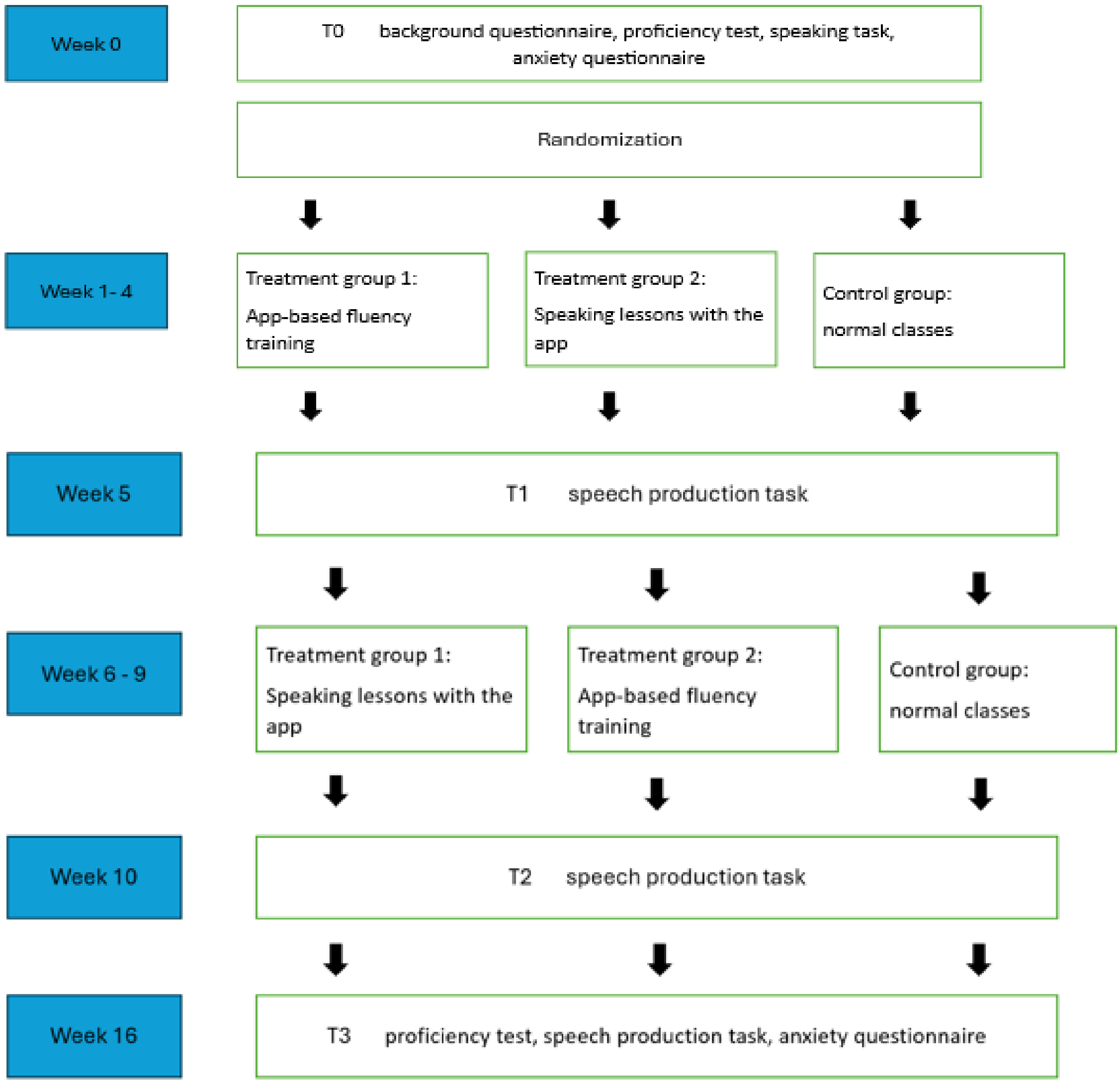
As shown in Table 1, the three groups were comparable in terms of age and anxiety at the start of the study. However, a one-way ANOVA revealed significant differences between groups in years of Dutch instruction ($p < .001$) and proficiency scores ($p = 0.02$), with the control group reporting fewer years of instruction and lower proficiency.
Classes 1 to 5 were randomly assigned to experimental conditions (Groups 1 and 2). Classes 6 and 7 formed the control group. See supplementary data, Appendix K, for the descriptive statistics of the classes. Significant differences between classes were observed for all background variables (Years of Dutch instruction, Anxiety, and Proficiency). To account for this variation, these variables were included as fixed effects in the statistical analyses to control for individual and contextual variability (see also 4.2).

The fluency training and speaking lessons were completed individually by students during their Dutch classes, twice a week for 15 minutes per session over a period of four weeks. During these sessions, the teacher's role was limited to classroom supervision and answering questions. According to the teachers, all students completed the full set of lessons. In the control group, speaking practice took place through pair or group work based on coursebook exercises, which were similar in topic, complexity, and length to those in the experimental groups. Each lesson included approximately 15 minutes of speaking practice, ensuring that the overall amount of practice time was equivalent across all groups[10].

Data was collected over a period of approximately five months and comprised three phases: pre-testing (T0), intervention with the *Reppen* app, post-testing (T1, T2) and delayed post-testing (T3). Between T1 and the start of the second instruction, three weeks passed because schools closed early for the two-week Christmas break due to the COVID-19 pandemic. Therefore, speech recordings could not be made for 26 of the 89 participants at T1. Seven of the 89 participants were absent during T2, and six at T3. In total, there were 317 speech recordings and 50 complete data sets.

[10] Naturally, we could not control for variation in the types and durations of speaking tasks during regular lessons.

Figure 1 *Design Overview*



5.3.4 Data analyses

A total of 317 speech recordings (from the speech production tasks at T0,1,2, and 3), with an average duration of slightly more than one minute (M_{length} = 77.28 s; SD_{length} = 32.55 s; $Range_{length}$ = 10.51 to 120.00 s), were analyzed and transcribed. Four recordings shorter than 10 seconds were excluded, as fluency measures would be unstable. Using PRAAT (Boersma & Weenink, 2016) we measured silent pauses, articulation rate, length of pauses, phonation time, syllables, and filled pauses (De Jong et al., 2021). Micropauses, that is, silent pauses smaller than or equal to 0.25 seconds, were excluded (Riggenbach, 1991). The automatically determined silent and filled pauses were manually checked. Through careful listening, we determined the pause location (i.e., mid- versus end-clause). We selected fluency measures based on findings from studies on L2 fluency training without an app (Tavakoli et al., 2016) and on research examining the relationship between anxiety and fluency (Pérez Castillejo, 2019). Tavakoli et al. (2016) found no significant improvement in repair measures. We therefore decided not to include them. This selection process resulted in the fluency measures summarized in Table 2.

Table 2

Fluency Measures and their Operationalization

Measure	Operationalization
Speed fluency	
Speech rate per minute	Total number of syllables divided by total sample time, multiplied by 60.
Articulation rate per minute	Total number of syllables divided by total speaking time (excluding pauses > 250 ms), multiplied by 60.
Breakdown fluency	
Phonation time ratio	Total speaking time divided by total sample time, multiplied by 100[11].
Frequency of mid-clause pauses per minute	Total number of mid-clause pauses (silent and filled) divided by total speaking time, multiplied by 60.
Frequency of end-clause pauses per minute	Total number of end-clause pauses (silent and filled) divided by total speaking time, multiplied by 60

A linguistically trained MA student reexamined 31 recordings (10% of the data) for syllable count, pause type (filled or silent), and pause location (mid- or end-clause), showing high interrater reliability with the main rater (first author) (96% for syllable count, 96% for pause type, and 90% for pause location, respectively).

[11] This measure is calculated as a percentage rather than per minute to better compare our data with similar studies (e.g., Tavakoli et al., 2016).

5.4 Results

After confirming normality (Shapiro-Wilk $p > .05$) and homogeneity of variances (Levene $p > .05$) multiple one-way ANOVAs were conducted to determine whether Groups 1, 2 and the control group were different in terms of fluency and anxiety at the beginning of the experiment. Results indicated no significant differences across the three groups in speech rate ($F(2,86) = 0.38$, $p = 0.69$), articulation rate ($F(2,86) = 0.40$, $p = 0.67$), phonation time ratio ($F(2,86) = 1.73$, $p = 0.18$), and frequency of end-clause pauses ($F(2,86) = 2.03$, $p = 0.14$). However, frequency of mid-clause pauses did reveal a significant difference ($F(2,86) = 4.45$, $p = 0.02$) suggesting that learners in the control group used more mid-clause pauses at the start of the intervention. We considered this to be only a minor indication of group differences, not violating the assumption of homogeneity. Furthermore, no differences in anxiety ($F(2,86) = 0.42$, $p = 0.66$) were detected.

Collinearity between the fluency measures during the four test times was investigated through Pearson's *r* correlations. High correlations were found between speech rate and articulation rate at T2 ($r = .877$) and T3 ($r = .855$). Correlations between other measures were either lower or negligible (see supplementary data, Appendix L). Speech rate and articulation rate were included in further analyses, acknowledging the potential risk of multicollinearity when interpreting the data.

We ran general linear models using the *lm()* function in R (R Core Team, 2020) to test for differences in fluency development and anxiety between the experimental and control groups. Model comparisons were assessed using the ANOVA function in R to test whether adding fixed factors significantly improved the model. Adjusted R^2 values are reported for the final models and interpreted following Plonsky and Ghanbar (2018); R^2 values below 0.20, between 0.20 and 0.50, and above 0.50 are interpreted as small, medium, and large, respectively, in terms of the explained variance they represent.

5.4.1 Descriptive statistics

Table 3 shows means and standard deviations for FLA, proficiency, and fluency measures. Participants' FLCAS scores ranged at T0 from 1.36 to 3.97 ($M_{anxiety} = 2.34$, $SD_{anxiety} = 0.54$); and at T3 from 1.18 to 3.94 ($M_{anxiety}$ = 2.16, $SD_{anxiety}$ = 0.58). According to Horwitz et al. (1986), averages below 3 indicate a low anxiety level. Participants' LEXTALE scores ranged at T0 from 35 to 75% ($M_{proficiency}$ = 56.35%, $SD_{proficiency}$ = 8.29), and at T3 from 38.75 to 78.75% ($M_{proficiency}$ = 58.13%, $SD_{proficiency}$ = 8.27). Overall, participants had a higher speech rate, articulation rate, and phonation time ratio at T3 than at T0, and used fewer mid- and end-clause pauses.

Table 3

Descriptive Statistics for All Variables at Time 0 – 3

		Group 1	**FT**	**SL**		**Group 2**	**SL**	**FT**		**Control Group**			
Time		T0	T1	T2	T3	T0	T1	T2	T3	T0	T1	T2	T3
Measure		*n* = 31	*n* = 22	*n* = 30	*n* = 29	*n* = 33	*n* = 21	*n* = 30	*n* = 29	*n* = 25	*n* = 20	*n* = 22	*n* = 25
FLCAS scores	M	2.19			2.08	2.22			2.08	2.36			2.29
	SD	(0.52)			(0.56)	(0.42)			(0.45)	(0.55)			(0.61)
LexTALE scores	M	57.90			59.22	57.92			60.58	52.35			53.90
	SD	(8.08)			(6.91)	(8.04)			(7.30)	(7.80)			(9.30)
Speech rate	M	140.96	166.89	169.21	169.49	134.34	153.47	160.80	162.73	137.66	142.89	147.54	158.50
	SD	(33.62)	(32.49)	(30.86)	(29.76)	(28.48)	(28.20)	(28.41)	(36.81)	(29.00)	(29.93)	(35.36)	(24.81)
Articulation rate	M	186.41	207.68	205.88	207.88	190.85	198.20	200.88	198.11	193.78	185.75	192.41	203.91
	SD	(32.31)	(30.18)	(31.33)	(25.11)	(26.94)	(22.77)	(26.11)	(36.93)	(34.47)	(24.74)	(33.87)	(25.84)
Phonation time ratio	M	74.94	80.16	82.13	81.26	70.06	77.27	79.83	81.64	71.63	76.61	76.30	77.70
	SD	(9.96)	(10.55)	(9.88)	(9.21)	(9.51)	(9.75)	(8.33)	(9.57)	(12.76)	(10.11)	(9.12)	(7.04)
Freq.mid-clause p.	M	18.68	18.06	16.97	17.77	22.59	20.34	20.11	18.80	24.74	20.13	22.95	22.91
	SD	(5.78)	(6.68)	(7.20)	(6.54)	(6.60)	(7.88)	(8.17)	(7.15)	(10.82)	(7.22)	(6.48)	(8.49)
Freq. end-clause p.	M	23.45	16.63	15.50	14.08	25.38	16.56	15.86	15.57	20.06	17.99	19.18	19.20
	SD	(10.77)	(6.56)	(5.64)	(5.57)	(8.77)	(6.13)	(5.00)	(5.65)	(10.47)	(6.97)	(8.23)	(9.89)

FT=fluency training; SL=speaking lessons; p. = pause

5.4.2 Research question 1: the effect of app-based training on L2 fluency

We conducted linear regression analyses to test whether the app-based training (fluency training and speaking lessons) in general led to changes in fluency at T2, and whether these changes persisted at T3 (research question 1). In the models, fluency indices at T0 were included as a covariate to enable focused comparisons between the groups. The following fixed factors were sequentially added to the models: Group (both experimental groups vs. control group), Class (Reference Class 1 vs. experimental Classes 2–5 and control group: Class 6–7), Anxiety (pretest score), Proficiency (pretest score), and Years of Dutch instruction. The best-fitting models of the fluency variables are presented below. (See supplementary data, Appendix M for the complete models).

Speech rate

As presented in Table 4, both models showed a significant difference between T0 to T2 (F(9, 74) = 7.5, $p < 0.001$, $R^2 = 0.45$) and T0 to T3 (F(8, 72) = 7.54, $p < 0.001$, $R^2 = 0.42$), with medium effect sizes (Plonsky & Ghanbar, 2018). Model T0 to T2 shows a significant effect of Group, indicating that the control group had lower overall speech rate scores than the two experimental groups at T2; however, this effect was not significant at T3. Additionally, Class was a significant predictor of speech rate from T0 to T2; participants in Classes 2, 6 and 7 had lower speech rates compared to Class 1, whereas no significant differences were found for Classes 3, 4 and 5 (see Appendix M for the complete models). Furthermore, higher anxiety was associated with a lower speech rate.

Table 4

Results of Linear Models Predicting Speech Rate

	T0 -T2				**T0-T3**			
Predictors	*Estimates (SE)*	*t value*	*p*	*R^2*	*Estimates (SE)*	*t value*	*p*	*R^2*
Intercept	159.489 (25.65)	6.217	< 0.001	0.45	132.865 (15.00)	8.854	< 0.001	0.42
Speech rate 0	0.344 (0.12)	2.842	0.005		0.268 (0.11)	2.489	0.015	
Control group	-20.658 (10.37)	-1.992	0.05		-10.679 (9.94)	-1.074	0.286	
Anxiety	-0.471 (0.22)	-2.173	0.03					
Class 2	-55.464 (13.48)	-4.114	< 0.001		-52.098 (11.24)	-4.636	< 0.001	
Class 6	-36.023 (10.70)	-3.366	0.001					
Class 7	-20.658 (10.37)	-1.992	0.05					

Articulation rate

As shown in Table 5, the models detected significant differences between T0 to T2 ($F(4, 78) = 4.08$, $p < 0.01$, $R^2 = 0.13$) and T0 to T3 ($F(4, 75) = 3.5$, $p < 0.05$, $R^2 = 0.11$), with small effect sizes. There was a significant effect of Group: the experimental groups had a higher articulation rate than the control group at T2, but this effect was not significant at T3.

