UNIVERSITY OF TURKU

The Influence of a 7-Week Spoken English Course on the Use of Stalling Mechanisms:

A Mixed-methods Study of Fluency Development

Hanna Aittokumpu

Master's Thesis

Degree Programme for Language Learning and Teaching, Department of English

School of Languages and Translation Studies

Faculty of Humanities

University of Turku

September 2024

Master's Thesis

Degree Programme for Language Learning and Teaching, Department of English

Hanna Aittokumpu

The Influence of a 7-Week Spoken English Course on the Use of Stalling Mechanisms: A Mixed-methods Study of Fluency Development

47 pages, with 9 pages in appendices

This MA thesis studies the influence that a 7-week spoken English course has on spoken English development. The study examines the changes in the use of stalling mechanisms as indicators of fluency by investigating both group-level changes and individual variation during the course. The topic of fluency is current in Finland as oral L2 proficiency is considered increasingly more in teaching, with suggestions to include spoken language testing in the matriculation examination having already been made. As of now, Finnish upper secondary schools offer only one spoken English course as an elective (ENA8) that focuses on oral performance, and the aim of this thesis is to investigate how short-term structured instruction influences fluency development during that course. 33 students were examined for their use of stalling mechanisms by analysing speech samples that were recorded before and after the course. The speech samples were transcribed into written format and annotations were made for five stalling mechanisms: silent and filled pauses, drawls, filler words and repetitions. In the group level analysis, the development in the frequency of every stalling mechanism was analysed individually to discover statistically significant changes during the course. This was accompanied by a qualitative study based on extreme case sampling (Dörnyei 2007) to give more detailed insight into individual variation inside the group and short-term gains that are achievable.

The results show that group-level changes were mixed. Statistically significant development was observed in the frequency of silent pauses, drawls and Finnish influence on filled pauses, but not in the other stalling mechanisms. Individual variation within the group was substantial, with some students improving their fluency significantly, gaining improvements in the frequency, duration and distribution of silent pauses and other stalling mechanisms, indicating that while group-level development might be harder to detect, individually students can improve greatly during a short timeframe.

Key words: English, second language acquisition, SLA, fluency, temporal fluency, stalling mechanisms, upper secondary school

Table of contents

List of tables

List of figures

List of abbreviations

CAF   Complexity, accuracy and fluency

CEFR   Common European Framework of Reference for Languages

GDPR   General data protection regulation of EU

L2   Second language

PSM   Problem-solving mechanism

1   Introduction

Fluency is an important aspect in second language learning in an increasingly globalized world, where English serves as a lingua franca across different domains of society, such as business, education and international communication. In Finland, English is taught from the first grade onward, but opportunities for practicing the language outside of the classroom might be limited. In an environment such as this, fluency can pose significant challenges for people learning English and trying to develop their oral fluency for effective communication. Fluency is regarded as one dimension of second language (L2) proficiency, which is multidimensional in nature and involves more than the mastery of grammatical functions and vocabulary; it requires the ability to produce the second language smoothly and with native-like hesitation, pausing and speed (Housen, Kuiken & Vedder 2012, 2). According to Paul Lennon, fluency is: "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, 26). This definition could be applied to the written medium as well, as fluency is not necessarily limited only to speech. However, this thesis explores fluency in oral speech, and more specifically, how oral fluency develops and what kinds of problem-solving mechanisms students of a foreign language utilize during speech. This can offer valuable information for effective language instruction and assessment as well.

From the perspectives of language teaching and assessment in Finland, the topic of oral fluency is current, since suggestions for including spoken language testing in the matriculation examination have already been made (Ministry of Education and Culture, 2017). Previous studies concerning oral language development have mostly been longitudinal studies, focusing on development over a longer period of time. However, as oral fluency is going to be incorporated in official testing in the near future, it is beneficial to examine what kinds of changes can happen during a shorter timeframe. Studies concerning gains that are achieved during a short period of time have usually been conducted in study abroad contexts where learners immerse themselves in the culture and language of their choosing (e.g. Song 2020; Moneypenny & Aldrich 2024). However, this study focuses on formal classroom environment where studying is not as immersive in nature. As of now, Finnish upper secondary schools only offer one spoken English course as an elective (ENA8) according to the national core curriculum for general upper secondary education (LOPS, 2019). This is a 7-week course that students can take in upper secondary school, and the aims of the course read as follows:

> "The aim of the course is to provide the students with tools to develop their own oral language skills, for the student to deepen their understanding of the different factors influencing oral interaction, and to practice oral production that requires preparation. The course includes an oral language skill test provided by the Finnish National Board of Education" (LOPS 2019, my own translation).

This elective course (ENA8) is designed to develop oral communication skills, providing students with materials and methods to develop their language skills and opportunities to engage in both spontaneous speech and oral productions that require preparation (LOPS 2019). The course contents also are said to focus on dialogues more, meaning that conversational proficiency is at the forefront of the course, and the tasks are designed to be communicative.

This thesis explores L2 oral fluency, with the main aim of examining the development of breakdown fluency and the occurrence of stalling mechanisms, such as pauses, drawls and repetitions, during a 7-week spoken English course at a Finnish upper secondary school. By focusing on this, I offer empirical data on how short-term, structured language instruction can influence fluency development. These findings are relevant for teachers and people that design the curriculums for Finnish schools by giving insight into how students develop during a relatively short period of time. These findings can help teachers support students in overcoming challenges regarding fluency and enhancing their spoken English proficiency.

Adhering to Lennon's narrow sense of fluency (1990) and Segalowitz's three aspects of fluency – and more importantly – utterance fluency (2010), this study utilizes a mixed-methods approach to analyze both group level and individual level changes in the use of stalling mechanisms. The focus is on the frequency of silent and filled pauses, filler words, drawls and repetitions. Furthermore, I examine the changes – or possible improvements - that happen according to these fluency measures during an upper secondary school spoken English course. In other words, my research questions for this thesis are:

1. What kind of group-level changes and variation are there in the frequency of stalling mechanisms during a 7-week spoken English course?

2. What kind of individual changes and variation are there in the use of stalling mechanisms during a 7-week spoken English course?

This thesis begins with a theoretical framework section (Section 2) detailing relevant theories and terms crucial for this study. First, important definitions of fluency will be introduced, after which I will delve into utterance fluency and different measures used in analysing it. Finally, I will examine previous studies on the matter as well. In the materials and methods section of this thesis (Section 3), I will introduce the participants and the data, while including data elicitation methods used in the creation of this study. In the results section (Section 4), I will analyse and present the data gathered for this study, and then discuss them further in the discussion section (Section 5). I will conclude this thesis by summarizing the present study while considering ideas for future studies and ways for these results to be applied to language studies and assessment (Section 6).

2   L2 oral fluency

In this section, I will introduce previous literature and theoretical frameworks relevant to the present study. In section 2.1, I will start by examining how fluency is defined in L2 speech production, and after that, in section 2.2, I will focus on utterance fluency and different measures used in analysing and assessing fluency. Finally, in section 2.3, I give an overview of previous studies on fluency development.

2.1   Definitions of L2 oral fluency

A common objective for L2 learners is to become a fluent user of the target language. As a concept, fluency can be defined in many ways and approached from many perspectives. In L2 acquisition alone, fluency can be approached from the perspectives of L2 speaking, writing, listening, and reading (Lintunen, Mutta & Peltonen 2020, 1). However, in this present study, fluency is treated as an aspect of spoken proficiency and defined according to that. When talking about L2 learning and assessment, fluency is usually a key criterion in assessing language proficiency, and as such, it is a very important concept for learners and teachers alike. For example, in the Common European Framework of Reference for Languages (CEFR), fluency is listed as one of the qualitative aspects when assessing spoken language, corresponding with each of the common reference levels of the framework from basic user (A1) to proficient user (C2) (CEFR, 2024).

If aspects of fluency are taken into account when assessing L2 learner's language skills – and more specifically to this study, oral skills – then the question of what it means to speak an L2 fluently requires answering. Intuitively, we all might have an idea of the ease and effortlessness of speech when talking about someone who is fluent, but the question is quite complex. In addition to alluding to fluency as being the fluidity and effortlessness of speech, it can also refer to the ability to express ideas as well in the L2 as in the L1, having minimal accent in the L2, or to having a large vocabulary or little to no grammatical errors when speaking (Segalowitz 2010, 4). In a study done by Tavakoli and Hunter (2018), 84 L2 language teachers were investigated for their understanding of fluency and their practices regarding it via a questionnaire. When they were asked to define the main characteristics of fluent L2 speech, the teachers reported components such as "lack of hesitation, speed, fluidity, infrequent pauses" (Tavakoli & Hunter 2018, 338). In addition to this, aspects such as good pronunciation, a wide range of vocabulary and the correct use of grammar were cited as well (ibid.). These aspects correspond with the features that Segalowitz (2010, 4) mentions as well.

In Finland, one study focusing on learner perceptions of L2 fluency by Lintunen and Peltonen (2020) examined Finnish university students with English as their major. The students were asked for their perceptions with a questionnaire that had questions focusing on defining fluency. The majority of the answers discussed fluency in relation to general spoken proficiency. Fluency was also discussed with regard to pronunciation and "the natural flow of speech" and reading and writing skills. The students understood fluency as the "smoothness of communication as a whole" (Lintunen & Peltonen 2020, 22). From this study it is evident that the students examined fluency in a broader context relating to L2 spoken proficiency, although more specific aspects were cited as well, such as grammar, vocabulary, speed, pauses and hesitations (ibid.). In this study learners' perceptions are in accordance with other related studies on what sort of aspects define fluency (e.g. Tavakoli & Hunter 2018; Skehan 2009; Lennon 1990).

As mentioned before, fluency can be approached from the perspectives of writing, reading, and listening as well. For example, you can speak of fluency as the ability to read novels, poems and other difficult materials, give spontaneous speeches, and produce accurate translations from one language to another (Segalowitz 2010, 4). In everyday speech, fluency can be defined in a variety of different ways, but when researching fluency, it is important to define these different aspects and differentiate them from one another to accurately measure fluency. In order to do this – with my focus strictly being on oral fluency – relevant frameworks are examined next.

When we talk about L2 production and development, two issues are crucial: what aspects make a learner proficient in using an L2, and how can we measure this L2 proficiency reliably. Since the mid-1990s, the CAF-framework of complexity, accuracy and fluency has been the standard way of describing how multidimensional language performance is (Pallotti 2021, 201). Complexity, accuracy and fluency also play an important role in language testing and assessment (ibid.), meaning that these three components of L2 proficiency are seen as separate but interconnected aspects of language learning. Complexity refers to how lexically, morphologically, and syntactically complex the language production in question is, while accuracy means how accurately the language that is produced conforms to the target-language norms (Pallotti 2021, 202-203). Finally, fluency is defined as the "extent to which linguistic production is (and/or perceived as) fast and smooth" (ibid.). The relationship between these three dimensions can be seen as either complementary or competitive. According to Robinson's *Cognition Hypothesis* (also known as the Multiple Attentional Resources Model)*,* complexity and accuracy are complementary, and that increasing task complexity increases not only complexity but accuracy as well, possibly at the expense of fluency (Robinson 2001, 307). In contrast, in Skehan's *Trade-off Hypothesis* (Skehan 2009) the three dimensions are seen as competing, meaning that when one dimension has a higher

performance rate, it may lead to lower performance in the others (Skehan 2009, 511). From these three, especially accuracy and fluency have been seen as being at the opposite ends of a continuum where at each extreme, speech can be seen as accurate and disfluent, or inaccurate and fluent (Fulcher 2003, 26-27).

While this framework divides language proficiency into three dimensions (complexity, accuracy, and fluency), the present study only focuses on the dimension of fluency, even though in language testing and assessment, fluency has been a complex concept to define and measure. Historically – and in general use – the term fluency used to describe the learner's overall language proficiency, on the grounds of smoothness and native-likeness of speech (Housen, Kuiken & Vedder 2012, 4). However, recently a more "narrow" definition of fluency has been adopted by L2 researchers, and it is agreed upon that fluency itself is multidimensional (ibid.). According to Paul Lennon's well-established model, there are two different ways it can be understood: broad vs. narrow senses (Lennon 1990, 389). These have been called higher-order fluency and lower-order fluency later on (Lennon 2000, 25). In the 'broad' sense, fluency accounts for the overall language proficiency in an L2 – where being 'fluent' represents the highest point of language proficiency when assessed – while in the 'narrow' sense, fluency refers to an "isolatable component of oral proficiency" (ibid.), meaning, for example, the smoothness and speed of speech. According to this narrow sense of fluency, Lennon deemed fluent speech as being "unimpeded by silent pauses and hesitations, filled pauses ("ers" and "erms"), self-corrections, repetitions, false starts, and the like" (Lennon 1990, 390), while stating that a less talented learner's speech is, in contrast, slow, halting and confused (ibid.). Therefore, when examining fluency according to this narrow sense of the term, it mostly refers to the ease and effortlessness of speech. In the CAF-framework, fluency is seen as one aspect of oral proficiency, and it differs from other features such as accuracy and complexity. Lennon goes as far as to say that fluency might be able to cover for deficiencies in these other areas (Lennon 1990, 391). This narrow sense of fluency is particularly used in speech assessment and analysis by researchers. Skehan has further adapted the narrow sense of fluency to measure components such as: speed, breakdown, and repair fluency (Skehan 2009, 512-513). In his article, Skehan argues that while complexity, accuracy, and fluency (CAF) have proven to be useful measures of L2 performance, fluency needs to be reassessed in order to be measured effectively (Skehan 2009, 510). I examine Skehan's fluency dimensions later in more detail in section 2.2.1.

