

Intermediate and Advanced ESL Speakers' Pause and Repair Use in the Finnish
National Certificate of Language Proficiency Speaking Test: A Mixed Methods Study

-

Tytti Kim

MA Thesis

English, Language Specialist Degree Programme

School of Languages and Translation Studies

University of Turku

March 2021

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin Originality Check service.

UNIVERSITY OF TURKU

School of Languages and Translation Studies / Faculty of Humanities

KIM, TYTTI: Intermediate and Advanced ESL Speakers' Pause and Repair Use
in the Finnish National Certificate of Language Proficiency
Speaking Test: A mixed methods study

MA Thesis, 72 pages, 11 pages of appendices

English, Language Specialist Degree Programme

March 2021

This thesis examined intermediate and advanced ESL speakers' pause and repair use in the Finnish National Certificates of Language Proficiency speaking test, combining quantitative and qualitative research methods in an attempt to answer the following research questions: "Which quantitative measures of fluency distinguish between intermediate and advanced levels of proficiency?"; "how do pause and repair use distinguish between intermediate and advanced levels of proficiency?"; and "are pause and repair use interdependent?" 30 intermediate and 30 advanced Finnish L2 speakers taking a proficiency test in English were sampled from the Finnish National Certificates of Language Proficiency corpus (YKI corpus) and controlled for age, gender, and level of education. Their speech samples were transcribed, annotated, and analysed statistically. The quantitative results attained by analysing extreme cases showed that the temporal fluency measures of articulation rate, mean length of run, pause duration, frequency, and location as well as repair location distinguished between the intermediate and advanced levels of YKI, and that the largest difference between the two groups was in the mean lengths of run, which was used for extreme case sampling. These extreme samples from 12 participants were then analysed quantitatively for pause and repair use based on Nakatsuhara, Tavakoli & Awwad's typology (2019): pauses and repairs were divided into pauses and repairs related to access and retrieval difficulty, reformulations, and effective speech delivery. Qualitative results showed that while intermediate participants had slightly more pauses and repairs related to access and retrieval difficulty than advanced participants, advanced participants were successful in their lexical and structural search more often than intermediate participants. In contrast, intermediate participants had more pauses related to reformulations than advanced. As for repairs, both intermediate and advanced participants had comparable numbers of reformulation repairs, but reformulation repairs were more common in the speech of intermediate and advanced participants with low mean lengths of run. In addition, the results showed that advanced participants used more of their pause and repair opportunities for more effective delivery. Advanced participants also used pauses and repairs to navigate socially and culturally difficult topics. Finally, the results showed that pause and repair use are interconnected: pause and repair use co-occurred and were used to achieve similar effects. The results suggest that pause and repair use, based on which inferences of L2 speakers' cognitive fluency can be made, should not be overlooked in fluency studies or proficiency testing. It is suggested that pause and repair use should be studied further and included into definitions and operationalisations of L2 fluency.

Keywords: second language fluency, Finnish national certificates of language proficiency, YKI, L2 proficiency, language testing, L2 English, English as a Second Language

Table of Contents

1 Introduction	1
2 Fluency as an indicator of oral L2 proficiency.....	6
2.1 Approaches to fluency	6
2.2 Operationalizing fluency in the context of language testing.....	9
2.2.1 Temporal fluency.....	12
2.2.2 Beyond temporal fluency	14
2.2.3 Emerging research	16
2.3 Fluency and the Finnish National Certificate of Language Proficiency.....	18
3 The present study	24
3.1 Study design	24
3.2 Data and methodology	27
3.2.1 Participants.....	27
3.2.2 Material and transcriptions.....	28
3.2.3 Quantitative data and extreme case sampling	30
3.2.4 Qualitative data.....	32
4 Results	34
4.1 Quantitative fluency measures distinguishing between intermediate and advanced levels of YKI, and extreme case sampling	34
4.2 Pause and repair use on intermediate and advanced levels of YKI	36
4.2.1 Pauses.....	37
4.2.1.1 Pauses related to access and retrieval difficulty	38
4.2.1.2 Pauses related to reformulation and rescue	43
4.2.1.3 Pauses related to effective delivery.....	45
4.2.2 Repairs	48
4.2.2.1 Repairs related to access and retrieval difficulty	49
4.2.2.2 Repairs related to reformulations.....	50
4.2.2.3 Repairs related to effective delivery.....	51
4.3 The interdependence of pause and repair use	53
5 Discussion	55