Table 5

Results of Linear Models Predicting Articulation Rate

	T0 -T2				T0-T3			
Predictors	*Estimates (SE)*	*t value*	*p*	*R^2*	*Estimates (SE)*	*t value*	*p*	*R^2*
Intercept	164.447 (40.71)	4.040	< 0.001	0.13	111.472 (34.80)	3.204	0.001	0.11
Articulation rate 0	0.408 (0.13)	3.195	0.002		0.359 (0.10)	3.426	< 0.001	
Control group	-21.312 (8.97)	-2.377	0.019		-0.328 (7.38)	-0.044	0.964	
Anxiety	-0.270 (0.24)	-1.129	0.262		-0.047 (0.19)	-0.246	0.806	
Proficiency	-0.328 (0.50)	-0.651	0.517		0.476 (0.42)	1.132	0.261	

Phonation time ratio

As presented in Table 6, both models showed a significant difference between T0 to T2 (F(4, 78) = 11.11, p < 0.001, R^2 = 0.33) and T0 to T3 (F(9, 71) = 11.14, p < 0.001, R^2 = 0.51), with medium and large effect sizes, respectively. Group significantly predicted phonation time ratio, with the experimental groups showing higher phonation time ratio than the control group at both T2 and T3, suggesting the effect of the app-based training was retained over time. Anxiety was significant at both time points, with higher anxiety associated with lower phonation time ratio. Furthermore, Class was a significant predictor: Classes 2 and 7 had lower phonation time ratios than Class 1 at T3.

Table 6

Results of Linear Models Predicting Phonation Time Ratio

	T0 -T2				T0-T3			
Predictors	*Estimates (SE)*	*t value*	*p*	*R²*	*Estimates (SE)*	*t value*	*p*	*R²*
Intercept	51.816 (13.01)	3.981	< 0.001	0.33	90.267 (7.27)	12.412	< 0.001	0.51
Phonation time ratio 0	0.575 (0.11)	5.212	< 0.001		0.063 (0.07)	0.813	0.419	
Control group	-5.837 (2.67)	-2.183	0.032		-11.136 (2.50)	-4.451	< 0.001	
Anxiety	-0.151 (0.07)	-2.099	0.039		-0.134 (0.04)	-2.887	0.005	
Class 2					-21.439 (3.03)	-7.072	< 0.001	
Class 7					-11.136 (2.50)	-4.451	< 0.001	

Frequency of mid-clause pauses

Both models yielded significant differences between T0 to T2 ($F(8,74) = 4.09$, $p < 0.01$, $R^2 = 0.21$) and T0 to T3 ($F(8, 72) = 6.10$, $p < 0.01$, $R^2 = 0.31$), with medium effect sizes, showing a small reduction of number of mid-clause pauses. There was no significant effect of Group, indicating no clear difference between the groups. Furthermore, a significant effect of Class was found (see Table 7).

Table 7

Results of Linear Models Predicting Frequency of Mid-clause Pauses

	T0 -T2				T0-T3			
Predictors	*Estimates (SE)*	*t value*	*p*	*R²*	*Estimates (SE)*	*t value*	*p*	*R²*
Intercept	13.462 (2.48)	5.44	< 0.001	0.21	7.955 (2.24)	3.553	< 0.001	0.31
Mid-clause pauses 0	0.249 (0.09)	2.597	0.011		0.342 (0.08)	3.905	< 0.001	
Control group	2.282 (2.50)	0.913	0.364		2.022 (2.53)	0.800	0.426	
Class 2					6.555 (2.75)	2.382	0.019	
Class 4					5.928 (2.27)	2.608	0.011	
Class 6					9.474 (2.25)	4.213	< 0.001	

Frequency of end-clause pauses

Both models showed a significant difference between T0 to T2 (F (8,72) = 4.85, $p < 0.01$, R^2 = 0.25) and T0 to T3 (F (9,71) = 11.14, $p < 0.01$, R^2 =0.18), with medium and small effect sizes, respectively (Table 8). Group had a significant effect: the experimental groups used fewer end-clause pauses at both T2 and T3 than the control group, indicating the effect of the app-based training was retained over time. Class was also a significant predictor.

Table 8

Results of Linear Models Predicting Frequency of End-clause Pauses

	T0 -T2				T0-T3			
Predictors	*Estimates (SE)*	*t value*	*p*	*R^2*	*Estimates (SE)*	*t value*	*p*	*R^2*
Intercept	10.680 (2.25)	4.745	< 0.001	0.25	11.452 (4.50)	2.544	0.013	0.18
End-clause pauses 0	0.228 (0.07)	3.217	0.001		0.034 (0.08)	0.437	0.663	
Control group	6.550 (2.04)	3.210	0.001		9.303 (2.41)	3.863	< 0.001	
Class 2					6.337 (2.78)	2.280	0.025	
Class 7	6.550 (2.04)	2.040	0.002		9.303 (2.41)	3.863	0.0002	

5.4.3 Research question 2: the effect of fluency strategies on L2 fluency

To answer the second research question, we compared the two experimental groups (FT+SL and SL+FT) with each other. We conducted linear regression analyses to test whether the app-based fluency training led to changes in fluency variables at T1. Due to differences at T0, we included scores at T0 as a covariate to enable focused comparisons between the groups. The following fixed factors were sequentially added to the models: Group (FT+SL vs. SL+FT), Class, Anxiety (pretest score), Proficiency (pretest score) and Years of Dutch instruction. A significant effect was found only for articulation rate. Table 9 presents the results. The model detected significant differences between T0 and T1 (F (8,38) = 6.38, p < 0.001, R^2 = 0.34 (medium effect size). The FT+SL group made significantly more progress between T0 and T1 than the SL+FT group, indicating an effect of fluency training. See Appendix N (supplementary data) for the results of other fluency variables.

Table 9

Results of Linear Models Predicting Articulation Rate

	T0 - T1			
Predictors	*Estimate*	*t value*		*R^2*
Intercept	100.291 (38.35)	2.615	0.013	0.34
Articulation rate 0	0.5477 (0.13)	4.349	< 0.001	
SL+FT	-14.514 (6,87)	-2.111	0.041	
Anxiety	-0.1844 (0.24)	-0.779	0.440	
Proficiency	0.295 (0.45)	0.651	0.518	

5.4.4 Research question 3: the effect of app-based training on FLA

To evaluate the third research question, linear regression analyses were conducted to test whether the app-based training led to changes in anxiety at T3. Again, we included scores at T0 (Anxiety) as covariate. The following fixed factors were sequentially added to the models: Group (both experimental groups vs. control group) and Class. Table 10 presents the results. The model detected significant differences between T0 and T3 (F (8,75) = 30.34, p < 0.001, R^2 = 0.71 (large effect size). Anxiety was significantly lower at T3. However, there was no significant effect of Group on Anxiety, suggesting that using the app did not significantly impact anxiety. There was a significant effect of Class.

Table 10

Results of Linear Models Predicting Anxiety

	T0 -T3			
Predictors	*Estimate*	*t value*	*p*	R^2
Intercept	11.229 (6.53)	1.720	0.08	0.71
Anxiety0	0.860 (0.07)	11.621	< 0.001	
Control group	-1.114 (3.71)	-0.300	0.765	
Class 3	-12.521 (4.20)	-2.981	0.003	
Class 5	-9.545 (4.22)	-2.260	0.027	

5.5 Discussion

The aim of this study was to investigate the effects of an app-based intervention on the development of L2 fluency and FLA in a classroom setting. Research Question 1 asked to what extent an app-based intervention in general affects L2 fluency. Our data reveal that, compared to the control group, the experimental groups showed statistically significant improvements in speech and articulation rate from pretest to posttest (T2), as well as in phonation time ratio and frequency of end-clause pauses from pretest to posttest (T2), and six weeks after the intervention, during the delayed posttest (T3). In line with Swain's (1985) output hypothesis, it can be argued that creating opportunities for practice with the app has contributed to fluency development. Firstly, participants were able to work independently and at their own pace in the app, so the time spent producing a message was not limited by the presence or influence of other classmates or the teacher. Presumably, this allowed all students equal opportunities to practice, which may have contributed to fluency development. Secondly, thanks to ASR, participants received immediate individual feedback on speech rate, length, and number of pauses, and in case of a low score, the advice to use fluency strategies (e.g., fillers or circumvention), which may have subsequently contributed to improvements in fluency. These findings suggest that the app-based intervention had a positive short-term effect on multiple fluency measures. At the delayed posttest six weeks later, however, only phonation time ratio and end-clause pauses still showed significant effects. Given that end-clause pauses are generally linked to conceptual rather than language-specific processing (Segalowitz, 2010), they may provide only indirect evidence of L2 fluency development. In contrast, the improvement in phonation time ratio may reflect enhanced fluency (Tavakoli et al, 2016; De Jong & Perfetti, 2011). Overall, these results suggest preliminary evidence of sustained effects, underscoring the importance of further research. The control group, following the standard curriculum, developed much slower on fluency over five months. These results are partly consistent with previous research of Hanzawa (2021) who investigated fluency development, without any intervention, in classroom settings over an academic year. Their data showed minimal development in articulation rate and frequency of mid- and end-clause pauses – similar to our findings. Together, this suggest that simply following the standard curriculum does not improve fluency to the same degree as using app-based training. It seems that working with the app, which provides immediate individual feedback on fluency - something that is difficult for a teacher to achieve in a classroom with 20+ students - supports fluency development. In addition to the differences between the treatment and control groups, class-related factors (e.g., group dynamics, student motivation levels) may have played a role in the fluency outcomes. Specifically, Class was a significant predictor for several fluency measures, with participants from Class 2 (the app-based training group) and from Classes 6 and 7 (the control group) showing lower fluency scores compared to the reference group (Class 1). However, the current dataset does not allow us to speculate about the specific factors that might underlying this variability.

Research Question 2 investigated the extent to which app-based specific teaching of fluency strategies affect L2 fluency. Based on the findings of Tavakoli et al. (2016) and Seifoori and Vahidi (2012), we expected that the group starting with fluency training (FT+SL), would show more progress on fluency measures than the group starting with the speaking lessons (SL+FT group). The results indicated that only the findings for articulation rate supported this expectation. Previous research (Tavakoli et al., 2016) found improvements in articulation rate after strategy training combined with awareness-raising activities. Our results support these findings and contribute a new aspect: L2 fluency development is observed after app-based fluency strategy training, complementing Tavakoli's study that involved teacher-led lessons. Articulation rate (syllables/total speaking time without pauses) is a speed fluency measure which reflects the degree of automatization of linguistic knowledge and ease of access to it (Pérez Castillejo, 2019) and refers to the formulation/encoding aspect (Levelt, 1989). Formulation, especially lexical retrieval and grammatical encoding, are challenging aspects of speech processing in L2 (De Bot, 1992). Given that participants in our study had on average low to intermediate proficiency levels, their ability to formulate might still be limited. It is likely that the instruction on use of fluency strategies (e.g., formulaic sequences) reduced the time in silence while processing, which may have resulted in an increase in articulation rate.

Research Question 3 addressed to what extent app-based training affects FLA. Our results indicated no significant effect of app-based training on FLA. It might be that our findings are caused by the participants' overall relatively low levels of FLA, which were lower than those measured in earlier research (e.g., Perez Castillejo, 2019). Notably, two classes within the experimental group reported significantly lower anxiety scores compared to the reference class at T3. While, these lower scores may reflect the positive effects of app use, such as increased learner autonomy or less stress when speaking, it is important to note that no overall group effect was observed. Therefore, other so far unknown factors may have played a role. Our study calls for further research on this matter.

Given that this study was performed in a high school setting, which enhances its ecological validity, and draws on authentic data, there are also several limitations. We will highlight the following three. To the best of our knowledge, this is the first study of its kind that draws on app-based fluency training. The specific technical innovations, drawing on ASR, caused the app to not always provide fully accurate analyses. As reported during debriefing, some participants experienced that the speech analysis did not always accurately match their speech recording. For example, the speech analysis indicated that there were many long pauses, when that was not the case. These are indeed unfortunate glitches, but with the rapid advancements in technological innovation, including ASR, it is likely that this will improve in the near future. An important direction for future research would be to conduct an empirical validation of the automated speech measures used in the fluency training. This could involve comparing the app-generated scores with independent human ratings to examine measurement reliability.

Secondly, due to COVID-19, we experienced some data loss, which might have impacted the results. On the other hand, self-paced learning with an app worked particularly well during the pandemic. Thirdly, the timing of the FLA measurements—taken at T0 and T3, six weeks after the training ended—limits the ability to attribute changes in FLA directly to the intervention, as other factors may have influenced the outcomes. Future research should include a posttest immediately after the training to better assess its immediate effects.

5.6 Conclusions and implications for teaching

This study introduced several innovations by investigating the use of ASR within an app to train fluency in the context of an underexplored setting (learning Dutch by German high school students). The findings indicated that app-based training led to enhanced speech and articulation rate, a higher phonation time ratio, and a reduced frequency of end-clause pauses. The results further indicated that specific app-based fluency training resulted in a significant improvement in articulation rate. However, the app-based training did not affect FLA.

We finish the paper by highlighting some pedagogical notes. First, the findings support the integration of app-based training into regular classes. The app offers L2 learners the opportunity for individual fluency practice at their own pace, at a time and place convenient for them, both within and outside the classroom, using the devices they already have and use daily. A second implication is that the app provides immediate, individual feedback on fluency, something that is difficult for a teacher to achieve, as large classes and the transient nature of speaking make it challenging to give direct feedback to every student. In this way, a digital tool can support both language teaching and language development.

6. General Discussion

6.1 Introduction

The purpose of this dissertation was to study the role of FLA on L2 fluency and the effect of app-based intervention on L2 fluency and FLA. FLA is examined from two perspectives: first, whether FLA leads to reduced fluency (Chapter 4), and second, whether an app-based training can reduce FLA (Chapter 5).

In this dissertation, four main research questions have been addressed:

1. To what extent is there a relationship between speed, breakdown and repair fluency measures and FLA and proficiency, and how does this interact with task complexity?
2. To what extent does an app-based intervention in general affect L2 fluency?
3. To what extent does specific app-based training of fluency strategies affect L2 fluency?
4. To what extent does app-based intervention in general affect FLA?

The main findings for each research question are summarized and discussed in Section 6.2; in 6.2.1 the relationship between FLA and L2 fluency, and in 6.2.2 effects of app-based intervention on L2 fluency and FLA. Section 6.3 discusses methodological challenges and choices in app-based research, followed by addressing theoretical and pedagogical implications in Section 6.4. Section 6.5 offers suggestions for further research, and the chapter concludes with a conclusion.

6.2 Summary of results

6.2.1 The relationship between FLA and L2 fluency

Chapter 4 adopted a psycholinguistic approach to L2 fluency and investigated which aspects of L2 fluency are related to FLA and proficiency among low-intermediate L2 learners and how this interacts with task complexity. So far, mainly the relationship between FLA and L2 fluency in less cognitively demanding tasks has been investigated. To gain a more comprehensive understanding of how FLA relates to L2 fluency in both simple and complex tasks, we used a broad range of fluency measures based on Skehan's (2009) classification of L2 utterance features into speed, breakdown, and repair fluency. Additionally, this study was intended as a preliminary study for the app intervention (see Chapter 5).