Adapting Lennon and Skehan, and according to research they have conducted themselves, Tavakoli and Hunter (2018) have suggested that Lennon's dichotomy of a broad sense and a narrow sense of fluency should be divided into four approaches instead (Figure 1). At the bottom of the pyramid is a very broad approach, meaning the general proficiency that one has in a foreign

language. Next is the broad approach, which encompasses L2 speaking ability, where a learner is "someone who can speak confidently and communicate their intended message well in the spoken mode" (Tavakoli & Hunter 2018, 343), incorporating aspects of pronunciation, accuracy, and the capability to hold a conversation (ibid.). Next is the narrow approach, which is similar to the CAF-framework in how it differentiates fluency from accuracy and complexity. It encompasses aspects such as the ease, flow, and continuity of speech (ibid.). Finally, at the top is the very narrow approach to fluency which contains aspects that are measurable to examine the speed, breakdown, and repair of speech. This is used in language research to assess fluency objectively by analysing measurable components, such as the frequency of pauses or other disfluencies.
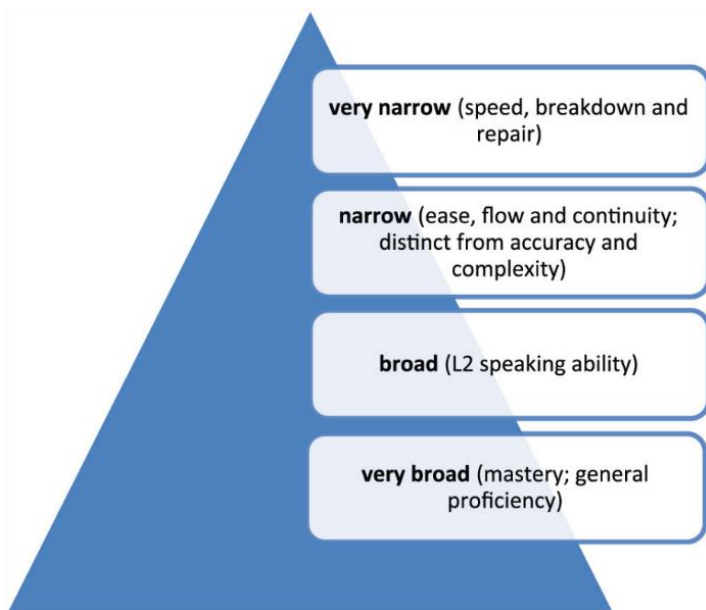
Figure 1. Four approaches to defining fluency (Tavakoli & Hunter 2018, 343)

Another important approach for fluency comes from Segalowitz (2010) with his three aspects of fluency, where he divides fluency into three separate categories: 1. cognitive fluency, 2. utterance fluency, and 3. perceived fluency. Cognitive fluency means that the speaker is able to efficiently utilize their cognitive processes that are responsible for producing speech, in order to keep the flow of speech going (Segalowitz 2010, 48). These cognitive processes use mechanisms that ensure that the planning, lexis, grammar and articulation of an utterance are done efficiently and quickly (ibid.). However, for L2 learners, processing and production are not as automatized compared to native speakers, and require more time (Dörnyei & Kormos 1998, 371). Utterance fluency "has to do with the features of an utterance. It refers to the temporal, pausing, hesitation, and repair characteristics" (Segalowitz 2010, 48) that can be found in speech. These characteristics can be individually

measured and analysed. Lastly, perceived fluency concerns the listener and the conclusions they draw from the speaker's speech (ibid.). It combines both previous aspects by assessing the speaker's cognitive fluency based on their utterance fluency. The aspect that I focus on in this study is Segalowitz's utterance fluency, which corresponds to Lennon's narrow sense of fluency as well (see 2.1.2), and in the next section, I examine utterance fluency in more detail by explaining how to measure different characteristic of fluency such as speed, breakdown, and repair. I will be using the terms 'temporal fluency' and 'stalling mechanisms' to examine these phenomena.

2.2   Fluency measurements

Recent descriptions of fluency have defined fluency as the "ability to produce the L2 with *native-like* rapidity, pausing, hesitation or reformulation" (Housen, Kuiken & Vedder 2012, 2, emphasis added). In L2 research, native-speaker performance – and especially rapidity – has been the point of comparison for fluent L2 speech and seen as something to aim for. However, even advanced learners of an L2 can easily be distinguished from native speakers – by native speakers and language learners alike. A study done by Abrahamsson & Hyltenstam (2009) researched nativelikeness by examining advanced bilingual Spanish and Swedish speakers and their perceptions on the matter. Native speakers of Swedish were used as judges, and the majority of learners who had started learning Swedish after the age of 12, were not perceived as native speakers (Abrahamsson and Hyltenstam 2009). Another study done by O'Brien (2014, 734) ascertained that even language learners can distinguish native from nonnative speech, stating differing aspects such as 'accentedness, fluency and comprehensibility' as factors distinguishing the two groups of speakers.

In addition, recent studies using a native speaker control group have noticed variation even in L1 fluency and how it contains pauses and hesitations as well (e.g. Peltonen & Lintunen 2016; Götz 2013). This goes to show that using native language as something to strive for has its own problems too, and suggestions have been made to use learners as their own control groups by comparing their L2 speech productions to their L1 speech, because L1 speaking styles have been reported to affect L2 speech fluency as well (e.g. De Jong et al. 2015; Huensch & Tracy-Ventura 2017; Peltonen 2018). One such study conducted by Peltonen (2018) has revealed that many L2 English fluency features correspond to features also exhibited in the L1, with moderate to strong correlations to temporal fluency features, and moderate correlations to stalling mechanisms, such as filled pauses and drawls. Even though 'nativeness' is hard – if not nearly impossible – to attain, and using nativeness as a metric can be questionable, it is still important to figure out what differentiates even the most advanced learners from native-like speech. To do this, it is helpful to compare the speech

of a learner to that of a native speaker with the help of fluency measures, to have a sense of what contributes to fluency or disfluencies when speaking. In the next section, I will first introduce temporal fluency measures, and after that, discuss problem-solving mechanisms – or stalling mechanisms – such as filled pauses, filler words, drawls, and repetitions, to illustrate how they are measured in fluency research.

2.2.1 Temporal fluency

As discussed in section 2.1, utterance fluency, which consists of characteristics of speech that can be measured, can be further divided into three sub-dimensions: speed, breakdown fluency and repair fluency (Skehan 2009, 512-513). L2 studies concerning fluency have primarily focused on data gathered from monologues, and analysed fluency based on various temporal fluency measures in order to examine an individual's level of L2 oral proficiency. L2 oral performance can be measured by using various fluency measures, and in her book, *Speech Production and Second Language Acquisition* (Kormos 2006)*,* Judit Kormos has provided an overview of the ten most frequently used temporal fluency measures when analysing the three dimensions of speed, breakdown, and repair. These measures include speech rate, articulation rate, phonation-time ratio and mean length of runs. Speech rate is calculated by dividing the total number of syllables by total time (including pause time) and multiplied by 60, which is then expressed in syllables per minute (Kormos 2006, 163). Articulation rate is similarly calculated, but pause time is excluded. It is the mean number of syllables per minute over the total amount of time (ibid.). Phonation-time ratio reflects how much of total speaking time is spent producing vocal sounds and is calculated by dividing total phonating time with total speaking time (Kormos 2006, 163). Finally, mean length of runs is the "average number of syllables produced in utterances between pauses of 0.25 seconds and above" (ibid.).

There are also measures for breakdown fluency that include aspects such as the number of silent and filled pauses per minute and the length of silent pauses. For repair fluency, the number of disfluencies such as false starts, repetitions, and replacements are measured. A false start means that there is a rejected sound or word(s) that were cut off, repetition means that words or longer utterances were repeated without any modifications, and replacement means words that were replaced by other words with no other modifications while speaking (Foster & Skehan 1999, 230). Many studies agree that best predictors of fluency are *speech rate*, meaning the number of syllables produced per minute, and *mean length of runs,* which is the average number of syllables produced in utterances between pauses of 0.25 seconds or above (Kormos 2006, 162). This present study does not use all of these measures but rather focuses on selected ones that are associated with the

breakdown and repair dimensions such as pausing and other stalling mechanisms. I explain the measures I chose for this study in more detail in section 3.3.

On the other hand, research findings considering the frequency of silent and filled pauses, in addition to disfluencies such as repetitions and restarts, are found to be somewhat ambiguous. Studies have discovered that there are strong correlations between the frequency, duration and location of pauses and fluent speech (Götz 2013, 18), studies with a small number of participants showing that the frequency of pauses did separate fluent speakers from less fluent learners (e.g., Freed 1995; Lennon 1990). However, one study with more participants concluded that the frequency of pauses did not correlate with the level of fluency, but that the duration of pauses and corrections correlated more strongly (van Gelderen 1994, 312-314). Peltonen adds that maybe the reason why results for filled pauses and repairs have been unclear might be because of their "context-dependent and multifunctional nature" (Peltonen 2017, 2). Maybe instead of looking at these characteristics as displaying disfluency, they can be approached as problem-solving mechanisms or speech management strategies. This perspective is adopted in the present study and discussed in more detail in the next section (see 2.2.2).

When assessing a speech sample by using these temporal fluency measures, it must be remembered that characteristics such as silent and filled pauses, repetitions and corrections can all be found in native speech as well. Therefore, in order to assess learner speech and see where the differences to native speech are, it is important to analyse the distribution of these features as well, instead of only the frequency of them (Lennon 1990, 392-393), meaning that the location of pauses and disfluencies is important when measuring L2 oral fluency. In less fluent learner's speech, disfluencies tend to occur in clusters of pauses, hesitations and repairs (Freed 1995, 2000) and within syntactic units (Lennon 1990) while more advanced learners and native speakers tend to pause between "major syntactic boundaries" (ibid.). This means that language learners exhibit more mid-clause pausing and hesitation than native speakers. This present study uses a mixed-methods approach that combines a frequency-based quantitative analysis of group level differences with a qualitative analysis of individual changes. This qualitative analysis focuses on the location and duration of pauses and other disfluencies as well, complementing the quantitative part of the study. As said before, this study also uses a perspective that disfluencies are seen as problem-solving mechanisms, and the next section examines this approach in more detail.

2.2.2 Stalling mechanisms as problem-solving mechanisms

Over the years, many studies have examined utterance fluency by focusing on temporal aspects and temporal fluency measurements. Studies have measured speech rate, the distribution and frequency

of pauses, repetitions and corrections, the amount of speech between pausing – and the length of pauses as well – in order to analyse fluency and its development effectively. Studies have also suggested that as fluency improves, hesitation phenomena evolve as well: less fluent learners have more unfilled and filled pauses, but when fluency develops, new kind of hesitation phenomena emerge, such as drawls and filler words (Freed 1995, 125). This is to reduce the time spent in silence and to keep the flow of speech going (e.g. Dörnyei & Kormos, 1998). This also broadens the view of individual fluency into interactional fluency, with learners resorting to different types of problem-solving mechanisms (PSMs) in order to process time pressure when communicating (Dörnyei & Kormos 1998). Research on temporal fluency has mostly regarded aspects such as repetitions as disfluency indicators, instead of seeing them as potential fluency-enhancing resources (Peltonen 2017) that help with problem-solving during speaking. In an illustration by Peltonen (2017, 3), temporal fluency is connected to stalling mechanisms and further into communication strategies as well (Figure 2). The left side of the figure contains the classic temporal fluency aspects of speed, breakdown and repair, of which the latter two are frequently thought of as displaying disfluency and are analysed by using different fluency measures. When approaching aspects such as filled pauses and repairs as problem-solving mechanisms that are connected to communication strategies instead of markers of disfluency, we can begin to refer to them as stalling mechanisms. As we can see here, there is significant overlap in these aspects of fluency, which further adds to the notion that fluency itself is multidimensional or multifunctional.
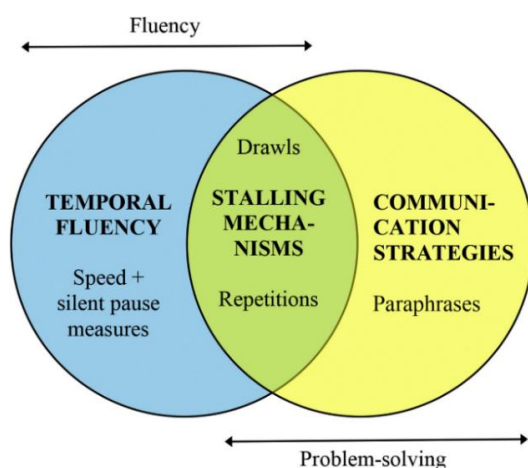

Figure 2. Connecting temporal fluency to stalling mechanisms and communication strategies (Peltonen 2017, 3).

An important basis for the discussion on stalling mechanisms is the psycholinguistic approach by Dörnyei and Kormos (1998) that distinguishes 'processing time pressure' as one of the main

problem sources of L2 communication, and in this approach, they have detailed various PSMs to combat it. In her book *Fluency in Native and Nonnative English Speech* (Götz 2013, 35), Sandra Götz calls PSMs 'speech management strategies' to battle planning pressure and enhance fluency. Processing time pressure is related to L2 speech processing and how it requires more attention and time in L2 speech production than in the native language (Dörnyei & Kormos 1998, 357), which in turn leads to employing different kinds of stalling mechanisms when planning and executing the message in question (ibid.). Producing speech in an L2 is not as automatized as producing speech in one's native language. Therefore, in order to keep the flow of communication going, learners need to avoid lengthy silences and extreme hesitations (Dörnyei & Kormos 1998, 368). Dörnyei and Kormos (ibid.) go on to detail different options that learners have when language production takes too much time. For the purpose of avoiding extreme hesitations created by planning and processing, learners may, for example, reduce or even abandon what they say. On the other hand, in order to keep communicating, learners might apply stalling mechanisms to gain more planning time for themselves (ibid.).

According to Dörnyei and Kormos (1998), there are different types of stalling mechanisms learners use when experiencing time pressure, such as: pauses, sound lengthening, fillers, and repetitions, but these can be organised under two main types: pauses and repetitions (Table 1). Pauses can be further divided into nonlexicalized and lexicalized pauses, of which non-lexicalized pauses can be seen as including silent pauses, filled pauses and sound lengthening, while lexicalized pauses contain filler words such as *well, like, you know, I mean* etc. Nonlexicalized filled and unfilled pauses do not need any additional processing but do not maintain fluency either because how hesitant and "disjointed" the speech becomes, meanwhile sound lengthening is a more effective way of "holding the floor" (Dörnyei & Kormos 1998, 370). The other main type, repetitions, can also be divided into self-repetition and other-repetition. Dörnyei & Kormos (1998, 371) add that stalling mechanisms are also used by native speakers for both problematic and non-problematic processing. However, the role might be more pronounced in L2 learning as processing and producing it is not as automatized and therefore requires more time (ibid.). In section 3.3, I will explain the specific measures of fluency I used in data analysis in more detail.