5.1 Quantitative fluency measures distinguishing between intermediate and advanced levels of YKI.....	55
5.2 Pause and repair use on intermediate and advanced levels of YKI	57
5.3 The interdependence of pause and repair use	61
5.4 Implications for the YKI speaking scale.....	62
5.5 Evaluating the present study	63
5.6 Suggestions for further research.....	65
6 Conclusion.....	67
List of References	70
Appendices	73

List of Figures

Figure 1. Study design.....	26
Figure 2. Participants' pause use by MLR and pause category (%).....	37
Figure 3. Participants' pause use by MLR and pause subcategory (%).....	38
Figure 4. Participants' repair use by MLR and repair use category.....	48
Figure 5. Participants' repair use by MLR and repair use subcategory (%).....	48

List of Tables

Table 1. Quantitative fluency features selected for the present study.....	30
Table 2. Statistical analysis: summary of quantitative results.....	35
Table 3. Participants' pauses.....	Appendix 2
Table 4. Participants' repairs.....	Appendix 3

Abbreviations

ADV	Advanced (level)
CEFR	Common European Framework of Reference
EC	End-clause, clause border (position)
EFL	English as a foreign language
ESL	English as a second language
INT	Intermediate (level)
L1	First language (native language)
L2	Second language (and any foreign language acquired after acquiring L1)
MC	Mid-clause (position)
MLR	Mean length of run
SLA	Second language acquisition
YKI	Finnish National Certificates of Language Proficiency

1 Introduction

Several second language proficiency tests test spoken language fluency as an aspect of spoken language proficiency and communicative language ability, and the CAF framework (Complexity, accuracy, and fluency; Skehan 1998) has been widely accepted as part of second language proficiency scales such as the Common European Framework of Reference for language proficiency, a criterion-based reference scale describing the achievement of language learners (Council of Europe 2011 [2001]). Research suggests that several fluency measures traditionally categorised under the dimensions of speed, breakdown, and repair fluency (Skehan 1998, 2000, Segalowitz 2010) are reliable indicators of spoken language fluency and language proficiency. Of these three dimensions, speed fluency measures such as speech rate and articulation rate are often cited as most accurately predicting oral proficiency (e.g. De Jong et al. 2013 and Kahng 2014). Many studies also include mean length of run (Kahng 2014) as well as the number of pauses, repetitions, and repairs (De Jong 2013) as good predictors of oral proficiency. De Jong summarises a body of research by stating that speed measures and pause measures, with the exception of total duration of unfilled pauses, have been found to be related to proficiency, as they variation in total pause duration has been thought to result from individual speaking styles (De Jong 2016, 206). Still, further research into fluency measures is required to determine which fluency measures best distinguish between different levels of proficiency: despite the widespread use of fluency in language proficiency scales, attaining a concise, analytical, and comprehensive description of this abstract concept on different levels of proficiency can be difficult: there are individual differences in both L1 and L2 speech with regard to speed and pause phenomena, as well as dysfluencies (e.g. De Jong 2016, 206). The traditional operationalisation of fluency into speed, breakdown, and repair fluency measures is also not without problems: fluency measures overlap and are, in some cases, interdependent. For example, Nakatsuhara, Tavakoli & Awwad suggest that intermediate and advanced L2 speakers' pause and repair use are interdependent (Nakatsuhara, Tavakoli & Awwad 2019, 37).

Pauses and repairs have been a topic of interest in fluency studies and contrastive linguistics from the 1970s and have been treated as simply subtracting from fluency, but there is also mounting evidence that the role of pauses and repairs in oral fluency is far from one-dimensional. One of the earliest descriptions of Finnish L2 English-speakers'

fluency is from Lehtonen, who describes Finnish L2 speakers' pauses "more or less sporadic as regards the syntactic structure of the English sentences" (Lehtonen 1979, 37, 48). Since then, there have been numerous attempts to quantify and describe the role of pauses (and later, repairs) in fluency, and many proficiency tests include a treatment of pauses and repairs in their fluency criterion. Fluency studies have gained popularity in recent years, new studies adding to the growing corpus of fluency studies, and each adding to the effort of forming a clear, analytical, and comprehensive description of fluency.