Based on Processing Efficiency Theory (Eysenck & Calvo, 1992), Attention Control Theory (Eysenck, 2010; Eysenck et al., 2007), Cognition Hypothesis (Robinson, 2001) and previous research (MacIntyre & Gardner, 1994; Pérez Castillejo, 2019, 2021; Sanaei et al., 2015; Bielak, 2022), we hypothesized that a cognitively more demanding task will negatively affect speed fluency of anxious learners, requires more pausing behavior, and will lead to fewer self-corrections and reformulations. To address the research question, German learners of Dutch completed the FLCAS questionnaire, a general proficiency test (LexTALE), and two speech production tasks (a simple task and a complex task) during a non-exam situation.

The results of the correlation analyses indicated that proficiency was positively related to speech rate and phonation time ratio, and negatively related to number of silent and mid-clause pauses, as well as number of pauses in the simple task. FLA had a negative impact on number of mid-clause pauses in the complex task, but with a small effect size. This result supports the theoretical claims regarding the negative effect of FLA in L2 speech processing in a cognitively demanding task. The findings are also in line with previous research (e.g., Pérez Castillejo 2019, 2021) that FLA primarily leads to issues in encoding and formulation rather than in conceptualization.

The results of the regression analyses showed that proficiency predicted number of mid-clause pauses and number of pauses, while task type predicted speech rate, and anxiety was not a significant predictor of any fluency measure. In Pérez Castillejo's studies (2019, 2021), FLA was a significant predictor of the fluency measures analyzed, while proficiency was not. Explanations for the contradictory results could be that in our study, the tasks were not performed during an exam situation, and the participants' anxiety levels were lower compared to those measured in previous research (Bielak, 2022; Pérez Castillejo, 2019, 2021; Saito et al., 2018).

Below, I discuss two unexpected results. Firstly, we expected a stronger effect of FLA on fluency in the cognitively demanding task, in line with Cognition Hypothesis, Processing Efficiency Theory, and Attentional Control Theory. When performing a cognitively demanding task, the need to meet the cognitive demands of a complex task may require increased attention to both conceptual planning of the message content and linguistic encoding. Regarding linguistic encoding, a complex task leads learners to use more elaborate and varied lexis and produce more complex linguistic structures. Anxiety may negatively affect performance by interfering with cognitive processes, such as encoding and formulation, which require attention. One explanation for the lack of a stronger effect of FLA in the complex task may be related to participants' generally low anxiety level, which could mean that the cognitive interference of anxiety was not strong enough to have an effect on L2 speech processing.

Another explanation could be that the complex task may not have been complex enough. The differences in task complexity between the simple and complex tasks, operationalized by the number of conceptual elements (Robinson, 2001) and personal concrete versus abstract topics (de Jong et al., 2012), apparently did not impose a high processing load. Therefore, L2 speech processing during the complex task may have been less vulnerable to anxiety interference.

A third explanation could be that participants had the opportunity to process L2 linguistic material in the simple task before engaging in the complex task. Note that the tasks were not counterbalanced but were always presented in the same order. It could be argued that performance that happens second in a sequence of tasks may still benefit from language processing during the first task. According to McDonough and Trofimovich (2008), the earlier activation of linguistic material may facilitate later access to and use of related linguistic knowledge by freeing up attentional resources. This gain in speech processing efficiency may explain why anxiety did not impact performance.

The second unexpected result was a significant positive effect of the predictor task type on speech rate; the more complex the task, the higher the speech rate. This is contrary to the predictions of the Cognitive Hypothesis (Robinson, 2001) and the Limited Attentional Capacity Model (Skehan & Foster, 2001), which suggest that such tasks might negatively affect fluency.

Here, the previous use of language during the simple task could also explain why performance in the complex task was more fluent.

Another explanation might be that the topic of the complex task, discussing the use of mobile phones in class - a widely debated topic in schools - may be more interesting for high school students than the simple task in which they had to talk about what they did over the weekend. The descriptive statistics seem to reflect this (see Section 4.4), indicating that the complex task resulted in more phonation time. Furthermore, the descriptive statistics showed that the simple task led to more and longer end-clause pauses than the complex task. These pauses are usually associated with conceptualization (de Jong, 2016). You would expect that a complex task requires more attention from the speaker to conceptualize the message than a simple task. Apparently, the simple task of talking about weekend activities led to more pause behavior. This can be explained by the fact that, during the data collection phase, it was not possible from an organizational point of view for all participants to perform the speech production tasks immediately after the weekend. As a result, it is conceivable that when participants have to recount on Tuesday what they did during the weekend, there might be more and longer end-clause pauses, partly due to the effort of remembering what happened during the weekend.

Finally, participants' generally low anxiety levels in this study could be explained by the fact that they live relatively close to the Dutch border and that Dutch is a related language to German (participants' L1). Firstly, living close to the Dutch border likely provides German learners of Dutch with several advantages; there are opportunities for real-life exposure to Dutch through interactions with native speakers, cultural exchange and school exchange, and access to Dutch media. This frequent exposure might make the language feel more familiar and may reduce anxiety about speaking it. Secondly, it may have played a role that teachers who participated in the study indicated that their lessons are communicative rather than grammar oriented. This focus is also consistent with the *Kerncurriculum für das Fach Niederländisch* (2017) and is reflected in the teaching materials used, such as *Welkom in de klas* (2019). Such a communicative approach may also influence students' speaking anxiety, potentially reducing their fear of making mistakes while encouraging more spontaneous language production. Thirdly, German and Dutch belong to the Germanic language family.

Perhaps speaking a language related to one's L1 causes less anxiety compared to speaking an unrelated language like French. These explanations are speculative and require further research.

6.2.2 Effects of app-based intervention on L2 fluency and FLA

The results of the first study (Chapter 4), which showed that FLA correlates with the frequency of mid-clause pauses, together with earlier research on the relationship between FLA and fluency measures (e.g., Pérez Castillejo, 2019, 2021; Sanaei et al., 2015), guided the development of app-based fluency training (Chapter 3). Results from these studies indicated negative relationships between FLA and fluency measures that reflect speech formulation and encoding processes. More anxious learners showed a lower phonation time ratio, shorter runs between pauses, and more frequent mid-clause pauses. In addition, findings from studies on L2 fluency training without the use of an app were considered (Tavakoli et al., 2016; Seifoori & Vahidi, 2012). These studies suggest that strategy instruction may be an effective way to enhance fluency. For this reason, an app-based training was developed to teach fluency strategies. This training is hypothesized to improve fluency by equipping learners with strategies that help compensate for potential losses in fluency during speaking tasks. In addition, the app provides feedback via ASR on both speed fluency and breakdown fluency. For this purpose, speech rate and phonation time ratio were selected for the automatic fluency analysis, based on findings from L2 fluency training studies without an app (Tavakoli et al., 2016), research on the relationship between FLA and fluency (Pérez Castillejo, 2019), and practical considerations, as these measures could be reliably analysed by the software. The software used did not allow for analysis of pause location, so specific analysis of mid-clause or end-clause pauses was not possible.

A pedagogical approach to L2 fluency was adopted in Chapter 5. German learners of Dutch, divided into three groups, participated in this study. Group 1 engaged in app-based fluency training, followed by app-based speaking lessons without a specific focus on fluency; group 2 started with app-based speaking lessons and then engaged in app-based fluency training focusing on fluency; while the control group 3 attended regular classes with no app-based training. Oral data were analyzed for various measures of utterance fluency. We selected these measures (speech rate, articulation rate, phonation time ratio, frequency of mid-clause and end-clause pauses) based on previous research. Studies on L2 fluency training without an app reported improvements in speech rate, articulation rate, phonation time ratio, and end-clause pauses after training (Tavakoli et al., 2016; Seifoori & Vahidi, 2012), whereas research on the relationship between anxiety and fluency indicated negative relationships between FLA and phonation time ratio and frequency of mid-clause pauses (Sanaei et al., 2015; Pérez Castillejo, 2019, 2021).

The first research question investigated to what extent the app-based intervention in general influences L2 fluency. Given that the app provides opportunities for oral output and based on Swain's (1985) Output Hypothesis, we expected that using the app would result in increased fluency. This question was addressed by implementing an app-based intervention. Using a pretest-posttest-delayed posttest design, data were collected over a period of five months. Findings indicated that both experimental groups showed significant progress in speech and articulation rate from pretest to posttest (T2), as well as in phonation time ratio and number of end-clause pauses from pre- to post-delay test. The control group showed slower progress in fluency over the five-month period. These findings suggest that providing opportunities for practice with the app contributes to fluency development. Below, these results will be further discussed in light of the literature on TELL/app-based training, fluency development and L2 speech production process.

App-based training and fluency development

Previous research has already shown that the use of TELL has a positive impact on the development of L2 fluency, focusing on peer-to-peer conversational activities (Tecedor & Campos-Dintrans, 2018; Wang et al., 2018; Grimshaw & Cardoso, 2018). The positive effect of using TELL on the development of L2 fluency, as reported in the literature, is extended by our results to individual app-based training. One explanation for the development of L2 fluency is that <u>all students</u> were able to practice speaking at their own pace with the app, providing equal opportunities for everyone to speak. This contrasts with classroom speaking instruction, where often only a few students speak while the rest listen. The latter is confirmed in the debriefing questionnaire (Chapter 3.3), where participants indicated that the app allowed them to speak a lot of Dutch in succession, something they felt does not often happen in regular classes. In addition, the app may have created a safe environment where no one is listening in, potentially making students more comfortable speaking in the L2, which may have encouraged them to speak more in the L2.

Furthermore, all students received immediate individualized speech analyses and instructional feedback via ASR. In contrast, classroom speaking instruction often adopts a one-size-fits-all approach, addressing general challenges without personalized feedback due to class size and the transient nature of speech. According to the debriefing questionnaire (see Chapter 3.3), students appreciated the individual feedback, which may have contributed to increased L2 speaking and consequently the development of L2 fluency.

Output hypothesis and fluency development

L2 learners often possess less linguistic knowledge and slow speed of access to that knowledge compared to L1 speakers. As a result, for L2 speakers, two of the three processes—formulating and articulating—are not as automatized as they are for L1 speakers. Consequently, these processes demand a lot of working memory capacity and often cannot occur in parallel with other speaking processes. This results in disfluencies such as a slower speech rate and more pauses than L1 speakers (e.g., Kahng, 2014; Riazantseva, 2001). To improve fluency, it is necessary to increase the linguistic knowledge of L2 speakers and make that knowledge as accessible as possible. In the literature, it is assumed that speaking practice allows more automatized access to linguistic knowledge (Kormos, 2006). Swain's (1985) output hypothesis also suggests that students learn through output (i.e., oral production) and production practice in their L2, as they need to experiment with their hypotheses about the language and its structures, to see how the language functions and to learn through trial-and-error. Output is therefore crucial for fluency development (Swain, 1985; Nation & Newton, 2008).

Our results support the predictions of Swain's (1985) output hypothesis in that the 32 speaking tasks in the Reppen app (16 speaking tasks in fluency training and 16 in TBI speaking lessons) apparently provided students with sufficient opportunities to speak in the L2 and develop fluency. In addition, in the app, students were more or less ‘forced’ to produce output, as they could not skip speaking tasks and scroll through the app without speaking. During speaking tasks, they usually had to speak for at least 30 seconds before the app could provide speech analysis and feedback. These features of the app may have contributed to the development of fluency.

L2 speech production and fluency development

The results showed that app-based training led to fluency gains in speech rate, articulation rate, phonation time ratio, and number of end-clause pauses. Speech rate (i.e., number of syllables divided by sample time, including pauses) is likely to reflect cognitive efficiency in carrying out all aspects of L2 speech production (Levelt, 1989,1999): content planning, linguistic encoding and articulation (Towell et al. 1996; Götz, 2013; Kormos, 2006). Increases in speech rate are often seen as an indication of more automatic speech processing. Articulation rate is discussed below. Phonation time ratio was calculated as percentage of time spent speaking and pausing. Towell et al. (1996) argued that the proportion of speaking time is related to automaticity in speech processing because a higher phonation time ratio signals a reduced need to pause for planning or formulating a message. In addition, the results of the present study indicate a reduction in end-clause pauses. Previous research has shown that end-clause pausing is used by speakers for content planning, and activating background knowledge (Butterworth, 1975; Götz, 2013) Hence, it is closely connected with the conceptualization stage of L2 speech production (Levelt, 1989, 1999). The participants in the study had, on average, low to intermediate proficiency levels. For L2 speakers, especially at lower proficiency levels, formulation and encoding may exert greater demands on attention control than message conceptualization (De Bot 1992; Kormos 2006), because retrieving appropriate words from the mental lexicon and encoding syntactic structures in the L2 might require more cognitive effort than planning what to say. Through speaking practice with the app, it might be that linguistic knowledge became accessible in a more automatic way, thereby requiring fewer attentional resources for formulation and encoding (resulting in a longer phonation time ratio and higher speech rate). It is possible that due to this increased cognitive efficiency, more attentional resources became available for content planning, which subsequently proceeded faster and resulted in fewer end-clause pauses.

The second research question examined the effect of app-based instruction of fluency strategies on L2 fluency. Based on Swain (1985), Tavakoli et al. (2016) and Seifoori and Vahidi (2012), we expected that both types of instruction - fluency training and task-based speaking lessons - could improve participants' fluency, but that participants who started with fluency training would show more gains in fluency compared to those who began with task-based speaking lessons. This is because participants who start with fluency instruction are better equipped to complete speaking tasks and to compensate for any loss in fluency than the group starting with task-based speaking lessons. The results for articulation rate supported our expectations; the group that began with fluency training made significantly more progress in articulation rate.

Why did fluency training succeed in increasing articulation rate? Participants in this study generally had low to intermediate language proficiency levels, which may mean that they have incomplete linguistic knowledge and slow access to that knowledge, making formulation less automatized. As a result, formulation requires a lot of working memory capacity and often cannot be performed in parallel with other speech processes. Articulation rate (syllables/total speaking time without pauses) is a speed fluency measure which reflects the degree of automatization of linguistic knowledge and ease of access to it (Pérez Castillejo, 2019, 2021) and refers to the formulation/encoding aspect of cognitive fluency (Levelt, 1989). One explanation for the increase in articulation rate might be the use of fluency strategies (such as formulaic sequences), which helped learners to avoid silent pauses and keep the communication channel open during difficulties (Dörnyei & Kormos, 1998). The improvement in articulation rate suggests that the use of fluency strategies enhanced the efficiency of the formulation processes.

The third research question examined to what extent app-based intervention in general affects FLA. To address this research question, participants completed the Foreign Language Classroom Anxiety Scale (the FLCAS; Horwitz et al. 1986) and a speech production task during both the pretest and delay posttest. Oral data were analyzed for various measures of fluency (e.g., frequency of mid-clause pauses and phonation time ratio). We used frequency of mid-clause pauses because previous research (Pérez Castillejo, 2019, 2021), as well as our own findings (Chapter 4), indicate that FLA correlates with frequency of mid-clause pauses. This suggests that learners with higher anxiety tend to use more mid-clause pauses. Additionally, we used the measure of phonation time ratio, as research (Sanaei et al., 2015; Pérez Castillejo, 2019, 2021) has found a link between phonation time ratio and FLA, suggesting that higher anxiety levels are associated with less phonation time. The findings showed that the app-based intervention did not significantly reduce FLA. One possible explanation for not finding an effect in the current study could be the relatively low levels of FLA among the participants.