Table 1. Problem-solving mechanisms (PSMs) related to processing time pressure (Based on Dörnyei & Kormos 1998, 369-370)

Type of PSM	Description
Pauses	
Nonlexicalized pauses: - unfilled pauses - umming and erring - Sound lengthening (drawl)	-Remaining silent while thinking -Using filled pauses (er, uh, umm) -lengthening a sound in hesitation
Lexicalized pauses: - Filler words	Using filling words to fill pauses, to stall, and to gain time in order to keep the communication channel open and maintain discourse at times of difficulty
Repetitions	
Self-repetition	Repeating a word or a string of words immediately after they were said
Other-repetition	Repeating something the interlocutor said to gain time

2.3 Previous studies concerning fluency

The empirical studies concerning fluency usually use one of three different approaches: either fluency is studied longitudinally, or by comparing fluent and nonfluent speakers, or correlating fluency assessments with temporal fluency measures (Kormos 2006, 162). Concerning the focus of the present study, I will focus on longitudinal studies first, after which I will examine short-term studies on the matter. Longitudinal studies are studies that take place over multiple months or even years and are used to research fluency development both quantitatively and qualitatively. For example, Towell et al. (1996) studied the development of fluency in 12 advanced learners of French over a period of four years by using temporal fluency measures in their mixed-methods study. The temporal fluency measures that were selected for the study were: speaking rate, phonation-time ratio, articulation rate, mean length of runs, and average length of pauses (ibid.). The students were required to complete oral and written tasks during their studies in the University of Salford, Great Britain, and the data were gathered by asking the participants to watch a movie and then retell the story in a recording booth (Towell et al. 1996). The recordings were taken during the second year of their studies and the third year of their studies. To compare their speech to their native language (English) the participants also recorded an English version during their fourth year of studies. The quantitative results reveal that there were improvements in speech rate, but this was mostly because of changes in the mean length of runs that, according to the researchers, appears to be the best

indicator of fluency development (Towell et al 1996). Performance in all measures except for phonation-time ratio and average length of pauses was not as high as in their native language (ibid.). A qualitative study revealed that "there is an increase in the length and complexity of the linguistic units which are uttered between pauses" (Towell et al. 1996, 112-113). This means that there is development in how quickly stored knowledge of the L2 can be accessed and processed for speech production (ibid).

Another longitudinal study by Hanzawa (2024) examined the impact a year-long L2 "learning experience" had on speed, breakdown and repair fluency measures. 50 Japanese university students were examined via a narrative task that was administered three times during the academic year and asked them to describe "the toughest or most challenging event they had experienced in the past few months" (Hanzawa 2024, 28). Temporal measures the researchers used were articulation rate, mid-clause and end-clause pause frequencies and lengths, and the frequency of repetitions and other corrections (Hanzawa 2024). The findings suggest that there were improvements in pause length (but not in frequency) in both mid-clause and end-clause positions, and that changes in repair and pause frequencies correlated strongly with the amount of L2 learning the participants had (ibid.). The study suggests that the effort the students dedicated to learning an L2 increased their use of repairs, which implies that as fluency develops so does the use of these problem-solving mechanisms.

In addition to this, studies comparing fluent and nonfluent speakers (e.g. Ejzenberg 2000; Hilton 2008) have suggested that higher language proficiency correlates with higher L2 speech fluency as well (Peltonen 2020, 18). Studies concerned with development have usually been longitudinal and during study abroad, creating a more immersive environment for language learning, but few studies have been focused on fluency development in formal school settings (Peltonen 2020, 20). In a study where short-term gains of fluency were inspected in a university setting (Peltonen & Lintunen 2024), the researchers used two cartoon strips for a picture description task and compared two groups: a control group and an experimental group. The data were gathered from an L2 English pronunciation course and both groups were given the same amount of teaching. However, only the experimental group received "targeted fluency activities" (Peltonen & Lintunen 2024, 47). The speech samples were transcribed, and annotations were made for seven fluency measures: articulation rate, duration of silent pauses, number of silent and filled pauses, number of repetitions, speech rate and mean length of utterance (Peltonen & Lintunen 2024). there were improvements in speech rate but no other changes concerning fluency were significant. However, there was great individual variation in the use of stalling mechanisms such as repetitions and filled pauses within the group (ibid.). Other previous studies have also shown that learner's exhibit high

individual variation especially in the use of filled pauses, and that they might have personal preferences in the choosing of stalling mechanisms as well (e.g. Götz 2013; Peltonen 2020; Wolk et al. 2020).

One study by Tavakoli, Cambell and McCormack (2016) studied university student's fluency development over a 4-week period by examining monologues. They had a control group that was focused on general speaking and listening skills, and an experimental group that – in addition to this – received "awareness-raising activities and fluency strategy training" (Tavakoli, Cambell & McCormack 2016, 447). Results reveal that both groups had higher articulation and speech rates, while pauses at clause boundaries also decreased (Tavakoli, Cambell & McCormack 2016, 459). Only the experimental group demonstrated improvements in the frequency of repair measures as well (ibid.). These studies suggest that there can be development and even improvements during short-term learning in formal settings as well. However, the type of instruction and activities might play a role in it.

Most language students list effective communication as one of the most important goals of language learning, but at the same time one of their greatest challenges (Maseda & DeFeo 2014, 110). In a study conducted by Maseda and DeFeo (2014) they researched how learning environments and learning materials can affect students' confidence and therefore fluency. The study took place at the University of Alaska Anchorage during the fall semester of 2010, meaning it was a study conducted over a shorter period of time. 14 students enrolled in a Spanish class that was taught by a native speaker of Spanish and focused on communicative tasks, interaction and authentic learning materials. The students achieved "gains in motivation and engagement, confidence and self-efficacy", which in turn contributed to language and fluency development (Maseda & DeFeo 2014, 116). The study used in-class instructor observations, student feedback, oral presentations and class discussions to gauge for fluency development during the course (Maseda &DeFeo 2014). Students from the study cited that they were nervous to speak but became braver and developed more trust in their abilities during the course, when at the beginning of the course the students displayed hesitations in their speech (Maseda & DeFeo 2014, 120-121). In addition to this, confidence is linked to cognitive fluency as well. Both breakdown fluency (pausing) and repair fluency are "related to the extent to which the learner is confident that what has been stored is reliable and the extent to which the learner has also created procedures which can be brought into operation to repair the situation when communication breakdown occurs" (Housen, Kuiken & Vedder 2012, 55). It might be that the focus of a short course such as the one in the study, and the one offered at Finnish upper secondary schools (which I discussed in the introduction of this

thesis), could be on developing students' confidence as communicators more, and that in turn develops their fluency as well.

According to Peltonen and Lintunen (2016, 209), contemporary studies concerning L2 oral fluency are "predominantly quantitative examinations that focus on broad, group level differences in fluency", with individual differences given less attention. Results of their study examining Finnish-speaking and Swedish-speaking learners propose that a quantitative analysis is sufficient for silent pauses and speed, but this should be accompanied by a qualitative investigation because "differences in the use of filled pauses and repair phenomena could only be revealed with a qualitative study" (ibid.). This present study examines the short-term gains of a 7-week course by conducting a mixed methods analysis such as the one conducted by Peltonen and Lintunen (2016). Group level tendencies are examined quantitatively with qualitative aspects added for individual variation. For example, subtle improvements in fluency might not be reflected in quantitative analysis and on group level but could be significant when individuals are analysed qualitatively. In the next section, I will introduce the participants and explain the data and the methods I used in analysing my data set.

3   Materials and methods

I start this section by introducing the participants in section 3.1, after which I will explain the data and the data collection methods in section 3.2. Finally, in section 3.3 I give an overview of the methods I used in data analysis. Regarding terminology, both terms 'unfilled pause' and 'silent pause' have been used previously in this thesis to describe the same stalling mechanism. Even though they are synonymous with each other, to avoid any confusion over the terminology, I will be referring to them as silent pauses in the following sections.

3.1   The participants

The participants of this present study were Finnish students from an upper secondary school located in southern Finland. At the time the data were collected, they were all students taking an elective course (ENA8) that focuses specifically on spoken English, in order to develop their oral language skills. I have discussed the aims of the course in question previously in the introduction. Recordings were gathered from 40 students in total, but due to some students withdrawing their consent from the study, and some admitting to having speech disorders that would have negatively impacted the reliability of the study, I have narrowed down the number of participants to 33. As the recordings were gathered before and after the spoken course, it means that overall, there are 66 speech samples to analyse (~1-3 minutes each) for changes in the use of stalling mechanisms. All the students reported Finnish as their native language and out of the 33 participants, 11 were male and 22 were female. Most of the students participating in the study were already 18 years old, only four of them being 17. According to guidelines concerning informed consent in data collection, all research must adhere to the general data protection regulation of EU (GDPR 2024). According to GDPR, children under 13 can only give consent with a parent's permission. The data for this study were gathered in 2014, which was before GDPR, but even so, informed consent was gathered for the study and all of the participants could give consent themselves.

A background questionnaire administered for the participants also asked for information such as how many English courses they had already completed up until that point in time, and the grade they had received from their last English course. All but two students had completed eight courses during their studies in upper secondary school, the two other ones having studied seven courses. According to this, the amount of English the participants had studied during their schooling is virtually the same. The range of the grades was from 6 to 10 with an average of 8.5. Overall, the group of participants is quite homogenous according to the background information, which could

possibly mean that the results will give a more accurate and unified point of view into Finnish upper secondary school students' development during a spoken English course such as this.

3.2   Data collection

The qualitative data consisted of recorded oral language samples gathered during two separate instances: at the start of the 7-week spoken course on the 8th of October 2014, and at the end of the same course on the 26th of November. There were two separate groups, but both were taught by the same teacher and followed the same contents during the course. One of the groups had 19 students and the other one had 21 students overall. However, as discussed in 3.1, there were some instances where it was necessary to exclude some participants from this study.

The questionnaire was also used for the purpose of linking the audio recordings with the right participants and their background information. All the participants were given a personal subject code (alphabetical and numerical) that was linked to the audio files so that not even the researchers could identify specific students. The students' names or other identifying and personal information were not asked at any point of this study, and participating in the study remained anonymous, as per good research ethics in non-experimental research (Byrne 2016). Participating in the study was completely voluntary. The participants could also withdraw their consent at any time, without any negative consequences as well. Due to this present study being given access to use the data that were gathered, I myself had no participation during the testing phase itself. However, I was given appropriate permission to the data by the University of Turku and have ensured that the handling of it is in accordance with proper research ethics.

A structured narration elicitation task was used to elicit speech samples from the participants, the task being a story retelling from Finnish to English. A structured task means that all the participants were given the same predetermined task in the form of a short story and explicit instructions, in order to ensure consistency in data collection. The task instructions asked the participants to imagine they saw a story in the newspaper that they wanted to tell their foreign friend about. The participants had three minutes to read and familiarize themselves with the text and approximately three minutes to retell the story in English. The text in question was taken from Maija Saleva's doctoral dissertation (1997, 93) and was the same before and after the course (see Appendix 1). To use a task that researchers have used previously to study fluency increases the validity of the study, as the task has already been used in successful research.

3.3 Methods of analysis

The present study is a mixed-methods analysis on the development of stalling mechanisms. This means that the data are analysed both quantitatively and qualitatively. Although contemporary studies concerning L2 oral fluency are mostly quantitative studies focusing on group-level phenomena, previous studies of the same nature have concluded that a mixed methods analysis might suit this type of a study better, in order to achieve a better understanding of learners' L2 oral fluency. As mentioned in section 2.3, Peltonen and Lintunen (2016) have concurred that a quantitative analysis is a suitable method for pausing phenomena, but a qualitative analysis can uncover differences in the use of filled pauses, fillers, and repair aspects of fluency. For these reasons, I will analyse the data using both methods for the purpose of examining both group level tendencies and individual differences as well. The methods of analysis are detailed next.

Researcher-imposed coding (Mackey & Gass 2011, 209) was utilized to analyze the data efficiently and accurately. First, all the audio recordings were transcribed in detail into written format and checked for accuracy. The transcription process was aided by creating rough drafts using a speech-to-text computer program *transcribe* from Microsoft, and then manually adding annotations for stalling mechanisms and pauses. AI was not used in the processing of research data as security is difficult to achieve with it. Instead, Microsoft transcribe was used because it is a tool that has been used in similar research before to create rough drafts for further annotation, and data remained secure because the files are processed locally. Concerning pausing and stalling mechanisms, this was then accompanied by a more detailed analysis using the software program Praat – an audio analysis tool (Boersma & Weenink, 1992). To aid in measuring silent pauses, a ready-made Praat script based on the article: "Praat script to detect syllable nuclei and measure speech rate automatically" (De Jong & Wempe 2009) was used, and after that they were checked for accuracy. This was accompanied by a manual annotation of stalling mechanisms for their duration, frequency and distribution. Frequency means the total number of occurrences and distribution means the location of the stalling mechanisms and whether they occur in mid-clause or end-clause positions. An example transcription has been provided in the appendices (Appendix 2). Annotations, for both the quantitative and qualitative analyses, that were made in the transcriptions fell into two different categories:

1. Duration, frequency, and distribution of silent pauses.
2. Frequency and distribution of other stalling mechanisms (drawls, filled pauses, filler words and repetitions).

The present study applies five fluency measures for quantitative data analysis to examine development in the use of stalling mechanisms. The chosen fluency measures are presented in Table 2. Silent pauses were examined for frequency because exhibiting frequent silences during speech is seen as less fluent. Silences longer than 0.25 seconds were considered silent pauses, silences that were shorter were considered micropauses and were not included in the analysis. Filled pauses (hesitations such as 'um', 'uh', 'er') were recorded for their frequency, as were the other stalling mechanisms: drawls, filler words and repetitions. It is important to add that, after a series of analyses of the audio recordings, I have elected to divide filled pauses and filler words further into two separate categories because of Finnish language transfer. In my analysis, I discovered that the students used both Finnish and English filler words, and phonemes usually not found in the English language such as /ø/, a sound most commonly uttered as 'öö'. These have been checked for statistical significance separately, to examine if there are changes in the use of Finnish while speaking English.

Before I could check for statistical significance in the data, I had to calculate the total number of each stalling mechanism and the total speaking time for both before and after the course. Firstly, the total number of these stalling mechanisms was calculated manually from the annotated transcriptions. After that, the total number was then divided by the total length of the recording in seconds and multiplied by 60. The average length of the recordings before the course came up to 99.64 seconds, while after the course, the recordings were significantly shorter, the average being 78.91 seconds. By calculating how many stalling mechanisms the students exhibited in a minute, any uncertainties about the difference in total length were erased. The quantitative analysis for group level tendencies is accompanied by a qualitative study even though much of the data gathered qualifies as both quantitative and qualitative data because: "Analysis in qualitative research is not a stage of research but begins during data collection and continues throughout the research process" (Mackey & Gass 2011, 191). This means that gathering the audio recordings and transcribing them into written format is already qualitative analysis of the data.