However, researchers have called for "moving beyond conceptualizing fluency solely as a temporal phenomenon" (Wright and Tavakoli 2016, Peltonen 2017). While fluency studies have branched out into studying speech rhythm, prosody, tone units and even gestures, many fluency studies still contrast L2 speakers against native speakers for temporal fluency measures. In contrast, interesting new views are being offered by the University of Turku, where fluency studies are a robust line of research: for example, Lintunen, Mutta & Peltonen have added new perspectives from psycholinguistics, sign language studies and L2 assessment into perspectives of fluency and SLA (Lintunen, Mutta & Peltonen 2019).

Recent studies are also broadening the scope of fluency studies and reconceptualising fluency by, for example, linking temporal fluency with problem-solving strategies in interaction (Peltonen 2017) and perceived fluency with speech rhythm (Salomaa 2019). Furthermore, the problem with temporal fluency measures usually consisting of speed measures, silent pauses and sometimes repairs, is that these measures are one-dimensional and seen as either adding to or subtracting from fluency. In contrast, Peltonen (2017) places repairs outside temporal fluency in interaction, placing them under *stalling mechanisms*, and finds that together with communication strategies, stalling mechanisms may in fact contribute to temporal fluency by compensating for local dysfluencies. As such, repairs are not a one-dimensional measure that simply subtracts from fluency. Furthermore, perceived speech rhythm and perceived fluency are closely related (e.g. Tominaga 2011, 53, Salomaa 2019, 49), and speakers' use of pause and repair opportunities is integral to the rhythm of speech.

However, if we simply keep score of L2 speakers' pauses and repairs, we tend to forget that speakers use pause and repair opportunities for many different purposes. As such, neither breakdown nor repair fluency are one-dimensional features of fluency: quantitative measures alone fail to comprehensively characterise pause and repair use on

different proficiency levels. For example, in a recent study, Nakatsuhara, Tavakoli & Awwad identify three categories of pause and repair use: access and retrieval, reformulations, and effective speech delivery (Nakatsuhara, Tavakoli & Awwad 2019, 26), positing that the use of pause and repair opportunities distinguishes between higher intermediate (B2) and lower advanced (C1) L2 speakers (Nakatsuhara, Tavakoli & Awwad 2019, 37-38). As such, combining a qualitative analysis of pause and repair use with a quantitative analysis of temporal fluency measures leads to a better conceptualisation of fluency, and a better conceptualisation of fluency in turn leads to better operationalisations of fluency in proficiency criteria. To summarise, it becomes apparent that further research into combining temporal measures with a qualitative analysis of L2 speakers' use of pause and repair opportunities is required, and that this effort requires a mixed-methods approach.

Following the work of Nakatsuhara, Tavakoli & Awwad (2019), the present thesis aims to contribute to the growing corpus of fluency studies by studying speech samples from 60 Finnish-speaking candidates taking the Finnish National Certificate of Language Proficiency in English, and combining a quantitative analysis of temporal fluency phenomena with a qualitative analysis of intermediate and advanced ESL speakers' pause and repair use. The objectives of this study are, first, to establish which quantitative measures best distinguish between intermediate and advanced levels of proficiency in the Finnish National Certificates of Language Proficiency, second, to study how pause and repair use differentiate between the two levels of proficiency, and finally, to study whether pause and repair are interconnected.

Following this, the present thesis seeks to answer the following research questions:

- 1 Which quantitative measures of fluency distinguish between intermediate and advanced levels of proficiency of the Finnish National Certificates of Language Proficiency?
- 2 How do pause and repair use distinguish between intermediate and advanced levels of proficiency of the Finnish National Certificates of Language Proficiency?
- 3 Are pause and repair use interdependent?

Regarding research question 1, I hypothesise that based on e.g. De Jong et al. 2013 and Kahng 2014, i) the advanced group has on average higher articulation rates than the intermediate group, ii) the advanced group has on average higher mean lengths of run than the intermediate group, iii) the advanced group has on average lower pause frequencies and total pause durations than the intermediate group, and that they have more on average pauses on the clause border than in mid-clause positions, and iv) based on Peltonen (2017, 10), advanced-level participants repair their speech more than intermediate-level participants.