In the current study, task-specific anxiety was measured with the FLCAS (Horwitz et al., 1986). Perhaps it would have been better to use a more task-specific questionnaire on FLA instead of the FLCAS, which determines general anxiety levels as experienced in the language classroom. The FLCAS was chosen due to its "consistently high reliability" (MacIntyre, 2017: 16) and its widespread use in research on FLA. It consists of 33 Likert scale items (ranging from 1 – I strongly disagree to 5 – I strongly agree) that assess anxiety in language learning situations, such as "I am afraid that my language teacher is ready to correct every mistake I make." A task-specific FLA measure could include questions like "How anxious were you while performing the task in Dutch?" and "What factors contributed to your anxiety during this session?" Future research could replicate the study using a different task-specific FLA measure.

6.3 Methodological challenges and choices in app-based research

In this section, I describe some challenges we encountered during the development and implementation of the app, so that future research may potentially benefit from this. The first challenge was finding funders to develop the app. Thanks to grants from the Taalunie and the Ems Dollard Region, the app could be developed. The development costs of the app and maintenance during the intervention amounted to € 38.000, which is low (Lastovetska, 2024). From experience, we have learned that finding sponsors takes a lot of time, in our case more than a year, so researchers should take this into account in their planning. In the budget, no money was reserved to pay for the annual licenses and maintenance of the app after the research was completed. We should have taken this into account. Consequently, the app is unfortunately offline at the moment.

The second challenge was the design of the app-based training, which was based on both evidence and expectations. For the content, I used research evidence regarding fluency training (Tavakoli et al., 2016; Seifoori & Vahidi, 2012), CALLA model for strategy instruction (Chamot, 2009), design of task-based instruction (Ellis, 2003; Willis, 1996), supplemented with my own knowledge and experiences as a language teacher, teacher educator, and author of various language courses (e.g., Reitsma, 2018; Reitsma, 2022), as well as input from language teachers, students, and specialists in information technology. The technological features were designed based on expectations of effectiveness: a colorful design, illustrations, and score visuals to motivate students to engage with the app. Whether this has actually been effective remains unknown; It has not been investigated to what extent the design and activities supported by technology contributed to increased student motivation and learning outcomes. Therefore, future research should also include an evaluation to determine whether the expectations regarding the positive impact of these technological features are indeed supported. This knowledge can benefit future app-based research.

A third challenge was of technological nature. During the intervention, there were occasional issues with Wi-Fi connections in schools, causing the app to freeze and speech recordings not to be captured, resulting in tasks needing to be redone. This caused annoyance among the participants. However, the main technological challenge lay elsewhere, namely, in the speech analyses, which were not always precise enough. For example, the speech analysis sometimes indicated that there were many pauses and long pauses in the speech recordings, while this was not the case. It should be noted that to our knowledge, the *Reppen* app is the first app to use ASR for the development of L2 fluency, so we could not rely on knowledge from previous studies. Azure Cognitive Services were used for the implementation of speech analysis within the Reppen app. This software allows speech to be converted into text. These speech analyses are based on a pre-trained model, where it should be noted that the more data available to the model, the more accurate the analysis. It can be assumed that these speech analyses for a smaller language like Dutch are less developed than for major languages such as English or Spanish. As a result, it is possible that the speech analyses for Dutch are less precise. Furthermore, the input of the pre-trained model is speech from native speakers of Dutch to create 'rules' for analyzing data. In my research, the participants were L2 learners of Dutch who often still struggled with Dutch pronunciation, so it is possible that words sometimes sounded different than expected by the system, which may have had consequences for speech recognition and ultimately for the feedback received by the participants regarding speech tempo (calculated as words per minute), where certain words may not have been recognized, resulting in a lower score than the participant expected. Moreover, the quality of the recording affects the speech analysis. When a recording is made in a noisy environment, it is difficult for the system to make an accurate speech analysis. If speech analyses are not accurate and thus the scores and feedback received by the participants via the app do not match reality, this can lead to less confidence in the analyses and the app. Therefore, technological improvement of speech analysis software, particularly for Dutch, is very important for the further development of apps.

A next step in developing the *Reppen* app would be to offer different learning routes for learners, i.e., more differentiation in instruction, activities and feedback. One possibility could be for learners to take a test before training begins, assessing their fluency level through speaking tasks and speech analysis in the app. After this, learners can follow a learning path tailored to their needs. For learners who struggle with fluency, this could mean additional instructions and engaging in more speaking tasks.

Finally, the *Reppen* app was developed to improve fluency in Dutch. Fluency training, which involves increasing awareness of fluency, teaching fluency strategies, and applying them in speaking tasks, is not only important for L2 learners of Dutch but also for learners of other foreign languages. With suitable ASR software, the training could also be adapted to other foreign languages in the future.

6.4 Theoretical and pedagogical Implications

The findings regarding the relationship between FLA and L2 fluency, as well as app-based intervention on L2 fluency and FLA, have both theoretical and pedagogical implications. These will be discussed in the following sections.

6.4.1 The relationship between FLA and L2 fluency

The findings from Chapter 4 contribute to our understanding of the relationship between FLA and L2 fluency. More specifically, they add to our knowledge of (1) in what way FLA affects L2 fluency in complex tasks and (2) the predictive power of FLA and proficiency on L2 fluency during a non-exam situation.

Our results inform scholars in the field of speech production about the effect of FLA on L2 fluency and the speech production processes in complex tasks. Previous studies (Pérez Castillejo, 2019, 2021; Sanaei et al., 2015) had already demonstrated that FLA interfered with speech processing hindering formulation and encoding more than conceptualization in simple tasks among low-intermediate L2 learners. Bielak (2022) demonstrated that FLA hinders speech processing during conceptualization in complex tasks among proficient L2 learners. The study reported in Chapter 4 adds that heightened anxiety can lead to issues during encoding and formulation in cognitively demanding tasks, rather than during conceptualization.

Our results also show that FLA was not a significant predictor of any fluency measure. Although proficiency was a significant predictor for two fluency measures, the outcomes of the regression models (see 4.4) indicate that the fluency measures examined here are not adequately predicted by L2 proficiency or FLA alone. Dewaele (2002) suggests, there are other cognitive variables (e.g., working memory efficiency or short-term memory capacity) that may contribute to the variation in L2 fluency. It would be interesting to further investigate if and how these factors contribute to the variation in fluency measures.

These insights from Chapter 4 can inform researchers, as well as teachers. Firstly, the finding that FLA may result in more frequent mid-clause pauses can help teachers understand the underlying reasons for the fluency profiles of L2 learners. Considering the frequent use of mid-clause pauses as a strategy of anxious L2 learners instead of disfluency may be a reason to introduce fluency strategies (Rood & de Jong, 2023). For example, if learners tend to fall silent, they can work on fluency strategies such as circumlocution or the use of formulaic sequences. Secondly, the results suggest that FLA may not be such a confounding variable during speaking tasks that are not part of formal tests. This could provide a reason for teachers to base grades for speaking skills also on classroom observations, rather than only on formal tests (Bielak, 2022).

6.4.2 Implications for app-based training on L2 fluency and FLA

The findings of Chapter 5 contribute to our understanding of app-based interventions on L2 fluency. Specifically, they enhance our knowledge of the effects of (1) app-based training on L2 fluency and (2) app-based fluency strategy training on L2 fluency.

App-based training on L2 fluency

Our results inform scholars in the field of L2 fluency about the effect of app-based training on L2 fluency. Previous studies had already shown that using TELL impacts L2 fluency development through conversation activities (Tecedor & Campos-Dintrans, 2019; Wang et al., 2018; Grimshaw & Cardoso, 2018). Our findings in Chapter 5 demonstrate that L2 fluency can also improve through individual practice using an app. This insight has important pedagogical implications.

The first implication is that the *Reppen* app offers L2 learners the opportunity for individual fluency practice at their own pace, at a time and place convenient for them, both within and outside the classroom, without classmates listening in. This means that with an app, all students have the opportunity to actively practice their speaking skills, instead of just a few students speaking while the rest listen, as often happens during language lessons. This was confirmed by participants in the debriefing questionnaire (see Chapter 3.3) that the app allowed them to speak a lot of Dutch in succession, as opposed to in regular classes.

Another implication is that using an app creates new pedagogical opportunities to broaden the reach of the classroom and consequently minimize one of the challenges of classroom teaching, namely time (Bione & Cardoso, 2020). This is particularly important for speaking, which requires a lot of practice (Everly, 2018), so that learners can test their hypotheses about what they are learning and consequently automatize their skills (Grimshaw & Cardoso, 2018). By expanding the classroom's reach, teachers can spend their classroom time on activities that may require human support e.g., feedback (Cardoso, 2022).

App-based fluency strategy training and L2 fluency

Previous studies had already demonstrated that learners can enhance L2 fluency through teacher-led strategy training (Tavakoli et al., 2016). Our findings indicate that learners can also improve fluency through fluency instruction via an app, which is an important contribution to the field of L2 fluency. Additionally, we have introduced an innovation, namely, providing immediate feedback on L2 fluency via ASR.

These results have important pedagogical implications. Firstly, the *Reppen* app offers a fluency training program that can be easily integrated into a curriculum for L2 learners from level A2 and above. It is theme-independent, making it easily applicable in language practice. Furthermore, this fluency training may also be of interest to teacher education, as L2 fluency and L2 fluency training have thus far received little attention in both teaching materials and teaching practices (Rossiter et al., 2010; Tavakoli & Hunter, 2018; see also Ch 2.3).

A second implication of using the *Reppen* app is that students, through explicit strategy instruction, develop metacognitive ability to identify their fluency issues, apply strategies to solve fluency problems, and evaluate whether their application was successful. Thus, working with the app helps students develop their metacognitive ability and thereby promotes their development into effective and independent learners (Chamot & Harris, 2019).

A third implication is that the *Reppen* app offers immediate individual feedback on fluency, which is challenging for a teacher to achieve due to large classes and the transient nature of speaking, making it difficult to give direct feedback to every student.

Finally, it is worth noting that during presentations about the study from Chapter 5, several teachers indicated they were inspired and interested in using the app in their language classes. Therefore, it appears that there is both interest and demand for the app.

6.5 Further research

This dissertation focused on the relationship between FLA and L2 fluency and app-based training of L2 fluency. In previous sections, some ideas for future research have been proposed already. In this section, two other interesting research topics are discussed.

6.5.1 Approaches to FLA

In the study from Chapter 4, FLA served as a measure of situation-specific anxiety experienced in the language classroom, measured several weeks before the performance of speech production tasks. It is possible that participants' anxiety increased during speech production tasks, especially for those learners who already had high situation-specific anxiety (Dewaele, 2002). Whether the relationship between fluency and the more situation-specific approach to FLA, in our study, differs from a more transient, dynamic approach to FLA (MacIntyre, 2012; Gregersen, MacIntyre, & Meza, 2014) could not be examined. In future research, instruments that measure more situation-specific experiences of anxiety could be used along with transient, task-specific measure of anxiety to obtain a more complete emotional portrait of participants and learning situations, allowing for more nuanced analyses.

6.5.2 Dialogue-based fluency training via an app

A follow-up study building on the results of the current study could focus on dialogue-based fluency training via an app. As one teacher noted in the debriefing questionnaire, after practicing individually with app-based training, students should engage more actively in conversations with dialogue partners. This follow-up study could investigate whether using an app where students learn fluency strategies and practice through dialogues leads to improvements in their fluency.

6.6 Conclusions

This dissertation investigated the relationship between FLA and L2 fluency and the effect of app-based training on L2 fluency development. The Reppen app was developed specifically for this dissertation with the goal of training L2 fluency using ASR. The latter is, to our knowledge, innovative in fluency development research.

The research has shown that FLA was negatively related to number of mid-clause pauses in a complex task. Furthermore, this dissertation has shown that app-based training in general leads to enhanced speech rate, articulation rate, phonation time ratio, and a reduced frequency of end-clause pauses. Additionally, app-based fluency training results in a significant improvement in articulation rate. An important implication of this research is that, since the control group, which followed regular classes, only showed slow improvement in L2 fluency over a period of five months, this research provides evidence for integrating app-based training into regular classes.

References

Abitzsch, D., & Sudhoff, S. (2019). *Welkom in de klas! 2*: Lesboek met luisterteksten voor smartphone/tablet. Klett.

Ahmadian, M. J. (2011). The effect of 'massed' task repetitions on complexity, accuracy and fluency: Does it transfer to a new task? *The Language Learning Journal, 39*(3), 269–280.

American Council on the Teaching of Foreign Languages (ACTFL). (2013). *The ACTFL proficiency guidelines 2012 – Speaking.* http://actflproficiencyguidelines2012.org/speaking

Anderson, N.J. 2003. Metacognitive Reading Strategies Increase L2 Performance. *The Language Teacher 27*, 20–22.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*(10), 829–839. https://doi.org/10.1038/nrn1201

Baran-Lucarz, M. (2022). Language anxiety. In T. M. Derwing (Ed.), *The Routledge handbook of second language acquisition and speaking* (pp. 83–96). Routledge.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Bielak, J. (2022). To what extent are foreign language anxiety and foreign language enjoyment related to L2 fluency? An investigation of task-specific emotions and breakdown and speed fluency in an oral task. *Language Teaching Research.*

Bione, T., & Cardoso, W. (2020). Synthetic voices in the foreign language context. *Language Learning & Technology, 24*(1), 169–186.

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research, 10*(3), 245–261. https://doi.org/10.1191/1362168806lr195oa

Boersma, P., & Weenink, D**.** (2016). PRAAT: Doing phonetics by computer [Computer program]. Version 6.0.23. Retrieved from http://www.PRAAT.org/

Borkovec, T. D. (1994). The nature, functions, and origins of worry. In G. C. L. Davey & F. Tallis (Eds.), *Worrying: Perspectives on theory, assessment and treatment* (pp. 5–33). John Wiley & Sons.

Bosker, H. R. (2014). *The processing and evaluation of fluency in native and non-native speech.* Netherlands Graduate School of Linguistics.

Bosker, H. R., Pinget, A.-F., Quené, H., Sanders, T., & De Jong, N. H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing, 30*(2), 159–175.

Botes, E., Dewaele, J.-M., & Greiff, S. (2020). The foreign language classroom anxiety scale and academic achievement: An overview of the prevailing literature and a meta-analysis. *Journal for the Psychology of Language Learning, 2*, 26–56.

Brown, H. D. (2007). *Teaching by principles: An interactive approach to language pedagogy.* Pearson Longman.

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate et al. (Eds.), *Researching pedagogic tasks: Second language learning, teaching, and testing* (pp. 23–48). Longman.

Cardoso, W. (2022). Technology for speaking development. In *The Routledge Handbook of Second Language Acquisition and Speaking* (pp. 299-313). Routledge.

Cattell, R. B., Knapp, R. R., & Scheier, I. H. (1961). Second-order personality factor structure in the objective test realm. *Journal of Consulting Psychology, 25*(4), 345.

Calvo, M., & Eysenck, M. (1996). Phonological working memory and reading test anxiety. *Memory, 4*, 289–305. https://doi.org/10.1080/096582196388960

Cardoso, W. (2022). Technology for speaking development. In *The Routledge handbook of second language acquisition and speaking* (pp. 299–313). Routledge.

Chambers, F. (1997). What do we mean by fluency? *System* 25 (4): 535 – 544.

Chamot, A. U. (2005). Language learning strategy instruction: Current issues and research. Annual Review of Applied Linguistics, 25, 112–130.

Chamot, A. U. (2009). *The CALLA handbook: Implementing the cognitive academic language learning approach.* Pearson Education.

Chamot, A. U., & Harris, V. (Eds.). (2019). *Learning strategy instruction in the language classroom: Issues and implementation.* Multilingual Matters.