Table 2. Fluency measures selected for this study

Fluency measure	Operationalization
Number of silent pauses per minute	The total number of silent pauses over 0.25 seconds in duration
Number of filled pauses per minute	The total number of non-lexicalized filled pauses such as *'um', 'uh', 'er'*

Fluency measure	Operationalization
Number of drawls per minute	The total number of sound-lengthening in hesitation
Number of filler words per minute	The total number of filled pauses such as *'like', 'I mean', 'well'*
Number of repetitions per minute	The total number of words or longer utterances that were repeated immediately without modification

3.3.1 Quantitative analysis

I analyzed the data quantitatively by utilizing descriptive and inferential statistics with the help of SPSS 29. I compared the frequency of pauses and other stalling mechanisms in the pre- and post-course recordings to provide insight into the general tendencies that the participants had on a group level. Before examining the data for statistical significance, each stalling mechanism had to be individually checked for normal distribution, which was done by utilizing the Shapiro-Wilk Test of Normality. The test uses a p-value of > 0.05, meaning that any p-value above that is considered normally distributed (Larson-Hall 2016, 120). This test is significant in ensuring the validity of the results of these tests and also in how the data is processed afterwards: if the data are normally distributed, parametric tests are appropriate to use for further analysis. However, if the data are not normally distributed, then we must use non-parametric tests.

Additionally, this study examines the same group after a period of time instead of data from two independent groups, which indicates the need for the Paired Samples t-test instead of the independent samples t-test (Larson-Hall 2016, 269) when searching for statistically significant differences in the data. This test is often used when there are two measurements taken from the same group. For the Paired Samples t-test the p-value needs to be lower than 0.05 to be considered statistically significant. Table 3 shows the test of normality p-values of every stalling mechanism, and they were calculated both before and after the 7-week course. As can be seen, only silent pauses did not deviate from normal distribution, which means that I will be using the Paired Samples t-test. However, filled pauses, drawls, filler words and repetitions were not normally distributed, meaning we must use a different, non-parametric test for statistical significance called Related-Samples Wilcoxon Signed Rank Test (Larson-Hall 2016, 286).

Table 3. Test of Normality p-values of stalling mechanisms

Stalling Mechanism:	Shapiro-Wilk Test of Normality p-value, before the course:	Shapiro-Wilk Test of Normality p-value, after the course:
Silent pauses	.639	.096
Filled pauses	< .001	.029
Filled pauses (Finnish influence)	< .001	< .001
Drawls	.024	.017
Filler words (English)	< .001	< .001
Filler words (Finnish)	< .001	< .001
Repetition	< .001	< .001

3.3.2 Qualitative analysis

In order to analyze the data more comprehensively, three students with contrasting performances were selected to compare the least and the most fluent students from the group. Three cases were selected, because if there are too many cases, then "less intensive scrutiny and presentation of each one are possible" (Friedman 2011, 108). This added to the fact that analyzing 66 speech samples qualitatively is not realistic for the scope of this study and too time-consuming means that fewer cases were more appropriate. Examining students that represent extremes, such as in Dörnyei's "extreme case sampling" (Dörnyei 2007, 128), it is possible to inspect the whole fluency range of the group and have a clearer view of individual differences as well. Two cases were selected based on the extreme case sampling by Dörnyei (2007) and the third case was selected because they showed the most progress during the course, which demonstrates that significant improvement during a short period of time is achievable.

The three cases were chosen based on the quantitative data. Using excel, calculations were made to examine how many stalling mechanisms each participant had before and after the course, in order to determine the best candidates for the qualitative analysis. From the group, one case represents the participant that struggled the most with speaking by not being able to produce more than a few seconds of speech before the course. One participant was the most constant during the course in terms of time spent speaking and the total number of stalling mechanisms, and the third one was a student that had the most improvement overall during the course. According to the quantitative study, they improved their performance by 37.4 %. After analyzing the transcriptions quantitatively and the three cases for the qualitative analysis were chosen, the transcriptions were then analyzed qualitatively in more detail for pauses and other stalling mechanisms. The qualitative data analysis was done by examining the same stalling mechanisms that I used for group level

changes but focusing on more than their frequency. This is done by analyzing the distribution and location of them as well, while comparing pre- and post- course speech samples for individual variation and changes.

4   Results

In this section, I present the results of my mixed-methods study. First, I will start with the results of the quantitative analysis based on group-level tendencies and changes in section 4.1. After that, I present the results of the qualitative analysis in section 4.2 to give further insight into individual development and differences in the use of stalling mechanisms and pausing phenomena.

4.1   Group level changes

For group level differences, I examined 5 different stalling mechanisms. Each of the stalling mechanisms and total number of them per minute are presented in Table 4. The total number for most of these stalling mechanisms has decreased during the course, but it is important to check for statistical significance of each of them. To examine the statistical significance of silent pauses after the test for normality, a Paired Samples t-test was conducted. These values gave a p-value of .003, marking it as statistically significant. This indicates that the use of silent pauses has decreased during the time of the course (see Figure 3). From the figure it can also be seen that there is only one outlier with a relatively high frequency of silent pauses after the course, but because the same outlier is not present before the course, individual development of this participant cannot be seen from this figure. However, it can be seen that on average, the group spent less time in complete silence after the course.

Table 4. Total number of stalling mechanisms per minute

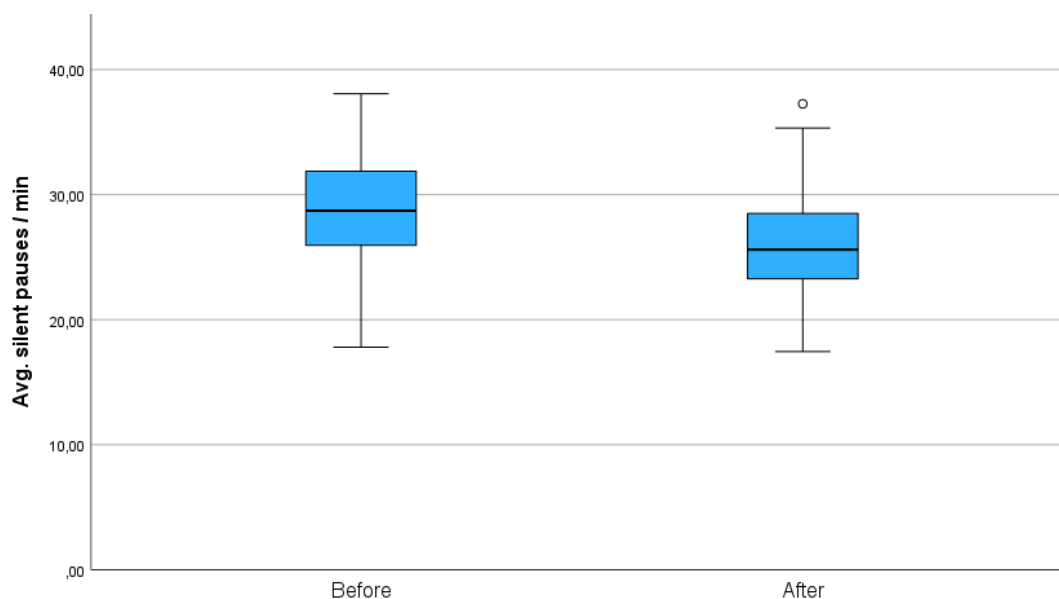| Stalling Mechanism	Group average per minute of recording, before the course	Group average per minute of recording, after the course
Silent pauses	29.03	26.57
Filled pauses	4.71	4.59
Filled pauses (Finnish transfer)	1.42	0.78
Drawls	5.33	3.89
Filler words (English)	0.35	0.48
Filler words (Finnish)	0.07	0.14
Repetition	0.95	0.90

Figure 3. Boxplot for average silent pauses per minute

There was substantial within-group variation in the use of filled pauses. Whereas some students only had a few cases of using filled pauses, some students preferred them over other stalling mechanisms and used them frequently, some over 20 times in their recordings. The group average of filled pauses before the course was 7.82 with the standard deviation, SD = 7.69. After the course the group average was 6.03 (SD = 4.99). For filled pauses with Finnish transfer, the group average before the course was 2.36 (SD = 4.70) and after the course it was 1.03 (SD = 2.62). The standard deviation being close to – or over – the average means that the results are widely spread out over the average and that there is a high level of variation. In terms of development, this signifies that during the course the group average decreased as did the standard deviation, meaning that the students used less filled pauses after the course. There were fewer outliers as well, which means that those who used filled pauses frequently during one recording used them less during the other. It is important to note that not all students used filled pauses less after the course. Some of them used very few filled pauses before the course, but increasing their usage after the course, indicating that there is individual variability in the use of stalling mechanisms as well.

For filled pauses – and filled pauses with Finnish transfer, non-parametric tests to calculate for statistical significance were used. For this, I used the Related-Samples Wilcoxon Signed Rank Test. While both gave a p-value of over 0.05, marking them as not statistically significant, I have reason to believe that with filled pauses with Finnish transfer, it might be a false negative result. As can be seen from Table 5, the p-values of both categories are over 0.05. However, when looking at the raw numbers for filled pauses with Finnish transfer, the number before the course was 78 and

after the course it was 34. Then, calculating how many of these filled pauses per minute they had respectively, it can be seen from Figure 4 that after the course the number has significantly decreased, with the students making almost half the number of /ø/ sounds than they did before. This added to the fact that 'regular' filled pauses stayed virtually the same – 4.71 per minute before the course and 4.59 after for a 2.5% decrease – means that there has been some improvement in this regard with Finnish transfer decreasing during the course.

Table 5. P-values of filled pauses

Stalling Mechanism:	Related-Samples Wilcoxon Signed Rank Test p-value
Filled pauses	.964
Filled pauses with Finnish Influence	.124
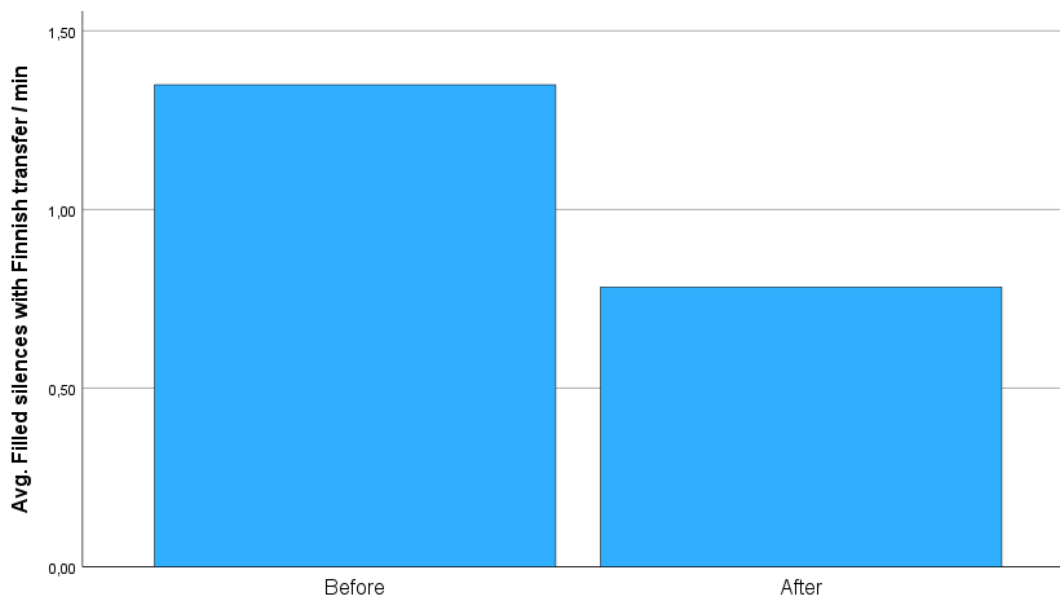
Figure 4. Bar chart for average filled silences with Finnish transfer per minute

Drawls had notable variation inside the group as well, similar to filled pauses. Even though all students used drawls, most of them used them less after the course. And just as filled pauses, some students only used them a few times, while others used them as frequently as 24 times before the course and 16 times after. Before the course the group average of drawls was 8.82 (SD = 4.74) and after the course it was 5.88 (SD = 4.70). These number indicate that the use of drawls has decreased during the course. However, after the course the standard deviation is greater, which means that

there is a higher level of variation in the use of drawls. This is highlighting the fact that there are significant individual differences in the use of stalling mechanisms. The change in the use of drawls is statistically significant as well. The test of normality indicated that it is not normally distributed, which called for the use of a non-parametric test. The Related-Samples Wilcoxon Signed Rank Test calculated the p-value to be .004, which is a significant result. From Figure 5 it can be seen that the participants used fewer drawls after the course, but there are students that deviate from the group average.

For filler words, Finnish filler words and repetitions, the Related-Samples Wilcoxon Signed Rank Test p-value was over .05 for all of them, marking them as not statistically significant (Table 6). The number of repetitions stayed almost the same before and after the course and filler words were used the least of all stalling mechanisms by the students. Only some students used filler words, and usually they were the same ones before and after the course. Finnish filler words did increase after the course by 30 %, total number increasing from 4 to 6. English filler words such as: 'well', 'yeah', 'like', 'okay', 'anyway', 'you know', 'I mean', 'I guess' were used, however, there were some profanities as well. Finnish filler words that the students used were short words such as 'eiku', 'eikä ku', 'niinku', 'jotain', 'ei vaa'. There were some profanities in Finnish as well. All the profanities were calculated as filler words in both categories because of their disruptive nature.
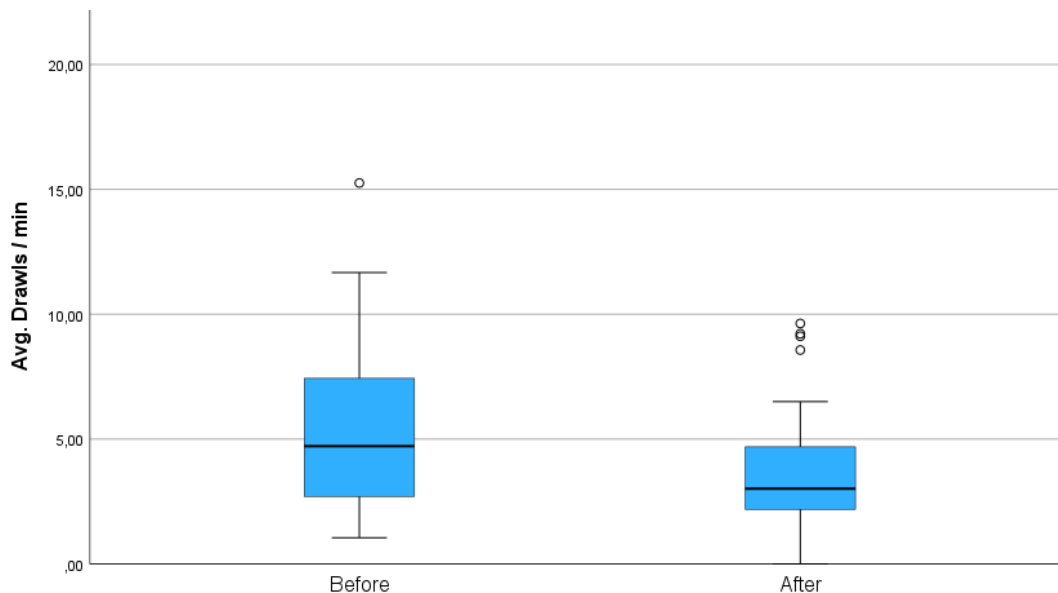

Figure 5. Boxplot for average drawls per minute

Table 6. P-values of filler words and repetitions

Stalling mechanism	The Related-Samples Wilcoxon Signed Rank Test p-value
Filler words	.480
Finnish Filler words	.249
Repetition	.611

4.2 Individual variation

The three cases were selected by analysing the quantitative data for participants who fit the criteria for the extreme case sampling and then one more who, purely numerically, had the most progress during the course. These numbers were then contrasted with their transcriptions to see if they were a good fit for the qualitative analysis. I will be using the original participant codes for these students that were used in data gathering as well. Case 1: OCS-MO7 was chosen as they struggled the most before the course, but then improved significantly during the course by producing almost four times the speech they did at the beginning. When the recording before the course only yielded 6.19 seconds of speech, the total speaking time after the course was 23.71 seconds (Table 7). This participant is a good choice for qualitative analysis because purely numerically speaking, the quantitative data marked them as having more stalling mechanisms after, which made it seem like they worsened during the course. However, by producing more speech, they applied more stalling mechanisms in order to keep the flow of speech going and getting their message across. This already tells us that they improved during the course.