Furthermore, regarding research question 2: in line with results from Nakatsuhara, Tavakoli & Awwad (2019, 26-27), I expect to find that based on the qualitative analysis, advanced and intermediate participants use roughly similar numbers of pause and repair opportunities for access and retrieval difficulty (i.e. for pauses and repairs related to lexical and structural search and monitoring language), but advanced participants are expected to be successful in their lexical and structural search more often than intermediate participants. Furthermore, I expect the advanced group to utilize fewer pause and repair opportunities for pauses and repairs related to reformulations (i.e. pauses and repairs related to reformulating ideas and utterances and rescuing ungrammatical utterances), and more pause and repair opportunities to improve the effectiveness of their speech (i.e. adding information, examples and justifications, offering opinions and comments, shifting topics and signalling dispreference for the topic). Conversely, I expect the intermediate group to use more pause and repair opportunities for pauses and repairs related to access and retrieval difficulty and reformulations, and fewer pause and repair opportunities to improve the effectiveness of their speech. As such, I expect the two groups to compensate for different areas of their speech, namely that the intermediate group is more concerned with self-monitoring and correcting, and the advanced group with the effectiveness of their speech and their overall communicative prowess.

As for Research question 3, I expect intermediate and advanced L2 speakers' pause and repair use to be interconnected as per Nakatsuhara, Tavakoli & Awwad (2019, 37). First, I expect pauses and repairs to co-occur. Second, while excessive use of repairs is usually considered a sign of disfluency, I also expect advanced participants to use repairs to maintain speech and to avoid unfilled pauses (e.g. Peltonen 2017, 10). Finally, I expect pauses and repairs to be used for similar purposes in speech.

The present study is first and foremost indebted to previous studies in the fields of second language acquisition and language testing: in specific, it owes much to CAF

and fluency studies. This theoretical background, including influences, central terms, frameworks, methods, and studies, is examined in section 2, which provides a theoretical background to the present study by introducing, contextualising, and localising terminology, theories, models, and studies central to the present thesis before moving on to examining fluency in the context of language testing in the Finnish context. The present study, including the choice of research methodology, data and analyses, is presented in section 3, whereas section 4 presents the quantitative and qualitative results of the present study. These results are then discussed in section 5. Finally, conclusions are presented in section 6.

2 Fluency as an indicator of L2 oral proficiency

This section examines fluency as an indicator of L2 oral proficiency by defining and conceptualising oral L2 fluency in SLA and its role in the CAF framework, before finally discussing the future of CAF. However, before examining fluency as an indicator of proficiency, this section explains some central terminology choices made for the sake of clarity. In fact, any examination or explanation of an aspect of oral L2 proficiency would be impossible without first carefully defining i) *second language* and ii) *proficiency* in the context of second language acquisition.

i) Second language. In Second Language Acquisition, languages acquired after the acquisition of native language(s) or L1s are commonly referred to as second languages or L2s (e.g. Ortega 2009, 5). Languages acquired after the native language(s) are called second languages regardless of when they were acquired, and regardless of the learner's level of skill, ability or proficiency in that language. However, it must be noted that some definitions distinguish between languages acquired after first languages based on their status in the speaker's environment, dividing them into second languages, which are often official languages in the speaker's environment, and foreign languages, which have no official status (Pietilä & Lintunen 2014, 10-11). Still, this division is not untested. For example, the status of English in Finland may in fact be close to that of a de facto second language despite having no official status in the country (Leppänen et al. 2011, 168). This is true for many other countries globally: given the status of English as a global lingua franca, English has been described "an almost essential means for international communication" (Pietilä & Lintunen 2014, 9). Pietilä and Lintunen also predict that "[f]or future job markets, English may be taken for granted much like mathematical skills or IT skills, and so the importance of proficiency in other languages will increase" (*ibid*). Keeping this in mind, this thesis favours the terms L2 English and L2 proficiency over EFL (English as a foreign language) and EFL proficiency. The Finnish context of language acquisition and language testing is examined further in Section 3.2.

ii) Oral L2 proficiency. In second language acquisition and language testing, proficiency generally means the ability or skill in the second language. In second language acquisition and language testing, complexity, accuracy, and fluency are treated as separate but interrelated components of proficiency. The CAF framework is discussed in detail in Section 2.2. In the context of the present thesis, oral L2 proficiency is assessed on a criterion-based scale comprising six proficiency levels and six criteria, fluency being

one of them. The speaking scale of the Finnish National Certificates of Language Proficiency is discussed further in Section 2.3.