Chamot, A. U., Barnhardt, S., El-Dinary, P. B., & Robbins, J. (1999). *The learning strategies handbook.* Longman.

Chang, M.-M., & Hung, H.-T. (2019). Effects of technology-enhanced language learning on second language acquisition: A meta-analysis. *Educational Technology & Society, 22*(4), 1–17. https://www.jstor.org/stable/26910181

Chen, Y., Carger, C. L., & Smith, T. J. (2017). Mobile-assisted narrative writing practice for young English language learners from a funds of knowledge approach. *Language Learning & Technology*, 21(1), 28–41. doi:10.1080/09588221.2018.1544151

Corda, A., Koenraad, T., & Visser, M. (2012). Spreekvaardigheidsdidactiek en ICT. *Levende Talen Tijdschrift*, *13*(1), 36-44.

Côte, S., & Gaffney, C. (2018). The effect of synchronous computer-mediated communication on beginner L2 learners' foreign language anxiety and participation. *The Language Learning Journal.* https://doi.org/10.1080/09571736.2018.1484935

Council of Europe. (2011). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment.* Council of Europe.

Darling-Aduana, J., & Heinrich, C. J. (2018). The role of teacher capacity and instructional practice in the integration of educational technology for emergent bilingual students. *Computers & Education, 126*, 417–432. https://doi.org/10.1016/j.compedu.2018.08.002

De Bot, K. (1992). A bilingual production model: Levelt's 'speaking' model adapted. *Applied Linguistics, 13*, 1–24.

De Jong, N. H., & Perfetti, C. A. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning, 61*(2), 533–568. https://doi.org/10.1111/j.1467-9922.2010.00620.x

De Jong, N. H. (2016). Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching,* 54, 113–132.

De Jong, N. H. (2018). Fluency in second language testing: Insights from different disciplines. *Language Assessment Quarterly*, *15*, 237–254.

De Jong, N. H., & Bosker, H. R. (2013). Choosing a threshold for silent pauses to measure second language fluency. In R. Eklund (Ed.), *Proceedings of the 6th workshop on disfluency in spontaneous speech* (DiSS) (pp. 17 – 20). Royal Institute of Technology (KTH).

de Jong, N.H., Pacilly J. & Heeren, W. (2021) PRAAT scripts to measure speed fluency and breakdown fluency in speech automatically, *Assessment in Education: Principles, Policy & Practice, 28:4,* 456-476. https://doi.org/10.1080/0969594X.2021.1951162

De Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2015). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics, 36*(2), 223–243.

De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2012). 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). John Benjamins. https://doi.org/10.1075/lllt.32.06jon

Dewaele, J.M. (2002). Individual differences in L2 fluency: The effect of neurobiological correlates. In Vivian Cook (ed.), *Portraits of the L2 user*, 219–250. Multilingual Matters.

Derwing, T. M. (2019). L2 fluency development. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 246–259). New York: Routledge.

Derwing, T. M., & Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: A 7-year study. *Language Learning, 63*(2), 163–185.

Derwing, T. M., Rossiter, M. J., & Munro, M. J. (2002). Teaching native speakers to listen to foreign-accented speech. *Journal of multilingual and multicultural development*, *23*(4), 245-259.

Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition,* 31, 533–557.

Dörnyei, Z., & Kormos, J. (1998). Problem-solving mechanisms in L2 communication: A psycholinguistic perspective. *Studies in Second Language Acquisition, 20.* https://www.jstor.org/stable/44485756

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

Everly, P. (2018). Expanding pronunciation instructional time beyond the classroom: Microsoft Office 2016 OneNote Class Notebook as an interactive delivery platform. *TESOL Journal*, *10*(2). doi: 10.1002/tesj.421

Eysenck, M. W. (2010). Attentional control theory of anxiety: Recent developments. In Aleksandra Gruszka, Gerald Matthews & Blazej Szymura (eds.), *Handbook of individual differences in cognition: Attention, memory and executive control,* 195–204. Springer.

Eysenck, M. W. & M.G. Calvo. (1992). Anxiety and performance: The processing efficiency theory. *Cognition and Emotion 6*(6). 409–434.

Eysenck, M.W., Santos Nazanin Derakshan, R. & Calvo. M.G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion* 7(2). 336–353. .https://doi.org/10.1037/1528-3542.7.2.336

Fillmore, C. (1979). On fluency. In C. Fillmore, D. Kempler, & W. S.-Y. Wang (Eds.), *Individual differences in language ability and language behavior* (pp. 85–101). Academic Press.

Foster, P. (2020). Oral fluency in a second language: A research agenda for the next ten years. *Language Teaching*, *53*(4), 446-461.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second language acquisition*, *18*(03), 299-323.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21*(3), 354–375.

Freed, B. F. (2000). Is fluency, like beauty, in the eyes (and ears) of the beholder? In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 243–265). University of Michigan Press.

Fulcher, G. 2000. The 'communicative' legacy in language testing. *System* 28(4): 483–497.

Gatbonton, E., & Segalowitz, N. (1988). Principles for promoting fluency within a communicative framework. *TESOL Quarterly, 22*, 473-492. doi:10.2307/3587290

Golonka, E., Bowles, A., Frank, V., Richardson, D., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning, 27*(1), 70–105. https://doi.org/10.1080/09588221.2012.700315

González-Lloret, M. (2017). Technology for task-based language teaching. *The handbook of technology and second language teaching and learning*, 234-247.

González-Lloret, M., & Ortega, L. (2014). Towards technology-mediated TBLT: An introduction. In M. González-Lloret & L. Ortega (Eds.), *Technology-mediated TBLT: Researching technology and tasks* (pp. 1–22). John Benjamins. https://doi.org/10.1075/tblt.6.01gon

Götz, S. (2013). *Fluency in native and nonnative English speech* (Vol. 53). John Benjamins.

Grimshaw, J., & Cardoso, W. (2018). Activate space rats! Fluency development in a mobile game-assisted environment. *Language Learning & Technology, 22*(3), 159–175. https://doi.org/10.125/44662

Guillot, M. (1999). *Fluency and its teaching.* Multilingual Matters.

Hanzawa, K. (2021). Development of second language speech fluency in foreign language classrooms: A longitudinal study. *Language Teaching Research.* https://doi.org/10.1177/13621688211008693

Hartsuiker, R. J., & Notebaert, L. (2010). Lexical access problems lead to disfluencies in speech. *Experimental Psychology, 57*(3), 169–177.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, *77*(1), 81-112.

Hauck, M. & Hurd, S. (2005). Exploring the link between language anxiety and learner self-management in open language learning contexts. *European Journal of Open and Distance Learning*.

Hewitt, E., & Stephenson, J. (2012). Foreign language anxiety and oral exam performance: A replication of Phillips's MLJ study. *The Modern Language Journal*, 96, 170–189.

Horwitz, E. K. (2010). Foreign and second language anxiety. Language Teaching, 43(2), 154–167. https://doi.org/10.1017/S026144480999036X

Horwitz, E.K. (2013). *Becoming* a *language teacher*: *A practical guide* to *second language learning* and *teaching*, 2nd edition, Pearson.

Horwitz, E. K. (2017). On the misreading of Horwitz, Horwitz and Cope (1986) and the need to balance anxiety research and the experiences of anxious language learners. In C. Gkonou, M.

Daubney & J.-M. Dewaele (Eds.), *New insights into language anxiety: Theory, research and educational implications* (pp. 31–47). Multilingual Matters.

Horwitz, E. K., Horwitz, B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal, 70*, 125–132. https://www.jstor.org/stable/327317

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied linguistics*, *30*(4), 461-473.

Huang, L. F., & Gráf, T. (2020). Speech Rate and Pausing in English: Comparing Learners at Different Levels of Proficiency with Native Speakers. *Taiwan Journal of TESOL*, *17*(1), 57-86.

Huensch, A., & Tracy–Ventura, N. (2016). Understanding second language fluency behaviour: The effects of individual differences in first language fluency, cross-linguistic differences, and proficiency over time. *Applied Psycholinguistics*, 38, 755–785

Huensch, A., & Tracy–Ventura, N. (2017). L2 utterance fluency development before, during, and after residence abroad: A multidimensional investigation. *Modern Language Journal*, 101, 275–293.

Hunter, A.-M. (2017). *Fluency development in the ESL classroom: The impact of immediate task repetition and procedural repetition on learners' oral fluency* (Unpublished doctoral dissertation). University of Surrey.

IELTS. n.d. *IELTS Speaking band descriptors*. http://www.ielts.org/PDF/UOBDs_SpeakingFinal.pdf

Jiang, M. Y. C., Jong, M. S. Y., Lau, W. W. F., Chai, C. S., & Wu, N. (2023). Exploring the effects of automatic speech recognition technology on oral accuracy and fluency in a flipped classroom. *Journal of Computer Assisted Learning, 39(1*), 125-140.

Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning, 64*, 809–854. https://doi.org/10.1111/lang.12084

Kahng, J. (2022). Fluency. In T. M. Derwing, M. J. Munro, & R. I. Thomson (Eds.), *The Routledge handbook of second language acquisition and speaking* (pp. 188–200). Routledge.

Kerncurriculum für das Gymnasium Schuljahrgange 6-10: Niederländisch. Hg. Niedersächsisches Kulturministerium. 2017..

Kircher, T. J., Brammer, M.J., Levelt W.J., Bartels, M. & McGuire, P.K. (2004). Pausing for thought: engagement of left temporal cortex during pauses in speech. *NeuroImage* 21(1): 84–90.

Kormos, J. (1999). Monitoring and self-repair in L2. *Language Learning, 49*, 303–342.

Kormos, J. (2000). The role of attention in monitoring second language speech production. *Language Learning, 50*(2), 343–384.

Kormos, J. (2006). *Speech production and second language acquisition.* Erlbaum.

Kormos, J. (2015). Individual differences in second language speech production. In J. W. Schwieter (Ed.), *The Cambridge handbook of bilingual processing* (pp. 369–388). Cambridge University Press.

Kormos, J. & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32, 145–164.

Lambert, C., Kormos, J., & Minn, D. (2017). Task repetition and second language speech processing. *Studies in Second Language Acquisition, 39*(1), 167–196.

Lastovetska, A. App development cost: Understand your budget to build powerful apps [Internet]. 2024 [cited 2024 Feb 19]. Available from: https://mlsdev.com/blog/app-development-cost

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behaviour Research Methods*, 44 (2), 325 - 343. https://doi.org/10.3758/s13428-011-0146-0

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40*, 387–417.

Lennon, P. (2000). The lexical element in spoken second language fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 25–42). University of Michigan Press.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* MIT Press.

Levelt, W. J. M. (1999). Producing spoken language: a blueprint of the speaker. In C. Brown & P. Hagoort (Eds.), *Neurocognition of language*. (pp. 83-122). Oxford University Press.

Levelt, W.J.M., Roelofs, A., & Meyer, A.S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1 –37.

Lynch, T., & Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research*, *4*(3), 221-250.

MacIntyre, P. D. (2017). An overview of language anxiety research and trends in its development. In C.Gkonou, M. Daubney & J.-M. Dewaele (Eds.), *New insights into language anxiety: Theory, research and educational implications* (pp.11–30). Multilingual Matters.

MacIntyre, P. D., & Gardner, R. C. (1989). Anxiety and second language learning: Towards a theoretical clarification. *Language Learning*, *39*(2), 251–275.

MacIntyre, P. D., & Gardner, R. C. (1991). Methods and results in the study of anxiety in language learning: A review of the literature. *Language Learning*, *41*, 85–117.

MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning, 44*(2), 283–305. https://doi.org/10.1111/j.1467-1770.1994.tb01103.x

Maclay, H. and C. E. Osgood. 1959. Hesitation phenomena in spontaneous English speech. *Word* 15: 19–44.

Mak, B. (2011). An exploration of speaking-in-class anxiety with Chinese ESL learners. *System, 39*(2), 202-214.

McClain, D. M. (2016). Artificially intelligent computer assisted language learning system with AI student component. *Knowledge Exchange: Senior Research Projects*.188. https://knowledge.e.southern.edu/senior_research/188

McCrocklin, S. (2016). ASR-based dictation practice for second language pronunciation improvement. Journal of Second Language Pronunciation,5(1), 98–118.processing in the second language. *Language Learning* 44(2). 283–305. https://doi.org/10.1075/jslp.16034.mcc

McDonough, K. & Trofimovich, P. (2008). *Using priming methods in second language research*. London: Taylor & Francis.

Michel, M. (2017). Complexity, Accuracy and Fluency in L2 production. In S. Loewen, & M. Sato (Eds.), *Routledge Handbook of Instructed Second Language Acquisition* (pp. 50-68). (Routledge handbooks in applied linguistics). Routledge.

Michel, M., Kuiken, F., & Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *International Review of Applied Linguistics in Language Teaching, 45*(3), 241–259.

Mora, J. C., & Valls-Ferrer, M. (2012). Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *Tesol Quarterly*, *46*(4).

Nattinger, J.R., & DeCarrico, J.S. (1992). *Lexical phrases and language teaching*. Oxford University Press.

Peltonen, P. (2017). Temporal fluency and problem-solving in interaction: An exploratory study of fluency resources in L2 dialogue. *System, 70*, 1-13. http://dx.doi.org/10.1016/j.system.2017.08.009

Pérez Castillejo, S. (2019). The role of foreign language anxiety on L2 utterance fluency during a final exam. *Language Testing,* 1 – 19 https://doi.org/10.1177/0265532218777783 .

Perez Castillejo, S. (2021). Prior processing, foreign language classroom anxiety, and L2 fluency. *International Review of Applied Linguistics in Language Teaching*. https://doi.org/10.1177/0265532218777783

Pickering, L. (2001). The role of tone choice in improving ITA communication in the classroom. *TESOL Quarterly, 35*, 233–249.

Plonsky, L., & Ghanbar, H. (2018). Multiple regression in L2 research: A methodological synthesis and guide to interpreting R2 values. *The Modern Language Journal, 102*(4), 713-731.

Plonsky, L., & Oswald, F. L. (2014). How Big Is "Big"? Interpreting Effect Sizes in L2 Research. *Language Learning,* 64(4), 878-912.

R Core Team. (2020). R: *A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing.

Riazantseva, A. (2001). Second language proficiency and pausing. *Studies in Second Language Acquisition, 23*, 497–526.

Reitsma, F. (2018). *77 kommunikative Spiele: Wortschatz und Grammatik – Niederländisch*, Verlag Ernst Klett Sprachen.

Reitsma, F. (2022). *Nederlands voor buitenlanders,* 7th edition, Uitgeverij Unieboek/het Spectrum

Richards, J. C., & Schmidt, R. (2010). *Longman dictionary of language teaching and applied linguistics* (4th ed.). Routledge.

Riggenbach, H. (1991) Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes* 14(4). 423 – 441. https://doi.org/10.1080/01638539109544795

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics, 22*(1), 27–57. https://doi.org/10.1093/applin/22.1.27

Robinson, P. (2005). Cognitive complexity and task sequencing: A review of studies in a Componential Framework for second language task design. *International Review of Applied Linguistics in Language Teaching* 43 (1): 1–33.

Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. *International Review of Applied Linguistics*, 45, 193–213.

Rood, M. M. S. J., & de Jong, N. H. (2023). A mixed-methods study on the effects of manipulated speaking anxiety on L2 utterance and L2 cognitive fluency. *Journal of the European Second Language Association*, 7(1), 60–74. https://doi. org/10.22599/jesla.96

Rossiter, M. (2003). " It's Like Chicken but Bigger": Effects of Communication Strategy in the ESL Classroom. *Canadian Modern Language Review*, *60*(2), 105-121.