Case 2: OCS-FO6 was numerically one of the most consistent in their performance during the course. The participant's total length of the recordings and total speaking times stayed constant during the course, deviating only slightly from the other (Table 7). When choosing the participant with the most consistent performance, instead of examining only the frequency of existing stalling mechanisms, their location and duration was considered as well. While there were students who, on average, used fewer stalling mechanisms during the course, case 2 exhibited shorter silent pauses, and stalling mechanisms were located mostly in end-clause positions rather than in mid-clause positions. These factors determined that case 2 was chosen for the qualitative analysis. Case 3: OCS-F18 was the one with the most improvement (37.4 %) during the course when the different stalling mechanisms were calculated per minute, making them an interesting case for the qualitative analysis for individual variation. I will be analysing these cases individually starting with Case 1.

Table 7. Total lengths and speaking times before and after the course

Participant code	Total length before the course	Total speaking time before the course	Total length after the course	Total speaking time after the course
Case 1: OCS-MO7	23.60s	6.19s	67.30s	23.71s
Case 2: OCS-FO6	70.24s	53.41s	66.09s	49.42s
Case 3: OCS-F18	92.57s	62.43s	68.37s	52.50s

On its own, there is not much to analyse in the pre-course recording of case 1, and example 1 shows the complete transcription before the course. However, when we examine the post-course recording as well, we can see that improvement has happened in total time spent speaking. As mentioned before, before the course there was only 23.60 seconds worth of recording of which 6.19 seconds was spent talking. This means that most of that time was spent in silence, and as we can see from the transcription (Example 1), the duration of individual silent pauses is quite long, with the last one being over 7 seconds. In the transcription, drawls are marked with colons and short silent pauses that are less than 0.25 seconds long are marked with periods inside parentheses. We can see that case 1 elongates their vowels in the beginning to obtain more time to figure out what they want to say. Interestingly, they exhibit short silent pauses between words before the course as well, such as *wo(.)ma:n* and *para(.)chu:te*, however these have disappeared completely from the after course recording, giving way for a more fluent enunciation of words.

(1)   (. . . .) Wo(.)ma:n from (0.30) Florida: (3.66) did para(.)chu:te jump:ing on he: she:s (1.85) p- eighty-six (1.54) year (1.20) birthday party (1.66) {umm} (7.20) *h* *background noise*

After the course, case 1 spent significantly more time speaking. Even though the recording itself is 67.30 seconds long and they talk only 23.71 seconds out of that, time spent talking is still four times more than before the course. The frequency of silent pauses has increased – with 17.8 per minute before and 29.4 per minute after the course, however with more speech this can be expected. The duration of most of them is shorter in comparison to before the course. In addition to this, the longest silent pauses are in end-clause positions with the average duration being 2.14 seconds, while mid-clause pauses are mostly shorter in duration, with the average being 1.10 seconds. There is some variation is the use of stalling mechanisms as well. When before the course they mostly used sound lengthening, in the after-course recording they have fewer drawls but have used repetition and filled pauses more. In the post-course recording they also have a

longer silence before a filled pause and then one immediately after it as well, probably trying to fill the silence with sound so it is not perceived as taking so long (Example 2).

(2)　… (3.04) {um} (0.83) She jumped because (0.86) wanted (1.97) {um} (0.44) money for (0.87) veteran (2.06) sport (0.38) Olympics. (5.98) {um} (5.61) she didn't keep (1.17) {umm} (1.70) jump (1.87) was danger:ous…

Case 2 is at the other end of the extreme case sampling because of how fluent and similar both recordings were, with few significant changes between the beginning and the end of the course. Times spent talking were only a few seconds apart – with 53.41 seconds before the course and 49.42 seconds after – and this subtracted from the total length of the recordings tells us that the total length of silent pauses were 16.83 seconds and 16.67 seconds respectively. The frequency of silent pauses was similar as well with 29 before and 26 after the course, and when calculated how many there were per minute, the numbers come out to 24.8 before and 23.6 after the course. However, when examining the duration of these silent pauses, it is worth noting that while before the course there were slightly more frequent, on average the duration of them was slightly shorter, with the average length being 0.58 seconds before and 0.64 seconds after. Regarding the distribution of silent pauses, from the transcriptions it is evident that while the end-clause silent pauses have remained the same, there are somewhat fewer mid-clause silent pauses.

There is some shift in other stalling mechanisms as well. While before the course the number of filled pauses was 9, after the course, there was only one filled pause, a phoneme /ø/. Filled pauses were always accompanied by a silent pause either before or after it, or both. Drawls stayed almost the same, with 6 occurrences before the course and 4 after the course. It is interesting to note that while before the course, the drawls were by themselves without any pausing near them, after the course the drawls were followed by a silent pause. Repetitions increased during the course. While before case 2 did not have any repetition, after the course they used repetition three times, repeating words ‘*with*’ and ‘*book*’, and ‘*during the seven*’ towards the end of the recording. There were also longer silent pauses accompanying these repetitions. From the post-course transcription (Example 3), it can be seen that case 2 had the same sort of short silent pause in the middle of the word ‘parachute’ as did case 1. Overall, there is some variation in stalling mechanisms and improvement in silent pauses during the course. However, case 2 has stayed quite consistent in their performance, which highlights the more fluent end of the spectrum quite nicely.

(3) …She had a master jumper with her (0.59) during the (0.50) seven (1.06) during the seven-minute jump a:nd (0.92) according to: (1.14) the Guinness (0.35) world record book- book (0.67) öö joyce is the oldest parac(0.26)chute jumper…

The final participant (case 3) for the qualitative analysis was the student that demonstrated the most improvement according to the quantitative analysis, improving their performance by 37.4 %. However, pure numbers cannot explain exactly how they have improved, which makes case 3 a good candidate for qualitative analysis. Before analysing the stalling mechanisms separately, it is important to note that even though the total length of the post-course recording is 24.2 seconds shorter than the pre-course recording, the contents of the speech have stayed consistent. Meaning that, even though case 3 had a shorter recording after, the contents were the same and no information has been omitted that would explain a shorter total length of recording. This indicates that something has changed in how fluent they are, and it is best demonstrated by examining the same section from the pre-course transcription (Example 4) and post-course transcription (Example 5). Whereas speaking this part lasted 26.6 seconds before the course, it only lasted 15.2 seconds after the course. At first the total length of silent pauses in this section is 8.07 seconds and after the course only 2.74 seconds. The number of other stalling mechanisms is also greater in the pre-course speech sample. There are clusters of filled pauses – some with Finnish transfer as well – with drawls and some repetition that are used to compensate for disfluencies such as mid-clause silent pauses. These also contribute to fluency while searching for the right words to use. All of these are lacking from the post-course speech sample, and this all contributes to a more fluent output.

(4) *pt* *hh* Joyce is a skilled (0.45) uh golf and tennis player? (0.34) who: (0.86) raised (0.86) uh money: (0.79) *pt* for the Olympic <uh who raised money with the jump> (0.42) for the Olympics of (0.58) old (0.30) athletics (0.74) uh she has always been a: hardass woman (0.96) ää (0.37) while she w- (0.40) öö she worked as a crime reporter in (0.47) Chicago Tribune? Tribune? Magazine? (0.53)

(5) *pt* Joyce is a skilled golf and tennis player who raised money for the Olympics of veteran athletics (0.39) with her jump. (0.86) She has always been a tough woman-woman. (0.74) While working as a crime journalist in (0.29) Chi-Chicago Tribune (.) magazine (0.46) *h*

There has been a significant change in the use of silent pauses for case 3. Even though case 3 spent less time speaking after the course than before (9.93 seconds), they still spent less time in silence after the course. The total length of silent pauses before the course was 30.14 seconds and

after the course 15.87 seconds. Before the course the number of silent pauses per minute was 29.8 and after the course 22.8, indicating that there has been a decline in the frequency of silent pauses. Concerning the duration of the silent pauses, they have been very consistent during the course, with an average of 0.63 seconds before and 0.61 seconds after. Even though the frequency of silent pauses has changed significantly, an even greater change has happened in the distribution of them. Before the course, there was significantly more mid-clause pausing, but in both recordings end-clause silent pauses have stayed the same. Additionally, in the post-course sample, mid-clause pauses were all less than 0.64 seconds long, whereas before the course some were as long as 1.14 seconds. This confirms that the improvement that has happened in silent pauses has happened in mid-clause pausing, contributing to fluency as well.

There was a slight decrease in the use of filled pauses with 5.83 per minute before and 5.27 after, however the use of the phonemes /ø/ and /æ/ has disappeared altogether from the post-course speech sample. Additionally, whereas before, case 3 used sound lengthening very frequently to bide for more time (11.02 times per minute), they used them only a few times after the course. Repetition has stayed the same with a few instances of use. Overall, there is more frequent hesitation in the pre-course speech sample that is not as frequent after the course, indicating that case 3 has improved significantly in speaking more fluently.

5 Discussion

In this section, the results of the present study are discussed in detail and in the light of previous studies. Group level changes are discussed first, after which individual variations in the use of stalling mechanisms are considered in more detail. The results are related to the research aim of the present study of what sort of variation or changes there are for stalling mechanisms on both group and individual levels during a 7-week spoken English course. Starting with the results for the frequency-based quantitative study, each stalling mechanism is discussed separately. After this, the qualitative results of the present study are discussed to complement the quantitative study, and to provide more insight into individual changes and improvements.

The results for the quantitative analysis for group level changes were mixed. For the present study, five stalling mechanisms were analysed for frequency before and after the course, however as said before, filled pauses and filler words were divided further into two separate categories because of Finnish language transfer that some students exhibited. Before discussing the individual stalling mechanisms, it is noteworthy that overall, as a group, the total length of the recordings was significantly shorter after the course than at the beginning. Whereas before the course the average length of the recordings was 99.64 seconds, after the course the average had decreased to 78.91 seconds. Most of the recordings were shorter after the course, with only a few that were close to the same length and only a couple that were longer. At this stage, it is difficult to draw any conclusions as to why this is. However, there have been some themes in the results worth investigating as to the reasons why.

The difference between the use of silent pauses before and after the course was statistically significant, meaning that on average, the students used fewer silent pauses after the course was over. This could contribute to a shorter total length of the recordings after the course. Whereas before the course, the total number of silent pauses per minute was 29.03, after the course it had decreased to 26.57. Statistical significance was not found for filled pauses, although, for filled pauses with Finnish transfer, there is reason to believe it was a false negative result. While 'regular' filled pauses remained the same, the total number of /ø/ sounds was reduced by 50% by the end of the course. There could be some explanations for the false negative result, mainly that my sample size might have been too small or the differences in my data might be too inconsistent for the test to work. After all, only some students exhibited filled pauses with influence from the Finnish language, with the majority either showing no change or changes that were too minimal to count. This could mean that the differences were not consistently distributed around the median and affected the outcome of the test. To further demonstrate this, the standard deviation for filled pauses

with Finnish transfer was large both before and after the course, indicating that the results were widely spread out around the average.

The least used stalling mechanism was filler words and the usage of them remained the same during the course. 15 students out of 33 used filler words and usually the same person used them before and after the course, highlighting the fact that in the use of stalling mechanisms, there are personal preferences as well. Concerning drawls, even though sound lengthening is seen as a more effective way of holding the floor and maintaining fluency than silent and filled pauses (Dörnyei & Kormos 1998, 370), the change in the frequency of drawls is statistically significant as well, and the results show that on average the group used less sound lengthening after the course. There are some students that deviate from the group average by not having any significant changes in this regard. Statistically significant changes in the use of filler words and repetitions were not detected after the course. Both remained virtually the same, meaning that even though drawls were used less after the course, none of the other stalling mechanisms were used more to compensate for it either. The use of drawls is simply less frequent after the course. To speculate why this is, it could be that the task being new at the beginning of the course might have increased the use of drawls at first. However, by the end of the course the task was already familiar and the participants more comfortable and confident in their capabilities and used sound lengthening more thoughtfully.

In the qualitative analysis, individual variation was examined in more detail for the purpose of providing more insight into the findings of the frequency-based quantitative analysis. The aim was to focus on aspects that were not possible to examine purely through numbers. In addition to this, the qualitative analysis demonstrated the variation inside the group, and by using extreme case sampling (Dörnyei 2007), the fluency range among the students was inspected. By analysing case 1 who struggled the most and case 2 that had one of the most fluent performances with little to no variation in the use of stalling mechanisms during the course, it is evident that there is substantial variation between students and their fluency levels within this group. It also highlights that not all students progressed at the same rate or in the same way during the course.

Case 1 is a good example of how a quantitative analysis can indicate that someone has worsened in their performance, when in fact it is quite opposite. It is only logical that as they spoke more, their use of stalling mechanisms increased as well, to help them get their message across. Numerically they used every stalling mechanism except for drawls more after the course. However, what we can take away from case 1 is that after the course, they applied themselves more and utilized these problem-solving mechanisms as an aid to speak longer (see also Dörnyei & Kormos 1998), which indicates that they have improved during the course. Case 1 goes to show that the course has been helpful in improving their fluency and capability in speaking a foreign language.

Case 2 gave a very consistent performance both times. When choosing this participant for the qualitative study, instead of only examining the frequency of silent pauses, the duration and distribution of them was also considered. While the frequency of silent pauses was higher than some students', the duration of them was significantly shorter. This is in line with some studies that have concluded that the frequency of silent pauses might not correlate with fluency, but the duration of them does (van Gelderen 1994). Case 2 represented the other end of the extreme case sampling by giving a very consistent performance both times. They represent the more fluent end of the spectrum, and with case 1 and case 2, it is apparent that there is significant variation in the use of stalling mechanisms and fluency levels within the group.