Finally, I have made a terminological choice between *L2 speaker* vs. *L2 learner*. In the present thesis, the two terms are used almost interchangeably, but use of the term *speaker* is favoured because the subjects of the present thesis are adults, many of whom use English with their family, at their workplace or in their studies without necessarily partaking in any formal or informal instruction. For many of them, their L2 may have already fossilized. The term is also effectively replaced by *candidate* while the National Certificates of Language Proficiency are being discussed. Finally, in the empirical section, the term *participant* replaces the terms mentioned above.

2.1 Approaches to fluency

The concept of fluency can be defined in broad and narrow terms. In fact, the first hurdle in approaching the concept of fluency is that the broad meaning of fluency in everyday uses is different to its uses in second language acquisition. Among broad definitions of fluency are everyday descriptions of “fluent” language use and among the latter the more field-specific definitions. One of the most cited definitions is Lennon’s distinction between broad (high-order) definitions of fluency, which refer to general oral proficiency in a second language, and narrow (low-order) definitions, where fluency is an operationalizable component of language proficiency with objective, quantifiable measures (Lennon 2000, 25). Many classic descriptions of spoken language fluency have to do with perceptions of cognitive fluency, or the apparent ease and smoothness of speech and the underlying psycholinguistic processes of speech planning and production (e.g. Lennon 1990). On the other hand, modern definitions of fluency define it along the lines of an “ability to produce the L2 with native-like rapidity, pausing, hesitation, or reformulation” (Housen, Kuiken & Vedder 2012, 2). The native speaker’s speech, with a special emphasis on speed, has long been the benchmark of L2 fluency: the native speaker is fast, effortless, and without *excessive* pauses (Lennon 1990, 360, emphasis added); and fluent speech is fast and effortless (Chambers 1997, 535). Most people can relatively easily differentiate between L2 speakers and native speakers, and it has been shown that even language learners can accurately identify more fluent speakers from less fluent speakers in their L2 (e.g. O’Brien 2014, 734).

In one of the early approaches to L2 fluency in the Finnish context, Lehtonen approached language proficiency in a quantitative way and touched on what later became

described as fluency by describing the L2 speech of Finnish learners of English as “too slow” with “more or less sporadic [pause placement] as regards the syntactic structure of the English sentences” (Lehtonen 1979, 37, 48). Internationally, attempts to quantify, measure and evaluate fluency and proficiency by studying the accurate and contextual use of language were made by e.g. Fillmore (1979), and on the other hand in communicative language teaching, a division between fluency and accuracy was emerging through attempts to study communicative L2 proficiency in the classroom (Housen, Kuiken & Vedder 2012, 2). Following this, Skehan (1998, 2009, Tavakoli & Skehan 2005) and Segalowitz (2010, 165), defined fluency as a component of L2 language proficiency comprising three dimensions: cognitive fluency, utterance fluency and perceived fluency. According to Segalowitz, cognitive fluency relates to the operational efficiency of underlying cognitive processes and cannot be measured directly, utterance fluency to acoustically measurable features of the actual utterances produced and can be measured directly via for example computer-assisted acoustic analysis and perceived fluency to the listener’s perception of the speaker’s utterance fluency, based on which inferences about the speaker’s cognitive fluency are made (Segalowitz 2016: 11-12). In other words, cognitive fluency is what happens in the mind of the speaker before and during the speech utterance; utterance fluency is the produced utterance, the kind of fluency which can be measured and analysed acoustically; and perceived fluency is how the hearers perceive the speaker’s fluency, i.e. the interpretation of the speaker’s cognitive fluency and how it manifests in their utterance fluency. As such, cognitive fluency affects utterance fluency, and together, cognitive fluency and utterance fluency affect perceived fluency (*ibid.*)

For years, fluency was under-defined in language teaching and testing, as pointed out by Fulcher (2003, 30). As such, fluency was also under-represented and under-defined in proficiency criteria. Since then, the CAF framework (Complexity, Accuracy, and Fluency) has been widely accepted as part of second language proficiency scales such as the Common European Framework of Reference for language proficiency, a criterion-based reference scale describing the achievement of language learners (Council of Europe 2011 [2001]). In addition, fluency is represented in the proficiency scales of the Finnish National Certificates of Language Proficiency.