Rossiter, M. J., Derwing, T. M., Manimtim, L. G., & Thomson, R. I. (2010). Oral fluency: The neglected component in the communicative language classroom. *Canadian Modern Language Review, 66,* 583–606.

Rubin, J., Chamot, A.U., & Anderson, N.J. (2007). Intervening in the use of strategies. In Cohen, A.D., & E. Macaro (Eds.), *Language learner strategies: Thirty years of research and practice* (pp. 141–160). Oxford University Press.

Saito, K., Ilkan, M., Magne, V., Tran, M., & Suzuki, S. (2018). Acoustic characteristics and learner profiles of low, mid and high-level second language fluency. *Applied Psycholinguistics, 39*, 593-617.

Sanaei, O., Zafarghandi, A. M., & Sabet, M. K. (2015). The effect of classroom anxiety on EFL learners' oral narratives fluency: The case of intermediate level students. *Theory and Practice in Language Studies*, *5*(7), 1390–1400. http://dx.doi.org/10.17507/tpls.0507.11

Satar, H. M., & Özdener, N. (2008). The effects of synchronous CMC on speaking proficiency and anxiety: text vs. voice chat. The Modern Language Journal, 92(4), 595-613. https://doi.org/10.1111/j.1540-4781.2008.00789.x

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in second language acquisition*, *14*(4), 357-385.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge University Press.

Segalowitz, N. (2010). *The cognitive basis of second language fluency.* Routledge.

Segalowitz, N. (2016). Second language fluency and its underlying cognitive and social determinants. *International Review of Applied Linguistics in Language Teaching*, *54*(2), 79-95.

Seifoori, Z., & Vahidi, Z. (2012). The impact of fluency strategy training on Iranian EFL learners' speech under online planning conditions. *Language Awareness*, 21(1–2). https://doi.org/10.1080/09658416.2011.639894

Skehan, P. (1996). *A framework for the implementation of task-based instruction.* Applied Linguistics, 17(1), 38–62.

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

Skehan, P. (2003). Task-based instruction. *Language teaching*, *36*(01), 1-14.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics, 30*(4), 510–532.

Skehan, P. (2014). Limited attentional capacity, second language performance, and task-based pedagogy. In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 211–260). John Benjamins. https://doi.org/10.1075/tblt.5.08ske

Skehan, P. (2015). Limited attention capacity and cognition: Two hypotheses regarding second language performance on tasks. In M. Bygate (Ed.), *Domains and directions in the development of TBLT: A decade of 191 plenaries from the international conference* (pp. 123–156). John Benjamins.

Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.). *Cognition and second language instruction* (pp. 183–205). Cambridge University Press.

Skehan, P., & Foster, P. (2007). Complexity, accuracy, fluency and lexis in task-based performance: A meta-analysis of the Ealing research. *Complexity, accuracy, and fluency in second language use, learning, and teaching*, *207*, 226.

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, 143–167.

Suzuki, S., Kormos, J., & Uchihara, T. (2021). The relationship between utterance and perceived fluency: A meta-analysis of correlational studies. *The Modern Language Journal*, *105*(2), 435-463.

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

Swerts, M. 1998. Filled pauses as markers of discourse structure. *Journal of Pragmatics* 30: 485–496.

Tavakoli, P. (2011). Pausing patterns: Differences between L2 learners and native speakers. *ELT Journal*, 65(1), 71–79.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. *International Journal of Applied Linguistics, 15*(3), 241–261.

Tavakoli, P., & Hunter, A. M. (2018). Is fluency being 'neglected' in the classroom? Teacher understanding of fluency and related classroom practices. *Language Teaching Research*, *22*(3), 330-349.

Tavakoli, P. & Wright. C. (2020). *Second Language Speech Fluency. From Research to Practice*. C.U.P.

Tavakoli, P., Campbell, C., & McCormack, J. (2016). Development of speech fluency over a short period of time: Effects of pedagogic intervention. *TESOL Quarterly*. http://dx.doi.org/10.1002/tesq.244

Tavakoli, P., Nakatsuhara, F., & Hunter, A.M. (2017). *Scoring validity of the Aptis Speaking test: Investigating fluency across tasks and levels of proficiency*, ARAGs Research Reports Online, AR-G/2017/7, British Council, London, available online at: www.britishcouncil.org/sites/default/files/tavako li_et_al_layout.pdf

Tavakoli, P., Nakatsuhara, F., & Hunter, A. M. (2020). Aspects of fluency across assessed levels of speaking proficiency. *The Modern Language Journal*, *104*(1), 169-191.

Tecedor, M. & Campos-Dintrans, G. (2018). Developing oral communication in Spanish lower-level courses: The case of voice recording and videoconferencing activities. *ReCALL*. https://doi.org/10.1017/S0958344018000083

Teimouri, Y., Goetze, J., & Plonsky, L. (2019). Second language anxiety and achievement: A meta-analysis. *Studies in Second Language Acquisition*, *41*, 363–387.

Tiberius, C., & Schoonheim, T. (2013). *A frequency dictionary of Dutch: Core vocabulary for learners*. Routledge.

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, *17*(1), 84-119.

Tyler, A. (1992). Discourse structure and the perception of incoherence in international teaching assistants' spoken discourse. *TESOL Quarterly, 26.* https://doi.org/10.2307/3586870

Varonis, E.M. and Gass, S. (1982) The Comprehensibility of Non-Native Speech. Studies in Second Language Acquisition, 4, 114-136.

Wang, J., An, N., & Wright, C. (2018). Enhancing beginner learners' oral proficiency in a flipped Chinese foreign language classroom. *Computer Assisted Language Learning,* 31(5–6), 570–521. https://doi.org/10.1080/09588221.2017.1417872

Welie, C. (2022). Spreken. In F. Kuiken & S. Andringa. *Handboek Nederlands als tweede taal in het volwassenenonderwijs*. Uitgeverij Coutinho.

Willis, J. (1996). *A Framework for Task-Based Learning*. Longman.

Willis, D., & Willis, J. (2007). *Doing Task-Based Teaching.* Oxford University Press.

Wood, D. (2010). *Formulaic language and second language speech fluency.* Continuum.

Wood, D. (2016). Willingness to communicate and second language speech fluency: An idiodynamic investigation. *System*, 60, 11–28.

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics,* 21, 463–489.

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

Wray, A. (2008). *Formulaic language: Pushing the boundaries.* Oxford University Press.

Wang Z. (2018). On-line time pressure manipulations: L2 speaking performance under five types of planning and repetition conditions. In P. Skehan (Ed.), Processing perspectives on task performance (pp. 27-62). John Benjamins.

Wood, D. (2010). *Formulaic language and second language speech fluency.* Continuum.

Wood, D. (2016). Willingness to communicate and second language speech fluency: An idiodynamic investigation. *System*, 60, 11–28.

Zhang, X. (2019). Foreign language anxiety and foreign language performance: A meta-analysis. *The Modern Language Journal*, *103*, 763–781.

Zuniga, M. (2015). *Attention and second language speech production: The relationship between attention and the production of self-initiated self-repairs*. University of Quebec at Montreal dissertation.

Appendix A: Screenshots of the Reppen App

Screenshots illustrating the Exercise 'Which Speaker is the Most Fluent?'

(English translation)

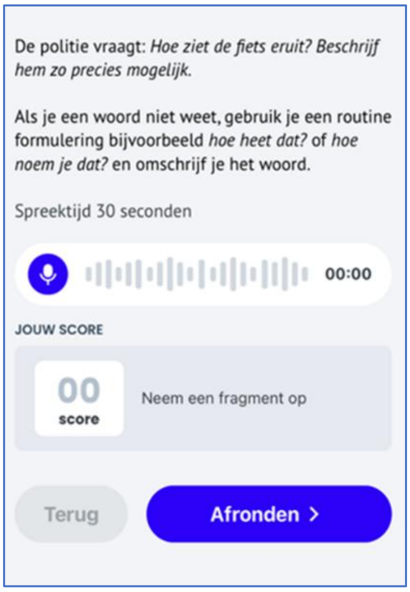
Fluency

Your friend has pain in her shoulders and neck. You know a few exercises. Look at the pictures. You will hear four speakers explaining what your friend should do (left).

Which speaker do you find the most fluent? Rank them in order. Speaker 1, 2, 3, 4 (middle).

Listen again to speakers 1 and 2. Which speaker do you prefer to listen to? Speaker 1 Speaker 2 (right).

Appendix B: Screenshots of the Reppen App



Strategy: Routine Phrases

You parked your mother's bike at the supermarket. Look at the picture. When you return, the bike is gone. You go to the police (left).
The police ask: What does your bike look like? Describe it as precisely as possible.
If you don't know a word, use a routine phrase such as "what's that called?" or "how do you say that?" and describe the word. Speaking time: 30 seconds (middle).

This is your score based on your recorded audio clip.

Speaking pace Spoken time

Tips

Just right Your speaking pace is great.

Good Fantastic! You say a lot and you don't make many pauses and/or your pauses are not too long (right).

Appendix C: Debriefing Questionnaire

I would like to hear from you about your experiences with the *Reppen* app. Please answer the following questions.

1.Working with the *Reppen* app was... (please check all that apply)

☐ enjoyable	☐ challenging	☐ refreshing
☐ interesting	☐ helpful	☐ a waste of time
☐ motivating	☐ boring	☐ educational

2. I feel that through working with the app, I… (please check all that apply).

☐ speak Dutch more fluently

☐ am less anxious

☐ am more confident when speaking Dutch

☐ know the components of fluent speech

☐ am aware of the impact of non-fluent speech on the listener.

3. Please check a box below for each statement to indicate the extent to which it applies to you. The scale ranges from (1) fully applies to (5) does not apply at all.

	fully applies to				does not apply at all
I was able to use …					
a. the feedback provided by the app in the speaking exercises.	☐	☐	☐	☐	☐
b. the tips provided by the app in the speaking exercises.	☐	☐	☐	☐	☐
c. the following strategies in the speaking exercises:					

Describing the unknown word	☐	☐	☐	☐	☐
Routine-phrases (e.g. "How do you say that?)	☐	☐	☐	☐	☐
Lexical fillers (e.g. 'Well"," You know")	☐	☐	☐	☐	☐

4.What did you like about working with the app? Please describe why.

5.What did you not like so much about working with the app? Please describe why.

6. Is there anything you would change about the app? If so, what and why?

Appendix D: Questions for the Teachers

1. What are your experiences using the Reppen app in your Dutch lessons? What positive and/or negative aspects have you experienced?

2. Do you have the impression that the use of the app has contributed to improving the students' speaking skills? If so, what improvements have you observed?

3. Do you think anxious students benefit from practicing speaking skills via the app? Why yes/no?

4. Do you have any suggestions for improving the app?

5. Would you use the *Reppen* app if it becomes available to a broader audience after my research? Why or why not?

Appendix E: Speaking Tasks

You will perform two speaking tasks. You will give a monologue for both tasks.

Simple task:

Read the first task.

What did you do last weekend?

-Talk about the activities you did.

You have a maximum of two minutes. If you finish earlier, you can stop speaking.

Do you understand the task? The task will be recorded.

(Lees de eerste opdracht door.

Wat heb je het afgelopen weekend gedaan?

-Vertel over de activiteiten die je hebt gedaan.

Je hebt maximaal twee minuten de tijd. Als je eerder klaar bent, kun je stoppen met spreken.

Begrijp je de opdracht? De opdracht wordt opgenomen.)

Complex task:

Read the second task.

In the Dutch lesson, you discuss the question:

Smartphones in the classroom: to ban or to allow?

What is your opinion? Also, provide arguments to convince the other students.

You have a maximum of two minutes. If you finish earlier, you can stop speaking. Do you understand the task? The task will be recorded.

(Lees de tweede opdracht door.

In de Nederlandse les discussiëren jullie over de vraag:

Smartphones in de klas: verbieden of toestaan?

Wat is jouw mening? Geef ook argumenten om de andere leerlingen te overtuigen.

Je hebt maximaal twee minuten de tijd. Als je eerder klaar bent, kun je stoppen met spreken. Begrijp je de opdracht? De opdracht wordt opgenomen.)

Appendix F: Demographische Daten und Fragen zur Sprachlerngeschichte

Beantworten Sie bitte ein paar allgemeine Fragen zu Ihrem Hintergrund.

1. Ihr Geschlecht: ☐ weiblich ☐ männlich

2. Wie alt sind Sie? _______________Jahre

3. Ihre Muttersprache(n)*: ______________________________

* Die Sprache die Sie von Geburt an (bzw. seit ihrem 4. Lebensjahr) sprechen.

4. Haben Sie Familienmitglieder oder enge Freunde, für die Niederländisch die Muttersprache ist und mit denen Sie sich regelmäßig auf Niederländisch unterhalten?

☐ ja ☐ nein

5. Seit wann lernen Sie Niederländisch? ________________ Jahre und ____________Monate.

6. Hatten Sie längere Aufenthalte (z.B. Schüleraustausch, Sommerkurse usw.) in einem niederländischsprachigen Land?

☐ ja ☐ nein

Wenn ja, wie lange?

____ Wochen/Monate/Jahre

Appendix G: Results of Linear Mixed-Effects Models

Results of mixed effects modeling analyses on factors affecting number of silent pauses

Fluency measure	**Fixed effects: Factor**	***Estimate***	***SE***	***p***	**Random effects: Factor**	**Variance**
Number of silent pauses	Intercept	52.63	14.04	**<0.001**	Participant	69.11
	Proficiency	-0.30	0.22	0.17		
	Anxiety	0.10	0.12	0.40		
	Task type	-1.05	1.49	0.48		

Results of mixed effects modeling analyses on factors affecting phonation time ratio

Fluency measure	**Fixed effects: Factor**	***Estimate***	***SE***	***p***	**Random effects: Factor**	**Variance**
Phonation time ratio	Intercept	50.50	11.32	**<0.001**	Participant	47.54
	Proficiency	0.28	0.18	0.12		
	Anxiety	-0.05	0.09	0.57		
	Task type	1.40	1.13	0.22		

Appendix H: Model Comparisons

Speech rate

Anova (model 1, model 2)
model1: speechratetotal1 ~ 1 + proficiency + anxiety + Task + (1 | ID)
model2: speechratetotal1 ~ 1 + proficiency + anxiety + Task + as.factor(School) + (1 | ID)

Model	**AIC**	**Deviance**	**Chisq**	**df**	***p***
Model 1	1042.7	1030.7			
Model 2	1044.4	1028.4	2.33	2	0.31

Number of pauses

Anova (model 1, model 2)
model1: numofpauminute1 ~ 1 + proficiency + anxiety + Task + (1 |ID)
model2: numofpauminute1 ~ 1 + proficiency + anxiety+ Task + as.factor(School) + (1 | ID)

Model	**AIC**	**Deviance**	**Chisq**	**df**	***p***
Model 1	936.80	924.80			
Model 2	940.36	924.36	0.44	2	0.80

Number of mid-clause pauses

Anova (model 1, model 2)
model 1: Numbermidclausepauzes1 ~ 1 + proficiency + anxiety + Task + (1 | ID)
model 2: Numbermidclausepauzes1 ~ 1 + proficiency+ anxiety + Task + as.factor(School) + (1 | ID)

Model	**AIC**	**Deviance**	**Chisq**	**df**	***p***
Model 1	716.45	704.45			
Model 2	718.74	702.74	1.71	2	0.42