Case 3 was the student with the most improvement during the course. Whereas before the course pauses and hesitations tended to occur in clusters, similar to other conducted studies (e.g. Freed 1995, 2000) and between syntactic units (e.g. Lennon 1990), after the course, they mostly exhibited pausing between major syntactic boundaries. There are also strong correlations between the frequency, duration and location of silent pauses and fluent speech (Götz 2013, 13), and with this case, the frequency and distribution of silent pauses improved significantly, indicating that they improved their fluency during this course. In the pre-course recording, case 3 used stalling mechanisms during word searches and to avoid spending time in silence in mid-clause positions. After the course, they still used stalling mechanisms for these purposes, although not as frequently. As speculated earlier, it could be a case of gaining more confidence during the course and knowing the text already that also contributed to maintaining fluency and improving as much as they did. The communicative nature of the course might have given them tools to improve their confidence as a communicator as well.

The data showed that there are significant differences in the use of stalling mechanisms. Individual variation was great in the use of drawls and filled pauses, which is in line with previous studies that have studied stalling mechanisms and repairs (e.g. Götz 2013; Peltonen 2020; Peltonen & Lintunen 2024). Added to this, it is safe to say that there are preferences in the use of stalling mechanisms, which is in line with previous studies as well (e.g. Götz 2013; Wolk et al. 2020) For example, filler words were mostly used by the same students during the course. From the qualitative analyses it is evident that stalling mechanisms have an important role in improving fluency and compensating for disfluencies such as silent pauses. They also help during searching for appropriate words. However, as case 3 showed, it is possible to improve in this regard as well. Even though the frequency of silent pauses and other stalling mechanisms lessened during the course, also the location of them changed. Pauses were located mainly at end-clause positions and stalling mechanisms such as drawls and repetitions were used less in mid-clause positions.

The findings from the quantitative analysis suggest that some group-level changes can be detected in relation to the use of stalling mechanisms. The frequency of silent pauses and drawls decreased as did Finnish influence on filled pauses. On other stalling mechanisms, however, the difference was too minimal to be deemed significant. A previous study on the matter has concluded that gaining more confidence and developing more trust in their capabilities, students become more fluent as well (Maseda & DeFeo 2014). Additionally, with confidence being linked to cognitive fluency and the ability to repair the situation when breakdowns occur (Housen, Kuiken & Vedder 2012), the combination of an already familiar text and increase in confidence might be crucial in explaining the changes in the use of stalling mechanisms. Although, as I was not specifically able to examine confidence as a factor, this is purely speculative. The data collection used the same story retelling elicitation method and the same text before and after the course, and it could be that using a familiar text can be a contributing factor for less hesitation, because the participants already knew what the text was about. This coupled with the aims of the spoken English course of providing tools for language development and providing communicative tasks for the students (LOPS 2019), and changes – and improvements – in confidence and fluency can occur.

6   Conclusion

The present study explored the development of fluency and the role of stalling mechanisms as problem-solving mechanisms during a 7-week spoken English course at a Finnish upper secondary school. The study examined both group-level tendencies the students exhibited and individual variation in the use of stalling mechanisms, with the focus on understanding how short-term structured teaching impacts students' L2 oral fluency. The data consisted of 66 speech samples taken before and after the course and combined a quantitative frequency-based analysis focusing on the group, and a qualitative analysis of three cases that were selected using Dörnyei's (2007) extreme case sampling. A mixed-methods approach was applied to answer the two following research questions: "What kind of group-level variation and changes are there in the frequency of silent pauses and other stalling mechanisms" and "What kind of individual changes and variation are there in the use of stalling mechanisms".

On a group level, the findings revealed that overall, the total length of the recordings was significantly shorter after the course. There were improvements in the use of silent pauses, with the students remaining in silence less frequently after the course. This indicates that the students were better able to maintain the flow of speech, which is a crucial aspect of fluency. There were also statistically significant changes in the use of drawls, the students using sound lengthening significantly less after the course. There was no increase in any of the other stalling mechanisms to compensate for this, meaning that the usage of drawls was more compacted after the course. The transfer of Finnish in filled silences (phoneme /ø/) was less frequent after the course although 'regular' filled silences (sound such as 'uh', 'umm') stayed the same, indicating that Finnish transfer in this regard lessened during the course. However, Finnish filler words did increase by 30 %. There were no notable changes to English filler words or repetitions on a group level. The important changes to silent pauses and drawls could explain why the total length of the recordings was shorter after the course, however the reason might consist of multiple different factors, such as increase in confidence or the material being familiar to the students, making them more comfortable during the second testing.

The qualitative analysis complemented the quantitative analysis by examining one of the least fluent students and one of the most fluent students with the most consistent performance, with the third case being the student that had the most improvement in their utterance fluency. The extreme case sampling (Dörnyei 2007) made it possible to inspect the whole fluency range of the group and have a clearer view of individual differences as well, while case 3 represented how much improvement is achievable during a short period of time. The qualitative analysis indicates that the

students did not progress at the same rate or in the same ways. Some students such as case 3 demonstrated remarkable development in their fluency, while others – although advancing in their skills – still continued to struggle after the course.

The study has some limitations, main ones being small sample size and external variables that could influence individual variation and development. A small sample size means that the findings cannot represent the whole of Finland's upper secondary school students, and this could limit the generalizability of the results to other student groups or learning contexts. Nevertheless, these results give valuable insight into Finnish upper secondary school students' use of problem-solving mechanisms and how much variation there can be inside the classroom and can inform instructors of their level of fluency. External variables might also play a role in how much students develop during a short course such as this. The study cannot account for students' exposure to English outside the classroom, their motivation to learn or anxieties related to speaking English. These factors could impact the results but are difficult to measure accurately or control. Addressing these limitations of the study, future research on the matter could opt for more 'triangulation' in the approach by incorporating questionnaires for the students before and after the course, to receive their input on their own learning as well. Additionally, perceived fluency (Segalowitz, 2010) could be more incorporated by having the teachers' input on the students learning as well. This study could also be replicated for university students studying English after they have had more instruction on spoken English, to see if differences in stalling mechanisms occur.

In conclusion, the aim of this study was to give insight into how fluency develops during a shorter timeframe. The results can notify teachers on how to support learners with their spoken L2 performance and help them overcome challenges regarding fluency. While the 7-week spoken English course was able to enhance fluency in some measures on a group level, the qualitative analysis revealed how significant individual variation inside the group actually is, and how varied the whole fluency range inside a group such as this can be. The focus of a short, spoken English course such as the elective offered at Finnish upper secondary schools, might be suitable for developing the students' confidence in communicating in another language, which in turn develops their fluency as well.

References

Abrahamsson, Niclas, and Kenneth Hyltenstam. 2009. "Age of Onset and Nativelikeness in a Second Language: Listener Perception Versus Linguistic Scrutiny". *Language Learning* 59 (2): 249–306. https://doi.org/10.1111/j.1467-9922.2009.00507.x.

Boersma, Paul, and David Weenink. 2024. *Praat: doing phonetics by computer* (version 6.4.07) [computer program]. Retrieved from http://www.praat.org/.

Byrne, David. 2016. *Research ethics.* Los Angeles, CA: SAGE Publications, Inc.

Common European Framework of Reference for Languages. 2024. Accessed 15th of April 2024. Retrieved from https://www.coe.int/en/web/common-european-framework-reference-languages/table-3-cefr-3.3-common-reference-levels-qualitative-aspects-of-spoken-language-use.

De Jong, Nivja, H., and Ton Wempe. 2009. "Praat script to detect syllable nuclei and measure speech rate automatically". *Behavior Research Methods* 41: 385–390. https://doi.org/10.3758/BRM.41.2.385

De Jong Nivja H., Rachel Groenhout, Rob Schoonen, and Jan H. Hulstijn. 2015. "Second language fluency: speaking style or proficiency? Correcting measures of second language fluency for first language behavior". *Applied Psycholinguistics* 36: 223–243. https://doi.org/10.1017/S0142716413000210.

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

Dörnyei, Zoltán, and Judit Kormos. 1998. "Problem-solving mechanisms in L2 communication: A psycholinguistic perspective". *Studies in Second Language Acquisition*, 20: 349–385. http://dx.doi.org/10.1017/S0272263198003039.

Ejzenberg, Roseli. 2000. "The juggling act of oral fluency: A psycho-sociolinguistic metaphor". In *Perspectives on Fluency*. Edited by Heidi Riggenbach, 287–313. Ann Arbor: University of Michigan.

Finnish National Agency for Education. 2019. Lukion opetussuunnitelman perusteet 2019 (LOPS) [National core curriculum for general upper secondary education 2019]. Retrieved from https://www.oph.fi/sites/default/files/documents/lukion_opetussuunnitelman_perusteet_2019.pdf

Foster, Pauline, and Peter Skehan. 1999. "The influence of source of planning and focus of planning on task-based performance". *Language Teaching Research*, 3(3): 215–247. https://doi.org/10.1177/136216889900300303.

Freed, Barbara. 1995. "What makes us think that students who study abroad become fluent?" In *Second Language Acquisition in a Study Abroad Context.* Edited by Barbara F. Freed. Amsterdam: John Benjamins Publishing Company.

Freed, Barbara. 2000. "Is fluency, like beauty, in the eyes (and ears) of the beholder?" In *Perspectives on Fluency*. Edited by Heidi Riggenbach, 243-265. Ann Arbor: University of Michigan Press.

Friedman, Debra A. 2011. "How to Collect and Analyze Qualitative Data." in *Research Methods in Second Language Acquisition*. Chichester, UK: John Wiley & Sons, Ltd, 180–200. Web.

Fulcher, Glenn. 2003. *Testing Second Language Speaking*. London; Routledge. https://doi.org/10.4324/9781315837376

GDPR. 2024. Accessed 30th of April 2024. Retrieved from https://gdpr.eu/what-is-gdpr/.

Götz, Sandra. 2013. *Fluency in Native and Nonnative English Speech*. Amsterdam: John Benjamins Publishing Company.

Hanzawa, Keiko. 2024. "Development of Second Language Speech Fluency in Foreign Language Classrooms: A Longitudinal Study." *Language Teaching Research* 28 (3): 816–838. https://doi.org/10.1177/13621688211008693.

Hilton, Heather. 2008. "The link between vocabulary knowledge and spoken L2 fluency". *Language Learning Journal* 36 (2): 153–166. https://doi.org/10.1080/09571730802389983.

Housen, Alex, Folkert Kuiken, and Ineke Vedder. 2012. "Complexity, accuracy and fluency". in *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*. Edited by Alex Housen, Folkert Kuiken and Ineke Vedder, 1–20. Amsterdam: John Benjamins Publishing Company.

Huensch Amanda, and Nicole Tracy-Ventura. 2017. "Understanding second language fluency behavior: The effects of individual differences in first language fluency, cross-linguistic differences, and proficiency over time". *Applied Psycholinguistics* 38 (4): 755– 785. https://doi.org/10.1017/S0142716416000424.

Kormos, Judit. 2006. *Speech Production and Second Language Acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Larson-Hall, Jenifer. 2016. *A Guide to Doing Statistics in Second Language Research Using SPSS and R*. London: Routledge.

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

Lennon, Paul. 2000. "The lexical element in spoken second language fluency". In *Perspectives on Fluency*. Edited by Heidi Riggenbach, 25–42. Ann Arbor: University of Michigan Press.

Lintunen, Pekka, Maarit Mutta, and Pauliina Peltonen. 2020. "Defining Fluency in L2 Learning and use" in *Fluency in L2 Learning and Use*, edited by Pekka Lintunen, Maarit Mutta, and Pauliina Peltonen. Bristol, UK: Multilingual Matters.

Lintunen, Pekka, and Peltonen, Pauliina. 2020. "What is Fluency? Learner Perceptions of the Concept" in *Fluency in L2 Learning and Use*, edited by Pekka Lintunen, Maarit Mutta, and Pauliina Peltonen. Bristol, UK: Multilingual Matters.

Mackey, Alison, and Susan M. Gass. 2011. *Research Methods in Second Language Acquisition: A Practical Guide*. West Sussex, England: John Wiley & Sons Ltd.

Maseda, Rebeca, and Dayna. J. DeFeo. 2014. "Finding Your "Spanish Voice" Through Popular Media: Improving Students' Confidence and Fluency". *The Journal of the Scholarship of Teaching and Learning.* 14 (3): 110–131. https://doi.org/10.14434/josotl.v14i3.5033.

Ministry of Education and Culture. 2017. Gaudeamus igitur - ylioppilastutkinnon kehittäminen. Opetus- ja kulttuuriministeriön julkaisuja 16. Helsinki. Retrieved from http://urn.fi/URN:ISBN:978- 952–263–462–7.

Moneypenny, Dianne Burke, and Rosalie Aldrich. 2024. "Study Abroad and Intensive Linguistic Immersion in Spain: Undergraduate Oral Proficiency for Short-Term Study Abroad". *Frontiers: The interdisciplinary Journal of Study Abroad* 36 (1): 640–659. https://doi.org/10.36366/frontiers.v36i1.781.

O'Brien, Mary. 2014. "L2 Learners' Assessments of Accentedness, Fluency, and Comprehensibility of Native and Nonnative German Speech". *Language Learning* 64 (4): 715–748. https://doi.org/10.1111/lang.12082.

Pallotti, Gabriele. 2021. "Measuring Complexity, Accuracy, and Fluency (CAF)." In *The Routledge Handbook of Second Language Acquisition and Language Testing*, edited by Paula Winke and Tineke Brunfaut, 201–210. New York: Routledge.

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

Peltonen Pauliina. 2018. "Exploring connections between first and second language fluency: A mixed methods approach". *The Modern Language Journal* 102: 676–692. https://doi.org/10.1111/modl.12516.

Peltonen, Pauliina. 2020. *Individual and interactional speech fluency in L2 English from a problem-solving perspective: A mixed-methods approach.* Doctoral Dissertation, University of Turku. Retrieved from https://urn.fi/URN:ISBN:978-951-29-8137-3

Peltonen, Pauliina, and Lintunen, Pekka. 2016. "Integrating quantitative and qualitative approaches in L2 fluency analysis: A study of Finnish speaking and Swedish-speaking learners of English at two school levels". *European Journal of Applied Linguistics* 4 (2): 209–238. https://doi.org/10.1515/eujal-2014-0018.

Peltonen, Pauliina, and Lintunen, Pekka. 2024. "Teaching L2 speech fluency with learner-corpus-based awareness-raising activities: Insights from a short-term intervention study". *International Journal of Learner Corpus Research,* 10(1): 39–66. https://doi.org/10.1075/ijlcr.00040.pel.