While the present study only examines fluency, fluency is part of a framework concerning the interaction of three dimensions of language proficiency. The CAF

framework distinguishes three components of language proficiency: Complexity, Accuracy, and Fluency (e.g. Larsen-Freeman 2009, 582).

The origins of CAF in second language acquisition can be traced back to the 1970s and attempts to approach language proficiency in a quantitative way (e.g. Fillmore 1979, Lehtonen 1979). However, it was not until towards the end of the 1990s that the three dimensions of proficiency were combined into a proficiency model by Skehan (1998 and 2000). With Skehan's influence, new working definitions were attained for the three dimensions of proficiency, which Housen, Kuiken and Vedder summarise as follows (emphases added):

[C]omplexity is commonly characterized as the ability to use a wide and varied range of sophisticated structures and vocabulary in the L2, **accuracy** as the ability to produce target-like and error-free language, and **fluency** as the ability to produce the L2 with native-like rapidity, pausing, hesitation, or reformulation[.]

Housen, Kuiken & Vedder 2012, 2

Still, the interaction of the different dimensions of the CAF can be problematised: An agreement of the interaction of complexity, accuracy, and fluency in the field of SLA remains to be established as these dimensions have been treated as either complete or competitive and contributive. The Limited Attentional Capacity Model (Skehan 1998) argues that L2 learners must prioritise the objects of their attention during task performance, whereas Robinson's Multiple Resources Attentional Model (Robinson 2001, 2005) argues that learners' complexity and accuracy are related, and that increases in task complexity increase both complexity and accuracy, possibly but not inevitably compromising fluency. Robinson (2003) also proposes in his complete and contributive theory, that complexity, accuracy and fluency may, depending on conditions imposed by the task, work together to either beneficially or detrimentally affect L2 performance (Robinson 2003, Housen, Kuiken & Vedder 2012, 6-7). In contrast, Skehan's later Trade-off Hypothesis (Skehan 2009) treats the three dimensions of proficiency as performance constraints, or as competitive dimensions of performance, where a higher performance in one of the dimensions may lead to lower performance in one or two of the other dimensions. Thus, simultaneous high performance in all three is unusual and results in trade-offs in attention for complexity, accuracy, and fluency (Skehan 2009, 511). According to Housen, Kuiken & Vedder, no empirical evidence has been found in direct

support of either model, which is partially due to challenges in conceptualizing and operationalizing these three dimensions (Housen, Kuiken & Vedder 2012, 6-7)

In sum, complexity, accuracy, and fluency can be treated as either competitive or complete dimensions of proficiency, but neither theory has been proven. Depending on the view, the CAF dimensions either compete with one another, producing performance constraints, or may, depending on the demands of the task, work together to aid or hinder performance, possibly but not inevitably at the expense of fluency. Either way, the relationship of complexity, accuracy and fluency is not straightforward. Furthermore, these proficiency dimensions are often studied in the context of monologues. Emerging research is calling for new perspectives.

2.2 Operationalising fluency in the context of language testing

In this section, the proficiency dimension of fluency is operationalised by first introducing its traditional categorisation of temporal fluency features into speed, pause and repair phenomena, then moving on to challenges in operationalising fluency and emerging research, before finally discussing fluency in the context of the Finnish National Certificate of Language Proficiency.

Before discussing operationalising fluency in language testing, the umbrella term of language testing must be covered. Language testing is a field of study under applied linguistics. The objectives of language testing vary – language tests can assess an individual's performance, proficiency, achievement, or aptitude in their first, second or foreign language. Such assessments may include formal or informal, high-stakes or low-stakes, anonymous or public, individual, or collective assessments. Researchers in applied linguistics, second language acquisition and educational sciences use a wide variety of tests to measure a multitude of phenomena related to second and foreign languages. For example, language testing can be a tool of language learning and assessment, or a means of describing and demonstrating proficiency. Language proficiency is one of the objectives of language learning, and as such, a topic of great interest in second language acquisition studies. Furthermore, language tests can exert institutional control on individuals by for example controlling entry to important social roles, thus acting as instruments of societal advancement and ranking (e.g. McNamara 2000, 6).