Phonation time ratio

Anova (model 1, model 2)
model1: PTR1 ~ 1 + proficiency + anxiety + Task + (1 |ID)
model2: PTR1 ~ 1 + proficiency + anxiety + Task + as.factor(School) + (1 | ID)

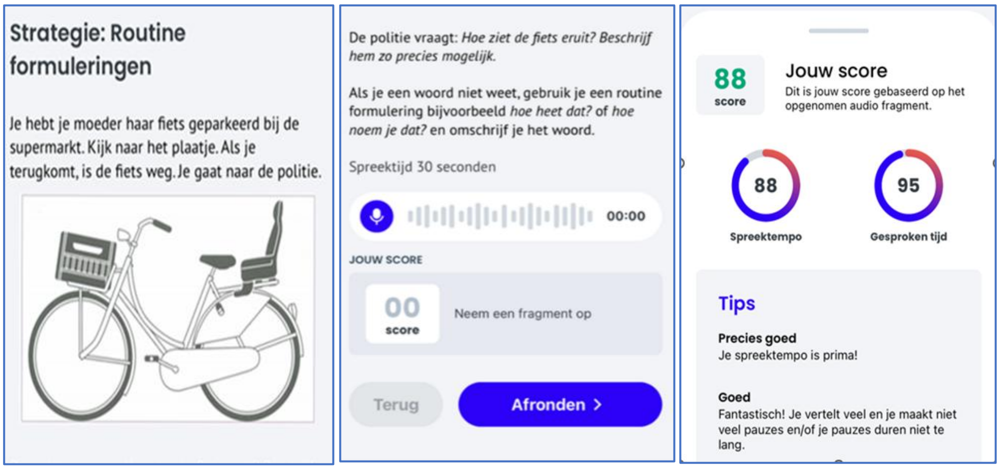
Model	**AIC**	**Deviance**	**Chisq**	**df**	***p***
Model 1	813.10	801.10			
Model 2	815.31	799.31	1.79	2	0.41

Silent pauses per minute

Anova (model 1, model2)
Model 1: sil.pausesminute1 ~ 1 + proficiency + anxiety + Task + (1 |ID)
Model 2: sil.pausesminute1 ~ 1 + proficiency + anxiety + Task + as.factor(School) +(1 | ID)

Model	**AIC**	**Deviance**	**Chisq**	**df**	***p***
Model 1	869.51	857.51			
Model 2	873.02	857.02	0.491	2	0.78

Appendix I: Sample Tasks from the Reppen App



Alt Text: An exercise in which a bicycle needs to be described, practicing fluency strategies of formulaic sequences and circumlocution (left). The recording must last at least 30 seconds before it can be analyzed (middle). Scores for speech rate and phonation time ratio with feedback tips (right).

Sample fluency training

(Part of week 1

Preparation: Awareness-raising of the importance of fluency and identify features of fluency that have an impact on the listener)

1. You will hear two speakers explaining how to cook potatoes. Listen to both recordings.
a. Which speaker do you prefer to listen to?

☐ Speaker 1 ☐ Speaker 2

(*You probably prefer listening to Speaker 1. Speaker 1 doesn't use long pauses, uses few "uh"s, has a good speaking pace, and the explanation is clear and understandable.*)

b. Listen again to the recordings and read the transcript.

Speaker 1:
Okay mm you have to uhh peel the potatoes and then uhh put them in plenty of boiling water for about thirty minutes, and then you pour off the water and then you can serve and eat the potatoes.

Speaker 2:
Uhh first you prepare the potatoes for the… uhh Topf (pot) then you have to put water in the… uhhm… yes fill the Topf… then the potatoes are boiled for twenty minutes… uhh then… you pour the water… out of the Topf and in the end you can… can eat the uhh potatoes.

c. Which speaker do you think is more fluent?
☐ Speaker 1 ☐ Speaker 2

d. Why is Speaker 1 more fluent? Because Speaker 1:
☐ Uses fewer pauses
☐ Says "uh" or "um" less often
☐ Makes fewer long pauses
☐ Has a good speaking pace: not too slow and not too fast

e. Speaker 2 couldn't come up with the Dutch word *pan* and used a lot of pauses and the German word *Topf* instead. What could Speaker 2 have done to sound more fluent? Think of two tips.

Tip 1: ……….
Tip 2: ………..

(Tip 1: She could have used a description instead of the German word *Topf*, for example: *a thing you use to cook potatoes in*.
Tip 2: Instead of using long pauses, she could have used filler words like *well*, *let me see*, or phrases like *what do you call that?* or *how do you say that?* These expressions give you time to think about what you want to say next or to find the right word.)

Sample speaking lesson

Describing Figures

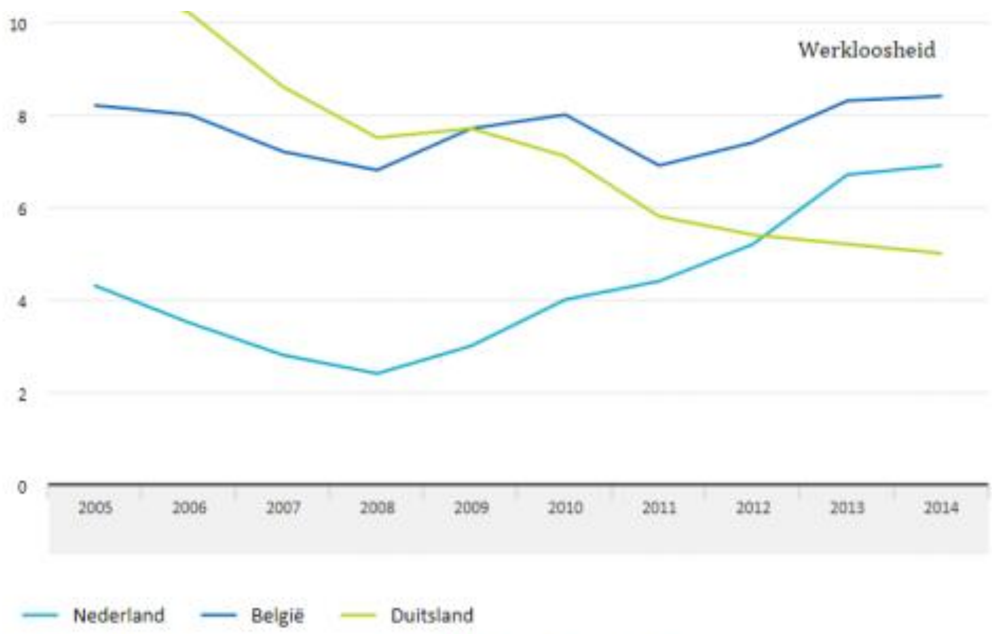
Pre-task

When describing a graph or diagram, you can use the following words and phrases:

The diagram/chart shows ...
The diagram/chart indicates that ...

Increasing		Decreasing	
to increase (has increased)	*the increase*	*to decrease (has decreased)*	*the decrease*
to rise (has risen)	*the rise*	*to fall (has fallen)*	*the fall*

Not increasing or decreasing
to remain the same/constant

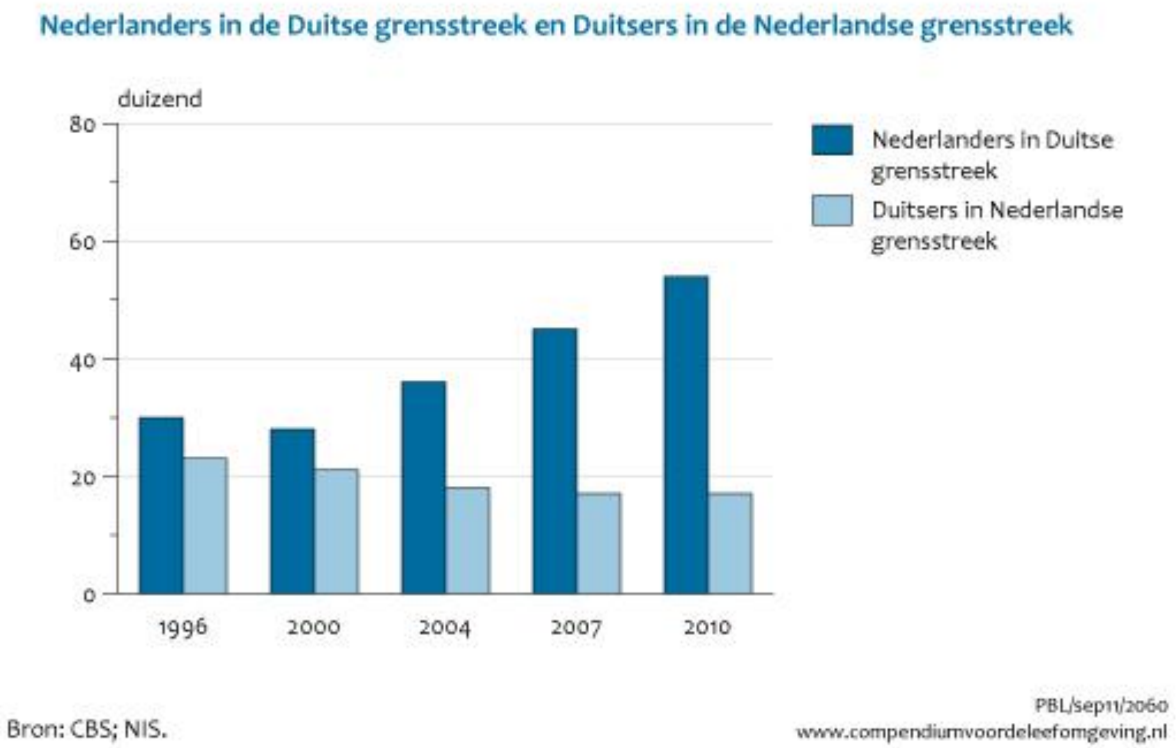


Grafiek 6: Werkeloosheid in Nederland, Belgie en Duitsland sinds 2005.

Description of Graph
This graph depicts the unemployment rates in the Netherlands, Belgium, and Germany from 2005 to 2014. The graph compares the unemployment rates among the Netherlands, Belgium, and Germany. In the graph, it can be observed that the unemployment rate in the

Netherlands decreased from 2005 to 2008 and then increased after 2008. However, in Germany, the unemployment rate declined significantly from 2006 to 2014. Meanwhile, during the same period, the unemployment rate in Belgium remained more or less constant.

1. Look at the diagram



Fill in the missing words.
The diagram........(a)........... the number of Dutch people living in the German border region and the number of Germans living in the Dutch border region between 1996 and 2010. From the diagram, it is evident that during this period, the number of Dutch people living in the German border region has ………..(b)…….significantly, while the number of Germans living in the Dutch border region has slightly ……(c)……….. What explanations are there for this ....(d)........of the number of Dutch people in the German border region?)

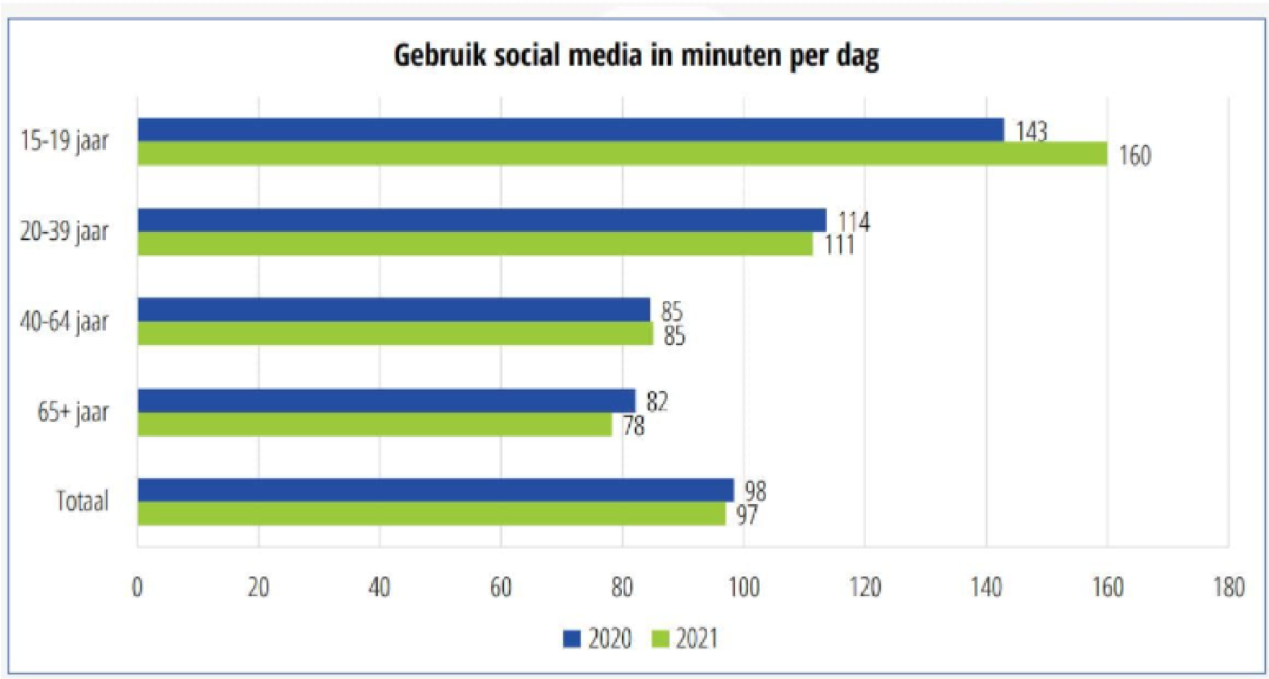
(Correct solution:
a. shows/illustrates
b. risen/increased
c. decreased/dropped
d. increase/rise)

When discussing a graph or diagram, you can use the following structure:

*Introduction: What does the graph/diagram depict?
*Main: What stands out?
*Conclusion: What can be concluded?

During- task

2. You present the following diagram in class



https://www.frankwatching.com/archive/2021/01/26/social-media-onderzoek-2021/

Prepare by first writing down keywords. Use the following format

Introduction: ……………………..
Main: ....
Conclusion: …………………..

Preparation time: 4 minutes
Speaking time: 1 minute

Post-task

App-based feedback		Advice
Variation in word choice	small	Try to use words like *increase, decrease, decline, remain constant* (see explanation describing figures).
	sufficient	You're doing well but try to vary even more in words describing the diagram (see explanation describing figures).
	large	Fantastic, an excellent variety!

Appendix J: Speech Production Tasks

1. In the city where you live, there are long traffic jams every day. You discuss this with a friend.

There are three solutions to this traffic jam problem:

-Build more roads
- Construct more bicycle lanes
-Improve public transportation

a. Discuss the advantages and disadvantages of these three solutions.
b. Choose the best solution and argue why this is the best solution.

2. In the Dutch class, the topic is about global warming. You discuss this with a friend.

Solutions to this problem are:
- Travel less
- Eat no more meat
- Buy local products

a. Discuss the solutions.
b. Which solution do you think is the best and why?

3. At your school, the school administration wants to offer half of the school subjects in English. What is your opinion about this? Discuss it with your friend.

a.Discuss the pros and cons of education in English. You can think of:

Advantages	Disadvantages
International contacts	Mother tongue also important
Job opportunities	Not all teachers speak English well

b. Share your opinion about education in English and provide an argument.

4. *Alcara* is a new music streaming service. It is possible to get a paid subscription or stream music for free. Your friend would like to subscribe to *Alcara*. What is your opinion about it?

a. Discuss the pros and cons of a subscription. You can think of:

Subscription advantages	Subscription disadvantages
No advertisements during streaming	Cost: € 9.99 per month

b. Tell what you would do and why.