Robinson, Peter. 2001. "Task complexity, cognitive resources, and syllabus design: a triadic framework for examining task influences on SLA," In *Cognition and Second Language Instruction*. Edited by Peter Robinson, 287–318. Cambridge University Press.

Saleva, Maija. 1997. *Now They're Talking: Testing Oral Proficiency in a Language Laboratory*. Doctoral Dissertation, University of Jyväskylä. Print.

Segalowitz Norman. 2010. *The Cognitive Bases of Second Language Fluency*. New York, NY: Routledge.

Skehan Peter. 2009. "Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis". *Applied Linguistics* 30: 510–532. https://doi.org/10.1093/applin/amp047.

Song, Jayoung. 2020. "The effects of a short-term study abroad program on developing students' intercultural competence and oral proficiency". *Linguistic Research* 37: 1–29. https://doi.org/10.17250/khisli.37..202009.001

Tavakoli, Parvaneh, Colin Campbell, and Joan McCormack. 2016. "Development of Speech Fluency Over a Short Period of Time: Effects of Pedagogic Intervention." *TESOL Quarterly* 50, no. 2: 447–471. https://doi.org/10.1002/tesq.244.

Tavakoli Parvaneh, and Ann-Marie Hunter. 2018. "Is fluency being 'neglected' in the classroom? Teacher understanding of fluency and related classroom practices". *Language Teaching Research* 22(3): 330–349. https://doi.org/10.1177/1362168817708462.

Towell, Richard, Roger Hawkins, and Nives Bazergui. 1996. "The Development of Fluency in Advanced Learners of French". *Applied Linguistics* 17 (1): 84–119. https://doi.org/10.1093/applin/17.1.84.

van Gelderen, Amos. 1994. "Prediction of global ratings of fluency and delivery in narrative discourse by linguistic and phonetic measures—oral performances of students aged 11–12 years". *Language Testing* 11 (3): 291–319. https://doi.org/10.1177/026553229401100304.

Wolk, Christoph, Sandra Götz, and Katja Jäschke. 2020." Possibilities and drawbacks of using an online application for semi-automatic corpus analysis to investigate discourse markers and alternative fluency variables". *Corpus Pragmatics* 5: 7–36. http://dx.doi.org/10.1007/s41701-019-00072-x.

Appendices

Appendix 1 The structured narration elicitation task: story retelling

Kuvittele, että näet seuraavan tarinan sanomalehdessä ja haluat kertoa sen ulkomaalaiselle ystävällesi englanniksi. Sinulla on 3 minuuttia aikaa tutustua tekstiin ja n. 3 minuuttia tarinan kertomiseen.

Päivänsankari hyppäsi

Floridalainen nainen juhli 86-vuotispäiväänsä hyppäämällä laskuvarjolla pienkoneesta 2900 metrin matkan. "Kaikki pitivät minua hulluna, mutta minua ei pelottanut hiukkaakaan", Manya Joyce sanoi neitsythyppynsä jälkeen. "Se oli todella ihanaa." Joyce on taitava golfin ja tenniksenpelaaja, joka keräsi hypyllä rahaa veteraaniurheilijoiden olympialaisia varten. Hän on aina ollut kovanaamainen nainen: toimiessaan rikosreportterina Chicago Tribune -lehdessä 1920-luvulla hän pelasi korttia rikollispomo Al Caponen porukan kanssa. Joyce ei pitänyt hyppyä mitenkään vaarallisena. Hänellä oli mestarihyppääjä seuranaan seitsenminuuttisen hypyn ajan. Guinnessin ennätysten kirjan toimituksen mukaan Joyce on vanhin laskuvarjohyppääjä, jonka kirjan toimituskunta tietää.

Appendix 2 Example transcription

{er} (.) a woman flom (.) Florida (0.30) {um} (0.35) celebrated her (.) eighty(.)sixth (0.35) {um}{*h*_0.82} year day? (0.34) by jumping (0.78) with (0.52) a (1.42) s- [{um} {*pt*_0.46} *hah* (1.63)] laskuvarjo? (0.39) {um} in (0.38) sma:ll plane (1.18) for (0.25) two: (.) m (.) hundred and nine(.)ty me:t-t- (.) nine hundred meters' (.) way. {*hh*_0.95} Everybody (0.33) thought I was crazy (0.41) but (0.65) I: wasn't scared at all said (.) Ma:nya Joyce (0.40) a:fter he first (.) jump. {*h*_0.59} It was (.) really nice (0.27) he (0.41) said (.) also. {*h*_0.87} {um} Joyce (.) is a (0.44) ver- (0.25) talented (0.26) golf and (.) tennis playler (0.30) so (0.68) who {*h*_0.67} raised (1.03) {um} (0.36) money (0.57) from his (.) jump (.) to (0.42) veteranian (.) °athletics° {*h*_1.15} Olym:pics (1.02) or for (.) veteranian (1.20) °athletics Olympics.° {*h*_0.65} {um} she: is (0.85) h- she has always been a tough (.) woman. (0.88) {um} when he was: (.) working (0.45) as a {*h*_0.70} crime reporter in Chicago (.) Tribute (.) magazine in (0.51) °nineteen twenty°? {*h*_0.89} she played (0.72) a ca:rd (0.27) wit:h (.) the crime (0.40) boss Al Capone(.)s (0.94) with (.) Al Capone's (.) group. [{*h*_0.96} {umh} (1.16)] Joyce didn't (0.48) {um} (0.30) took (0.57) the (0.53) jump very (0.33) dangerous {*h*_0.88} he has a master jumper (0.43) with her (1.15) {um} (.) all the (.) seven minutes. [{*h*_1.60} {mm} (0.75)] °after Guinness (0.85) *indistinct*°

Appendix 3 Finnish summary

Finnish Summary

Johdanto ja teoria

Tässä pro gradu -tutkielmassa tutkitaan vieraan kielen sujuvuutta tarkastelemalla viivytysmekanismien (stalling mechanisms) kehittymistä lukion toisen vuoden opiskelijoiden valinnaisella puhekurssilla. Englannin kielen opetuksen ja arvioinnin näkökulmasta puheen sujuvuus on Suomessa ajankohtainen aihe, koska Ylioppilastutkintoon on jo suunnitteilla puhutun kielen testaamisesta (Opetus- ja kulttuuriministeriö 2017). Suullisen kielitaidon kehittymisen tutkimukset ovat usein pitkittäistutkimuksia, jotka tutkivat kielen kehittymistä pitkällä aikavälillä, mutta koska puhuttu kieli sisällytetään virallisiin arviointeihin lähitulevaisuudessa, on tärkeää tutkia, minkälaisia muutoksia puhutun kielen sujuvuudessa tapahtuu lyhyemmällä aikavälillä. Lyhyen aikavälin tutkimuksissa usein tutkitaan ulkomailla opiskelevien vieraan kielen kehittymistä, mutta tämä tutkimus keskittyy formaaliin luokkahuoneympäristöön. Suomalaisissa lukioissa tarjotaan tällä hetkellä vain yksi englannin puheeseen keskittyvä moduuli, joka on valinnainen opiskelijoille: ENA8 Viesti ja vaikuta puhuen (LOPS 2019). Tämä on 7 viikon kurssi, joka on suunniteltu kehittämään suullista kielitaitoa tarjoamalla opiskelijoille materiaaleja ja menetelmiä, mitkä auttavat suullisessa vuorovaikutuksessa. Kurssilla on mahdollista harjoitella sekä spontaania puhumista että valmistelua tarvitsevaa suullista tuottamista ja tehtävät ovat suunniteltu kommunikatiivisiksi (LOPS 2019).

Tämän tutkimuksen päätavoitteena on tutkia viivytysmekanismeja kuten taukoja, äänteiden pidentämistä, toistoja ja täytesanoja, ja niiden kehittymistä kurssin aikana. Tutkimus antaa tutkimustietoa siitä, miten luokkahuoneopetus vaikutta sujuvuuden kehittymiseen ja miten oppilaat kehittyvät suhteellisen lyhyessä ajassa. Nämä havainnot voivat auttaa opettajia tukemaan oppilaita paremmin sujuvuuteen liittyvien haasteiden voittamisessa ja parantamaan heidän puhutun englannin kielitaitoaan. Tässä tutkimuksessa tutkitaan sekä määrällisin että laadullisin menetelmin eri viivytysmekanismien määriä, kestoja ja sijainteja ja tutkimuskysymyksiä on kaksi: ensin tutkin millaisia ryhmätason muutoksia ja vaihtelua on viivytysmekanismien määrissä kurssin aikana ja seuraavaksi tutkin yksilötasolla muutoksia ja vaihtelua viivytysmekanismien käytössä.

Sujuvuustutkimuksissa sujuvuus määritellään yleensä luonnolliseksi, nopeaksi ja sujuvaksi kielenkäytöksi, jossa ei ilmene paljon epäröintiä eikä taukoja (Tavakoli & Hunter 2018). Tässä tutkimuksessa käytettyjä viitekehyksiä ovat 1990-luvulta saakka käytetty CAF-viitekehys, mikä

erottelee sujuvuuden, tarkkuuden ja kompleksisuuden (fluency, accuracy, complexity) kielitaidon eri osa-alueiksi (Pallotti 2021), Lennonin laaja ja kapea sujuvuuden käsitys (broad vs. narrow senses, Lennon 1990) missä laaja käsitys kattaa kokonaisuudessaan kielitaidon tason ja kapea käsitys tarkoittaa yksittäisiä kielitaidon osa-alueita, ja lisäksi kognitiivisen sujuvuuden, puhetuotoksen sujuvuuden ja havaitun sujuvuuden viitekehys (three aspects of fluency, Segalowitz 2010), joista keskityn tarkemmin puhetuotoksen sujuvuuteen. Viimeinen tärkeä viitekehys on Dörnyein ja Kormosin (1998) psykolingvistinen viitekehys, jossa epäsujuvuudet nähdään ongelmanratkaisumenetelminä, mitkä auttavat vieraan kielen prosessointiin.

Aiemmissa tutkimuksissa natiivipuhujan puhetta on käytetty usein vertailun kohteena, mutta tutkimukset ovat osoittaneet, että jopa edistyneet vieraan kielen opiskelijat erotetaan helposti natiivipuhujista (Abrahamsson ja Hyltestam 2009; O'Brien 2014). Tämän lisäksi tutkimukset, joissa on käytetty natiivipuhujia kontrolliryhmänä, ovat havainneet variaatiota myös äidinkielen sujuvuudessa (Götz 2013; Peltonen ja Lintunen 2016). Natiivipuhujien puheessa on myös epäröintiä ja taukoja, mutta on silti tärkeää tutkia minkälaiset asiat erottavat natiivipuhujat jopa edistyneistä kielenoppijoista. Tähän on käytetty apuna erilaisia mittareita, joiden avulla puheesta mitataan sujuvuuden eri komponentteja. Sujuvuutta on tutkittu paljon määrällisin menetelmin. Lennonin kapeaa määritelmää on käytetty sujuvuustutkimuksissa paljon ja Tavakoli ja Hunter (2018) ovat edelleen jakaneet laajan ja kapean määritelmän neljään eri kategoriaan: hyvin laaja, laaja, kapea ja hyvin kapea. Näistä hyvin kapea määritelmä kattaa temporaalisen sujuvuuden eri komponentteja kuten nopeutta, taukoja ja korjauksia (Skehan 2009), joita sujuvuudesta voidaan määrällisesti mitata ja analysoida. Temporaalista sujuvuutta voidaan mitata erilaisten mittareiden avulla. Näitä mittareita ovat esimerkiksi puhe- ja artikulaationopeus, hiljaiset ja täytetyt tauot mistä mitataan määrää, sijaintia ja pituutta, sekä korjaukset eli esimerkiksi toistot ja uudelleenmuotoilut. Tässä tutkimuksessa ei käytetä kaikkia näitä mittareita vaan keskitytään taukoihin ja korjauksiin, joita ei kohdella epäsujuvuuden merkkeinä vaan ongelmanratkaisukeinoina, joita kielen opiskelijat käyttävät vähentääkseen hiljaisuuksia ja ylläpitääkseen puhumista. Tällöin niistä voidaan puhua viivytysmekanismeina (Dörnyei ja Kormos 1998). Viivytysmekanismeja ovat tauot, äänteiden venyttäminen, täytesanat ja toistot. Kuten mainitsin aikaisemmin, myös natiivipuhujat käyttävät viivytysmekanismeja puhuessaan, mutta vieraan kielen käyttö ei ole niin automatisoitunutta kuin natiivipuhujilla, jolloin viivytysmekanismien käyttö voi olla oppijoilla huomattavampaa koska puheen suunnittelu kestää kauemmin.

Aiemmissa lyhyen aikavälin tutkimuksissa on nähty kehitystä puheen nopeudessa (Peltonen ja Lintunen 20246; Tavakoli et al 2016), hiljaisten taukojen pituuksissa (Hanzawa 2024) ja artikulointinopeudessa (Tavakoli et al 2016). Kehitystä on myös havaittu taukojen ja korjausten

määrissä (ibid.). Tutkimuksissa on myös havaittu, että taukojen ja korjausten määrät ovat suoraan verrannolliset opiskelun määrään (Hanzawa 2024) ja että viivytysmekanismien käytössä on suurta yksilöllistä variaatiota (Peltonen ja Lintunen 2024). Nämä tutkimukset osoittavat, että sujuvuudessa voi tapahtua kehitystä lyhyessä ajassa, mutta ohjeistuksen tavalla ja aktiviteeteilla voi olla myös vaikutusta. Tutkimuksessa, jossa tutkittiin kuinka oppimisympäristö ja -materiaalit vaikuttavat opiskelijoiden itsevarmuuteen ja sitä kautta sujuvuuteen (Maseda ja DeFeo 2014) havaittiin, että opiskelijat kehittivät lukukauden kestäneen opiskelun aikana motivaatiotaan, itsevarmuuttaan ja tehokkuuttaan, mitkä vaikuttivat puhumisen sujuvuuteen. Opiskelijat raportoivat, että opintojen alussa he olivat hermostuneita, mutta rohkaistuivat opintojen aikana ja luottivat taitoihinsa enemmän, jolloin epäröinti väheni (Maseda & DeFeo 2014, 120–121). Itsevarmuus yhdistetään myös kognitiiviseen sujuvuuteen, ja tauot ja korjaukset ovat yhteydessä siihen missä määrin puhuja on varma omasta osaamisestaan, kun tapahtuu epäröintiä (Housen, Kuiken & Vedder 2012). Tämänlaisen lyhyen kurssin (kuten ENA8) tai opintojakson tarkoitus voikin olla opiskelijoiden itsevarmuuden lisäämisessä, mikä puolestaan lisää puheen sujuvuutta.