Following this, language tests may have serious implications for the future of an individual. In fact, many language tests have a gatekeeper role in education and employment: many academic institutions require a certain level of language command from international students in the language of instruction and require applicants to take a

proficiency test to prove their qualifications. Likewise, applying for citizenship and entry into certain professions requires a command of the official language. As such, proficiency tests are often high-stakes tests: they may have a strong beneficial or adverse effect on an individual's career or education opportunities, or even their application for citizenship (McNamara 2000, 6-8). Examples of high-stakes language tests include the Finnish National Certificate of Language Proficiency, the TOEFL, and North Atlantic Treaty Organisation's STANAG 6001.

Given the tests' gatekeeper role, careful attention is paid to test design: a speaker's competence only becomes known to the interlocutor through their performance. Technological advances in neurolinguistics, particularly in brain imaging, may one day allow researchers to examine their subjects' internal processes directly, but in the meantime, proficiency is assessed via performance in tests, some of which are designed to tap into the internal processes of language processing indirectly (e.g. Olkkonen 2017). Performance in proficiency tests is seen as synonymous with language proficiency, and yet it is widely known that certain factors (such as working memory, attention, personality, and fatigue) affect and inhibit performance in test situations. In sum, there is a mismatch between competence and performance, i.e. the implicit knowledge a user has of a language and what they do in communication (VanPatten & Benati 2010, 124-125).

SLA literature distinguishes two primary categories of language tests according to their purpose: achievement and proficiency tests (e.g. McNamara 2000, 6-8). Achievement tests measure an individual's achievement in a certain curriculum or course: as such, they are intrinsically associated with language instruction and correspond to and support a curriculum. In short, achievement tests measure what learners have learned as a result of instruction; or, how much and how well the learner has learned what they were taught. In contrast, proficiency tests are concerned with language use in "real" contexts with criteria that represent and emulate future use. "Real" is in quotes; while proficiency tests aim for performance in a "real-life task", a test situation is still different from naturally occurring language use. Proficiency tests compare candidates' achievement to pre-selected criteria independent of formal language instruction, i.e. what the learners can do in the target language, and how appropriately and fluently (e.g. McNamara 2000, 6-8).

Proficiency test scores are typically set to a criterion-based scale. For example, in the National Certificates of Language Proficiency, candidates are tested in two receptive skills (listening and reading comprehension) and two productive skills (speaking and

writing) and assessed on a six-point proficiency scale comprising basic, intermediate and advanced levels of proficiency (Finnish National Agency for

Education 2011). This scale is comparable to the Common European Framework of Reference for language proficiency, a criterion-based reference scale describing the achievement of language learners (Council of Europe 2011 [2001]) and in which the CAF framework is well-represented. As such, let us move on to operationalising fluency in the context of language testing.

2.2.1 Temporal fluency

Early attempts of operationalising L2 fluency include featuring fluency measures in quantitative cross-linguistic studies. One of the classics of fluency studies dates to 1979, when Lehtonen contrasted the speech of Finnish and Swedish learners of English for temporal patterns and pause phenomena using speech rate, articulation rate, pause time and pause percentage as parameters. He described the L2 speech of Finnish learners of English “too slow”, and the number and placement of pauses in their speech as “more or less sporadic as regards the syntactic structure of the English sentences” (Lehtonen 1979, 37, 48). Studies contrasting language learner and native speaker speech are the staple of SLA, applied linguistics and contrastive linguistics.

Besides cross-linguistic studies, the CAF framework (Complexity, Accuracy and Fluency) has been a driving force behind operationalising L2 fluency. One of the founding fathers of modern fluency studies is Peter Skehan, who defined fluency as one of the three essential constituents of second language proficiency (Skehan 1998, 2000). Since then, three subdimensions have in turn been distinguished within fluency, and this divide continues to influence fluency studies today. Following Skehan (2003, 2009; Tavakoli & Skehan 2005), fluency can be divided into three subdimensions: speed fluency, or the rate and density of linguistic units produced, breakdown fluency, or the number, length and location of pauses, and repair fluency, or false starts, misformulations, self-corrections and repetitions (e.g. Housen, Kuiken & Vedder 2012, 5).