Appendix K: Descriptive Statistics per Class

Variable	Class 1 N=22 Mean SD Range	Class 2 N=8 Mean SD Range	Class 3 N=12 Mean SD Range	Class 4 N=14 Mean SD Range	Class 5 N=8 Mean SD Range	Class 6 N=13 Mean SD Range	Class 7 N=12 Mean SD Range	F	*p* value
Age (years)	15.05 (0.21) 15-16	15.25 (0.46) 15-16	17.25 (0.45) 17-18	17.14 (0.36) 17-18	18.25 (0.46) 18-19	18.15 (0.55) 18-20	15.17 (0.57) 14-16	136.63	<.001
Years of Dutch Instruction	4 (0.00) 4	4 (0.00) 4	5.92 (0.29) 5-6	6 (0.00) 6	7 (0.00) 7	3.15 (0.55) 3-5	5 (0.00) 5	381.01	<0.001
Proficiency pretest score	56.30 (8.58) 40-71.25	57.34 (7.71) 46.25-70	61.98 (8.83) 46.25-75	56.96 (7.15) 43.75-70	57.97 (6.75) 47.50-66.25	55.10 (8.10) 43.75-70	49.38 (6.52) 35-58.75	2.75	0.017
Anxiety pretest score	2.54 (0.51) 1.63-3.67	2.20 (0.78) 1.48-3.97	1.87 (0.32) 1.36-2.48	2.05 (0.28) 1.58-2.70	2.44 (0.55) 1.79-3.48	2.41 (0.54) 1.61-3.42	2.31 (0.56) 1.36-3.18	3.03	0.01

Appendix L: Correlations between Fluency Measures

Correlations between fluency measures at Test Time 0

	Speech rate	Articulation rate	Phonation Time ratio	Frequency of mid-clause pauses	Frequency of end-clause pauses
Speech rate	1	.725**	.736**	-.227*	-.444**
Articulation rate		1	.092	-.099	-.184
Phonation time ratio			1	-.229*	-.518**
Frequency of mid-clause pauses				1	.239*
Frequency of end-clause pauses					1

Note: **Significant at the $p < 0.01$ level.
*Significant at the $p < 0.05$ level.

Correlations between fluency measures at Test Time 1

	Speech rate	Articulation rate	Phonation Time ratio	Frequency of mid-clause pauses	Frequency of end-clause pauses
Speech rate	1	.577**	.759**	-.289*	-.360**
Articulation rate		1	,193	-.099	-.103
Phonation time ratio			1	-.224	-.434**
Frequency of mid-clause pauses				1	.320*
Frequency of end-clause pauses					1

Note: **Significant at the $p < 0.01$ level.
*Significant at the $p < 0.05$ level.

Correlations between fluency measures at Test Time 2

	Speech rate	Articulation rate	Phonation Time ratio	Frequency of mid-clause pauses	Frequency of end-clause pauses
Speech rate	1	.877**	.770**	-.575**	-.281
Articulation rate		1	.513**	-.352**	-.004
Phonation time ratio			1	-.559**	-.490**
Frequency of mid-clause pauses				1	.242*
Frequency of end-clause pauses					1

Note: **Significant at the *p* < 0.01 level.
*Significant at the *p* < 0.05 level.

Correlations between fluency measures at Test Time 3

	Speech rate	Articulation rate	Phonation Time ratio	Frequency of mid-clause pauses	Frequency of end-clause pauses
Speech rate	1	.855**	.682**	-.344**	-.303**
Articulation rate		1	.216	-.164	.016
Phonation time ratio			1	-.397**	-,567**
Frequency of mid-clause pauses				1	.104
Frequency of end-clause pauses					1

Note: **Significant at the *p* < 0.01 level.
*Significant at the *p* < 0.05 level.

Appendix M: Results of Linear Models

Results of Linear Model Predicting Speech Rate

lm(formula = sr2 ~ sr0 + Class + Anxiety + Group, data = sr)
lm(formula = sr3 ~ sr0 + Group + Class, data = sr)

	T0 -T2				**T0-T3**			
	Estimates (SE)	*t value*	*p*	*Adjusted R^2*	*Estimates (SE)*	*t value*	*p*	*Adjusted R^2*
Predictors								
Intercept	159.489 (25.65)	6.217	< 0.001	0.39	132.865 (15.00)	8.854	< 0.001	0.37
Speech rate 0	0.344 (0.12)	2.842	0.005		0.268 (0.11)	2.489	0.015	
Control group	-20.658 (10.37)	-1.992	0.05		-10.679 (9.94)	-1.074	0.286	
Anxiety	-0.471 (0.22)	-2.173	0.03		-			
Class 2	-55.464 (13.48)	-4.114	< 0.001		-52.098 (11.24)	-4.636	< 0.001	
Class 3	0.226 (11.89)	0.019	0.984		4.808 (9.70)	0.496	0.621	
Class 4	-7.673 (10.65)	-0.721	0.473		4.443 (9.38)	0.474	0.637	
Class 5	1.644 (12.09)	0.136	0.892		6.626 (10.61)	0.625	0.534	
Class 6	-36.023 (10.70)	-3.366	0.001		-13.695 (8.95)	-1.530	0.130	
Class 7	-20.658 (10.37)	-1.992	0.05		-10.679 (9.94)	-1.074	0.286	

Model T0-T2: $F(9,74) = 7.5$, $p < 0.001$

Model T0-T3: $F(8,72)=7.541$, $p<0.001$

Results of Linear Model Predicting Articulation Rate

lm(formula = AR2 ~ AR0 + Group + Anxiety + Proficiency, data = ar)

lm(formula = AR3 ~ AR0 + Group + Anxiety + Proficiency, data = ar)

	T0 -T2				**T0-T3**			
	Estimates (SE)	*t value*	*p*	*Adjusted R^2*	*Estimates (SE)*	*t value*	*p*	*Adjusted R^2*
Predictors				*0.13*				*0.11*
Intercept	164.447 (40.71)	4.040	< 0.001		111.472 (34.80)	3.204	0.001	
Articulation rate 0	0.408 (0.13)	3.195	0.002		0.359 (0.10)	3.426	< 0.001	
Control group (Group2)	-21.312 (8.97)	-2.377	0.019		-0.328 (7.385)	-0.044	0.964	
Anxiety	-0.270 (0.24)	-1.129	0.262		-0.047 (0.19)	-0.246	0.806	
Proficiency	-0.328 (0.50)	-0.651	0.517		0.476 (0.42)	1.132	0.261	

Model T0-T2: (F (4,78)=4.08, p<0.01)
Model T0-T3: (F(4,75)=3,5, p<0.05)

Results of Linear Model Predicting Phonation Time Ratio

lm(formula = PTR2 ~ PTR0 + Group + Anxiety + Proficiency, data = ptrdata)

lm(formula = PTR3 ~ PTR0 + Group + Anxiety + Class, data = ptrdata)

	T0 -T2				**T0-T3**			
Predictors	*Estimates (SE)*	*t value*	*p*	*Adjusted R^2*	*Estimates (SE)*	*t value*	*p*	*Adjusted R^2*
Intercept	51.816 (13.01)	3.981	< 0.001	0.33	90.267 (7.27)	12.412	< 0.001	0.51
ptr0	0.575 (0.11)	5.212	< 0.001		0.063 (0.07)	0.813	0.419	
Control group (Group2)	-5.837 (2.67)	-2.183	0.032		-11.136 (2.50)	-4.451	< 0.001	
Anxiety	-0.151 (0.07)	-2.099	0.039		-0.134 (0.04)	-2.887	0.005	
Proficiency	-0.029 (0.15)	-0.196	0.844					
Class 2					-21.439 (3.03)	-7.072	< 0.001	
Class 3					-4.081 (2.58)	-1.581	0.118	
Class 4					-1.321 (2.42)	-0.546	0.586	
Class 5					0.849 (2.64)	0.322	0.748	
Class 6					-3.364 (2.23)	-1.505	0.136	
Class 7					-11.136 (2.50)	-4.451	< 0.001	

Model T0-T2: (F(4,78)=11.11, p<0.001)

Model T0-T3: (F(9,71)=11.14, p<0.001)

Results of Linear Model Predicting Frequency of Mid-clause Pauses

T0-T2 lm(formula = mp2 ~ mp0 + Group + Class, data = mp)

T0 -T3 lm(formula = mp3 ~ mp0 + Group + Class, data = mp)

	T0 -T2				**T0-T3**			
Predictors	*Estimates (SE)*	*t value*	*p*	*Adjusted R^2*	*Estimates (SE)*	*t value*	*p*	*Adjusted R^2*
Intercept	13.462 (2.48)	5.44	< 0.001	0.21	7.955 (2.24)	3.553	< 0.001	0.31
mp0	0.249 (0.09)	2.597	0.011		0.342 (0.08)	3.905	< 0.001	
Control group	2.282 (2.50)	0.913	0.364		2.022 (2.53)	0.800	0.426	
Class 2	6.613 (2.99)	2.209	0.030		6.555 (2.75)	2.382	**0.019**	
Class 3	-4.161 (2.62)	-1.586	0.116		4.272 (2.41)	1.773	0.080	
Class 4	-1.564 (2.36)	-0.662	0.510		5.928 (2.27)	2.608	**0.011**	
Class 5	2.038 (2.84)	0.719	0.474		3.715 (2.60)	1.427	0.157	
Class 6	4.317 (2.67)	1.615	0.110		9.474 (2.25)	4.213	**< 0.001**	
Class 7	2.282 (2.50)	0.913	0.364		2.022 (2.53)	0.800	0.426	

Model T0-T2: F (8,74) =4.09, p < 0.01

Model T0-T3: F (8,72) =6.10 , p < 0.01

Results of Linear Model Predicting Frequency of End-clause Pauses

lm(formula = ep2 ~ ep0 + Group + Class, data = ep)

lm(formula = ep3 ~ ep0 + Group + Class + Anxiety, data = ep)

	T0 -T2				**T0-T3**			
	Estimates (SE)	*t value*	*p*	*Adjusted R^2*	*Estimates (SE)*	*t value*	*p*	*Adjusted R^2*
Predictors								
Intercept	10.680 (2.25)	4.745	< 0.001	0.25	11.452 (4.50)	2.544	0.013	0.18
End-clause pauses 0	0.228 (0.07)	3.217	0.001		0.034 (0.08)	0.437	0.663	
Control group (Group2)	6.550 (2.04)	3.210	0.001		9.303 (2.41)	3.863	< 0.001	
Anxiety	-				0.026 (0.04)	0.600	0.550	
Class 2	-0.493 (2.34)	-0.211	0.833		6.337 (2.78)	2.280	0.025	
Class 3	0.532 (2.09)	0.255	0.799		0.486 (2.50)	0.194	0.846	
Class 4	-1.778 (1.89)	-0.939	0.350		-0.791 (2.31)	-0.342	0.733	
Class 5	-0.609 (2.30)	-0.265	0.791		-0.974 (2.60)	-0.375	0.708	
Class 6	2.313 (2.16)	1.073	0.286		1.854 (2.28)	0.813	0.418	
Class 7	6.550 (2.04)	2.040	0.002		9.303 (2.41)	3.863	0.0002	

Model T0-T2: (F(8,72)=4.85, p<0.01)
Model T0-T3: (F(9,71)=3,24, p<0.01)

Appendix N: Results of Linear Models

Results of Linear Model Predicting Speech Rate

lm(formula = sr1 ~ sr0 + Group + Class, data = sr)

Predictors	**T0 -T1** *Estimates (SE)*	*t value*	*p*	*Adjusted R^2*
Intercept	113.720 (25.31)	4.493	< 0.001	0.39
Speechrate0	0.272 (0.15)	1.873	0.06	
Control group	-2.003 (7.44)	-0.269	0.78	
Class 2	-21.854 (19.94)	-1.096	0.28	
Class 3	22.394 (18.75)	1.194	0.24	
Class 4	15.354 (18.71)	0.821	0.42	
Class 5	8.41 (19.43)	0.433	0.67	

F-statistic: 5.46 on 6 and 36 DF, p-value: 0.004

Results of Linear Model Predicting Phonation time ratio

lm(formula = ptr1 ~ ptr0 + Group + Class, data = ptrdata)

	T0 -T1			
Predictors	*Estimates (SE)*	*t value*	*p*	*Adjusted R^2*
Intercept	36.847 (10.85)	3.395	0.002	0.59
Phonation time ratio0	0.657 (0.14)	4.600	< 0.001	
Control group	0.189 (2.08)	0.091	0.93	
Class 2	-11.893 (5.43)	-2.189	0.04	
Class 3	-4.275 (5.14)	-0.832	0.41	
Class 4	-6.419 (5.27)	-1.217	0.23	
Class 5	-7.745 (5.42)	-1.430	0.16	

F-statistic: 11.27 on 6 and 36 DF, p-value: 0.001

Results of Linear Model Predicting Frequency of Mid-Clause Pauses

lm(formula = mp1 ~ mp0 + Group + Class, data = mp)

Predictors	**T0 -T1** *Estimates (SE)*	*t value*	*p*	*Adjusted R^2*
Intercept	6.049 (4.87)	1.242	0.222	0.36
Frequency of mid-clause pauses	0.495 (0.14)	3.484	0.001	
Control group	-0.122 (1.90)	-0.064	0.94	
Class 2	8.456 (4.82)	1.756	0.08	
Class 3	0.370 (4.55)	0.081	0.94	
Class 4	1.784 (4.51)	0.396	0.69	
Class 5	5.122 (4.70)	1.089	0.28	

F-statistics: 4.96 on 6 and 36 DF, p-value: 0.001

Results of Linear Model Predicting Frequency of End-Clause Pauses

lm(formula = ep1 ~ ep0 + Group + Class, data = mp)

Predictors	**T0 -T1** *Estimates (SE)*	*t value*	*p*	*Adjusted R^2*
Intercept	0.943 (4.68)	0.201	0.841	0.34
Frequency of end-clause pauses	0.348 (0.09)	3.845	0.0005	
Control group	-0.646 (1.59)	-0.406	0.687	
Class 2	10.918 (4.17)	2.612	0.013	
Class 3	7.130 (4.06)	1.754	0.099	
Class 4	7.516 (4.09)	1.839	0.074	
Class 5	8.878 (4.32)	2.057	0.050	

F-statistic: 4.64 on 6 and 36 DF, p-value: 0.001

About the author

Foekje Reitsma was born on 3 August 1966 in Dokkum, the Netherlands. After completing her secondary education at Oostergo in Dokkum, she studied Dutch and History at the Ubbo Emmius teacher training college in Leeuwarden from 1984 to 1989. After graduating, she studied Dutch Language and Literature at the University of Groningen, where she specialized in Applied Linguistics and Language Use (*Toegepaste Taalkunde* and *Taalbeheersing*) from 1989 to 1992. Following the completion of her studies, she worked for three years at the Language Centre of the University of Groningen. In 1995, she was appointed as a Lektorin at the Institute for Dutch Studies (*Institut für Niederlandistik*) at the Carl von Ossietzky University of Oldenburg. After several years, she specialized further in the didactics of Dutch as a foreign language.

In October 2019, Foekje Reitsma started a PhD position at the Institute for Dutch Studies at the University of Oldenburg. In her PhD research, she investigated the relationship between foreign language anxiety and L2 fluency, as well as the effects of an app-based fluency training on L2 fluency and foreign language anxiety.

Eigenständigkeitserklärung

Hiermit erkläre ich, dass ich die Dissertation selbstständig verfasst habe und dass ich alle benutzten Hilfsmittel vollständig angegeben habe.

Oldenburg, den 24.07. 2024 Foekje Reitsma