Aineisto ja tutkimusmenetelmät

Tutkimuksen osallistujat olivat suomalaisen lukion opiskelijoita, jotka olivat ilmoittautuneet englannin puhekurssille (ENA8). Äänitteet kerättiin 40 opiskelijalta, mutta koska osalla oli puhehäiriöitä ja osa perui suostumuksensa tutkimukseen, osallistujien määrä laskettiin 33:een. Tutkimuksen aineisto koostui 66 ääninäytteestä (1–3 minuuttia per äänite), joista 33 äänitettiin ennen 7 viikon puhekurssia ja 33 sen jälkeen. Kaikilla opiskelijoilla oli äidinkielenä suomi ja he olivat 17–18-vuotiaita äänittämisen aikana. Kaikki olivat myös suorittaneet 7 tai 8 englannin kurssia ennen puhekurssia. Viivytysmekanismien käyttöä mitattiin sekä määrällisin että laadullisin metodein ääninäytteistä, kuten sujuvuustutkimuksissa suositellaan (Peltonen ja Lintunen 2016). Äänitteiden keräämisessä opiskelijoita pyydettiin kertomaan suomenkielinen uutinen englanniksi (Liite 1). Heillä oli kolme minuuttia aikaa lukea teksti ja kolme minuuttia aikaa kertoa tarina englanniksi. Kyseinen teksti oli otettu Maija Salevan väitöskirjasta (1997), ja teksti oli sama molemmilla äänityskerroilla.

Äänitteet litteroitiin kirjalliseen muotoon ja apuna käytettiin Microsoftin Transcribe-ohjelmaa. Taukojen mittauksessa käytettiin Praat -puheanalyysiohjelmaa (Boersma ja Weenink, 1992) ja tauot, jotka olivat yli 0.25 sekuntia pitkiä, laskettiin mukaan analyysiin. Viivytyskeinojen määrät, kestot ja sijainnit lisättiin litteroituun tekstiin manuaalisesti ja tarkastettiin oikeellisuuden vahvistamiseksi. Näin näytteistä saatiin arvot mitä analysoitiin sekä määrällisesti että laadullisesti. Ryhmäkohtaisessa määrällisessä analyysissa käytettiin seuraavia mittareita:

- Hiljaisten taukojen määrä (frequency) minuutin aikana

- Täytettyjen taukojen määrä minuutin aikana (kuten suomen 'öö' tai englannin 'uh' ja 'um')

- Venytettyjen äänteiden määrä minuutin aikana

- Englanninkielisten täytesanojen määrä minuutin aikana (kuten 'like', 'I mean', 'well')

- Suomenkielisten täytesanojen määrä minuutin aikana (kuten 'eiku', 'niinku')

- Toistojen määrä minuutin aikana

Mittarit suhteutetaan jokaisen näytteen kestoon seuraavalla laskukaavalla: viivytysmekanismin kokonaismäärä jaetaan puhenäytteen kokonaiskestolla ja kerrotaan 60:llä niin saadaan lopullinen määrä minuutissa. Kvantitatiivisessa analyysissa aineistosta tutkittiin kurssin aikana tapahtuneita tilastollisesti merkittäviä muutoksia SPSS 29 käyttäen. Jokaisesta viivytysmekanismista tutkittiin ensin normaalijakauma, minkä jälkeen tehtiin joko parametrinen tai ei-parametrinen t-testi, jolla tutkittiin tilastollista merkittävyyttä.

Kvantitatiivisen analyysin perusteella valittiin kolme opiskelijaa kvalitatiiviseen analyysiin. Kolme osanottajaa valittiin, koska 66:en ääninäytteen kvalitatiivinen analysoiminen ei olisi ollut mahdollista tämän suuruisessa tutkimuksessa, minkä vuoksi kaksi opiskelijaa valittiin edustamaan ryhmän ääripäitä (extreme case sampling, Dörnyei 2007) ja viimeinen oli opiskelija, joka numeerisesti kehittyi eniten kurssin aikana (37.4 %). Ääripäiden analysoiminen tässä tutkimuksessa mahdollisti ryhmänsisäisen vaihtelun sekä yksilöllisten erojen tutkimisen. Kvalitatiivisessa analyysissä analysoitiin samoja viivytysmekanismeja kuin kvantitatiivisessakin, mutta sen sijaan että keskityttiin vain määriin, keskityttiin myös pituuksiin ja sijainteihin.

Tulokset ja päätelmät

Ääninäytteiden kvantitatiivisessa analyysissa keskityttiin ryhmäkohtaisiin muutoksiin kurssin aikana ja tulokset olivat vaihtelevia. Viittä viivytysmekanismia analysoitiin muutosten kannalta, mutta täytetyt hiljaisuudet ja täytesanat eroteltiin molemmat vielä kahteen kategoriaan koska osa opiskelijoista käytti suomea äänityksissään. Opiskelijat puhuivat paljon vähemmän kurssin jälkeen. Ryhmän keskiarvo ennen kurssia oli 99.64 sekuntia ja kurssin jälkeen keskiarvo oli laskenut 78.91 sekuntiin. Tähän on haastavaa löytää varmaa syytä, mutta tuloksissa löytyi teemoja mitkä voivat mahdollisesti selittää tätä. Hiljaisten taukojen käyttö kurssin aikana laski ja tulos oli tilastollisesti merkitsevä. Keskimäärin opiskelijat olivat hiljaa vähemmän mikä voi osin selittää lyhempää puheen kokonaispituutta. Täytettyjen taukojen (suomen- ja englanninkielisten äänteiden) käytön tulokset eivät olleet tilastollisesti merkitseviä, mutta suomenkielisten äänteiden kohdalla tilastollinen tulos

tulee tulkita varauksella. Vaikka englanninkielisten täytettyjen taukojen käyttö pysyi melkein samana kurssin ajan, suomen vaikutus (eli äänne /ø/) laski 50 % kurssin aikana. Tätä voi selittää se, että otos oli liian pieni tai erot eivät olleet tasaisesti jakautuneet mediaanin ympärille ja vaikuttivat tulokseen, koska vain osa opiskelijoista käytti /ø/ äännettä. Tätä osoittaa myös suuri keskihajonta.

Täytesanoja käytettiin viivytysmekanismeista vähiten ja niiden käyttö pysyi samana kurssin ajan, paitsi suomenkieliset täytesanat lisääntyivät 30 %. Myös toistojen käyttäminen pysyi samana kurssin ajan. Äänteiden venyttämisen käyttö kurssin aikana oli myös tilastollisesti merkittävää ja tarkoittaa sitä, että opiskelijat käyttivät sitä vähemmän kurssin jälkeen. Tätä voi selittää se, että tehtävä oli uusi ensimmäisellä äänityskerralla, mikä on saattanut nostaa äänteiden venyttämisen määrää aluksi, mutta tehtävän ollessa jo tuttu toisella äänityskerralla, sai käytön vähenemään. Täytetyissä hiljaisuuksissa ja äänteiden venyttämisessä oli ryhmänsisäisesti myös suurta vaihtelua. Täytettyjen hiljaisuuksien keskiarvo ennen kurssia oli 7.82 (keskihajonta = 7.69) ja kurssin jälkeen 6.03 (keskihajonta = 4.99). Äänteiden venyttämisessä keskiarvo oli kurssin alussa 8.82 (keskihajonta = 4.74) ja lopussa 5.88 (keskihajonta = 4.70). Keskihajonnat ovat suuret mikä viittaa suureen vaihteluun näiden viivytysmekanismien käytössä.

Kvalitatiivisessa tutkimuksessa tutkittiin kolmea opiskelijaa tarkemmin. Opiskelija 1 puhui vain 6.19 sekuntia kurssin alussa, kun taas kurssin lopussa hän puhui 23.71 sekuntia. Kvantitatiivisen, numeroihin perustuvan analyysin perusteella näyttäisi siltä, että opiskelija 1 huononsi suoritustaan kurssin aikana, koska hän käytti enemmän viivytysmekanismeja, mutta totuus on aivan päinvastoin: koska hän puhui enemmän, hän myös käytti viivytysmekanismeja ongelmanratkaisuun ja jatkaakseen puhumista. Tämä tarkoittaa sitä, että hän on kehittynyt kurssin aikana paljon. Opiskelija 2 oli ryhmän sujuvimpia puhujia, joka antoi molemmilla kerroilla hyvin sujuvan suorituksen. Näiden kahden tapauksen pohjalta on selvää, että ryhmän sisällä on suurta vaihtelua viivytysmekanismien käytössä sekä sujuvuustasoissa. Opiskelija 3 kehittyi eniten kurssin aikana, 37.4 %. Sisällöllisesti hänen äänitteensä pysyivät täysin samana, mutta kehitys tapahtui hiljaisten taukojen määrässä ja sijainnissa. Ennen kurssia hän käytti viivytysmekanismeja sanaetsinnöissä ja välttääkseen hiljaisuuksia, ja kurssin jälkeen hän käytti niitä edelleen, mutta ei niin usein.

Tulokset osoittavat, että viivytysmekanismien käytössä on merkittäviä eroja ja mieltymyksiä. Esimerkiksi vain osa käytti täytesanoja ja usein sama opiskelija käytti niitä ennen moduulia ja sen jälkeen. Kvalitatiivisesta tutkimuksesta käy ilmi, että viivytysmekanismeilla on tärkeä rooli sujuvuuden lisäämisessä ja ne auttavat puheen jatkuvuudessa. Taukojen määrä laski, mutta myös sijainnissa tapahtui muutosta ja hiljaisuuksia ei esiintynyt lauseiden keskellä niin usein kurssin jälkeen. Tulokset näyttävät, että joitakin ryhmäkohtaisia muutoksia voidaan havaita

viivytysmekanismien käytössä, mutta kvalitatiivinen analyysi osoittaa, että opiskelijat eivät edenneet samassa tahdissa tai samalla tavalla. Jotkut opiskelijat, kuten opiskelija 3, osoittivat huomattavaa kehitystä sujuvuudessaan, kun taas toisilla – vaikka etenivätkin taidoissaan – oli edelleen haasteita kurssin jälkeen.

Lopuksi

Tässä tutkimuksessa selvitettiin englannin kielen puheen sujuvuuden ja viivytysmekanismien kehittymistä 7 viikkoa kestävän puhekurssin aikana. Tutkimusaineistoa analysoitiin niin määrällisin kuin laadullisinkin menetelmin ja tutkimuskysymyksiä oli kaksi: 1. Minkälaisia ryhmätason muutoksia ja vaihtelua on viivytysmekanismien käytössä 7 viikon puhekurssin aikana? ja 2. Minkälaisia yksilötason muutoksia ja vaihtelua on viivytysmekanismien käytössä 7 viikon puhekurssin aikana? Aineisto koostui 66 puhenäytteestä, mitkä otettiin ennen kurssia ja sen jälkeen. Tutkimuksessa yhdistettiin kvantitatiivinen analyysi mikä keskittyi viivytysmekanismien määriin ja kvalitatiivinen analyysi mihin valittiin kolme opiskelijaa ryhmästä.

Ryhmäkohtaisesti tulokset paljastavat, että hiljaisten taukojen käytössä ja äänteiden venyttämisessä oli kehitystä, eikä niitä käytetty niin paljon kurssin jälkeen. Suomen vaikutus äänteen /ø/ muodossa väheni huomattavasti myös, mutta suomenkieliset täytesanat lisääntyivät 30 %. Englanninkieliset täytesanat ja toistot pysyivät samana kurssin aikana. Tärkeät muutokset hiljaisiin taukoihin ja äänteiden venyttämiseen voivat selittää miksi oppilaat puhuivat keskimäärin vähemmän kurssin jälkeen, mutta syyt voivat olla monimuotoisia. Esimerkiksi jo tuttu materiaali ja itsevarmuuden nousu voivat vaikuttaa siihen. Laadullinen tutkimus täydensi määrällistä tutkimusta tutkimalla vähiten sujuvaa opiskelijaa ja yhtä sujuvimmista opiskelijoista, jolla oli kaikkein johdonmukaisin suoritus. Kolmas tapaus oli opiskelija, jolla oli eniten kehitystä kurssin aikana. Ääritapausten tutkiminen mahdollisti ryhmän koko sujuvuusalueen tarkkailun ja antoi myös selkeämmän näkemyksen yksilöllisistä eroista. Tapaus 3 edusti sitä, kuinka paljon parannusta on saavutettavissa lyhyessä ajassa.

Tällä tutkimuksella on rajoitteita, joista merkittävimmät ovat pieni otanta ja ulkoiset muuttujat, jotka voivat vaikuttaa yksilölliseen vaihteluun ja kehittymiseen. Pieni otos tarkoittaa, että havainnot eivät voi edustaa koko Suomen lukiolaisia, ja tämä voisi rajoittaa tulosten yleistettävyyttä muihin opiskelijaryhmiin tai oppimiskonteksteihin. Tulokset antavat kuitenkin arvokasta tietoa suomalaisten lukiolaisten ongelmanratkaisumekanismien käytöstä ja siitä, kuinka paljon vaihtelua luokkahuoneen sisällä voi olla. Tulokset voivat kertoa opettajille tärkeää tietoa lukiolaisten sujuvuudesta. Myös ulkoiset muuttujat saattavat vaikuttaa siihen, kuinka paljon opiskelijat kehittyvät tällaisen lyhyen kurssin aikana. Tutkimuksessa ei voida ottaa huomioon oppilaiden

altistumista englannin kielelle luokkahuoneen ulkopuolella, heidän oppimismotivaatiotansa tai englannin puhumiseen liittyvää ahdistusta. Nämä tekijät voivat vaikuttaa tuloksiin, mutta niitä on vaikea mitata tarkasti tai hallita. Tutkimusta voisi kehittää tutkimalla sujuvuutta monipuolisemmin. Esimerkiksi lisäämällä kyselyt ennen ja jälkeen kurssin, joissa kysytään oppilaiden mielipiteitä heidän oppimisestaan ja ottamalla opettajan arvioinnin sujuvuudesta mukaan tutkimukseen tekisi tutkimuksesta monipuolisemman. Tutkimuksen voisi myös toistaa yliopisto-opiskelijoille, kun he ovat saaneet enemmän puheopetusta, ja verrata että onko viivytysmekanismien käytössä eroja. Yhteenvetona voidaan todeta, että vaikka 7 viikkoinen englannin puhekurssi pystyi ryhmätasolla parantamaan sujuvuutta jossakin määrin, laadullinen analyysi näytti, kuinka merkittävää yksilöllinen vaihtelu ryhmän sisällä todella on, ja kuinka monipuolinen sujuvuuden kirjo lukioryhmän sisällä voi olla.