As for the operationalisation of fluency, performance can be measured across these three temporal dimensions of fluency using measures like speech rate and articulation rate for speed fluency, pause frequency, duration and location for breakdown fluency, and the frequency of repairs for repair fluency (De Jong et al. 2013, 894). In addition, fluency is closely related to proceduralisation. For example, Towell (2012) posits that fluency is largely the outcome of how well appropriate procedures for

processing acquired linguistic knowledge have been created within the procedural memory: according to Towell (2012, 55), speed fluency relies on storage and recall (access and retrieval). Speed fluency is operationalised as measures relating to speech delivery rate and density. Some of the most used operationalisations of speed fluency are speech rate and articulation rate. Speech rate, or the rate of speech delivery, has been cited as the fluency measure most accurately predicting oral proficiency (e.g. De Jong et al. 2012 and Kahng 2014). Speech rate is calculated as number of syllables per minute or syllables per second (De Jong et al. 2013: 894). On the other hand, articulation rate is associated with the motor skill of the speaker, as well as the automatising of speech patterns (e.g. De Jong & Perfetti 2011). Articulation rate is calculated as syllables per minute by dividing the total number of syllables by the total duration of articulated sample before multiplying the result by 60.

On the other hand, according to Towell (2012, 55) pauses and repairs are “related to the extent to which the learner is confident that [the stored linguistic knowledge] is reliable, and the extent to which the learner has created procedures which can be brought into operation to repair the situation when communication breakdown occurs”. Similarly, De Jong et al. (2013) posit that perceived fluency is affected by the number of unfilled pauses, repetitions, and repairs. O’Brien (2014) included the number of filled pauses as affecting perceived fluency. Pause location is important as well: the number or frequency of mid-clause pauses has been shown to affect fluency and are shown to be frequent in L2 speech as opposed to native speakers’ regular use of end-clause pauses. As such, pause phenomena can be operationalised as the number or frequency, location, and duration of breakdowns (namely unfilled and filled pauses.) Still, repairs are usually operationalised in much the same way as the other temporal fluency variables: number or frequency, duration, and location. Again, repairs are thought to adversely affect perceived fluency (e.g. De Jong et al. 2013), but qualitative analyses by Peltonen & Lintunen (2016) and Peltonen (2017) suggest that a one-dimensional interpretation of fluency is not justified: repairs have a multidimensional role in L2 speech.

The most reliable operationalization of temporal fluency may in fact be mean length of run. Mean length of run has been shown to accurately predict L2 oral proficiency, making it a reliable fluency measure (e.g. Kahng 2014). This composite measure comprises aspects of all three dimensions of fluency and can be calculated as the mean of syllables per unbroken run.

Following this, researchers have called for expanding the conceptualization of fluency beyond temporal phenomena (Wright and Tavakoli 2016, Peltonen 2017). For example, Peltonen (2017), Segalowitz (2016), and Wright and Tavakoli (2016) are advocating for extending fluency studies to interaction and the social dimension of fluency instead of conceptualizing fluency solely as a temporal phenomenon (Wright and Tavakoli 2016, Peltonen 2017). Furthermore, additional concepts have been studied in connection to fluency, including rhythm, adequacy (Revesz 2016), and problem-solving and strategy (Peltonen 2017).

As a result, emerging research is moving beyond temporal fluency and also into studying L2 speakers' pause and repair use qualitatively.

2.2.2 Beyond temporal fluency

As above, dividing fluency into speed, breakdown, and repair phenomena is a well-established way of operationalising temporal fluency. However, categorising different measures into the three temporal fluency dimensions is not always completely straightforward: instead of being completely independent, some fluency measures overlap. For instance, the oft-used speed fluency measure speech rate also contains pause and repair data. As such, if a more analytical approach is desired, it may be worthwhile to use articulation rate or phonation-time ratio as an alternative measure, i.e. to remove pauses from speed data. Still, speech rate as a more holistic variable contains data about non-proceduralised lexical, syntactic, phonological and suprasegmental knowledge, which in turns allows the interlocutor to make inferences about the speaker's cognitive fluency.

Furthermore, according to Kormos (2006), language learners' access to L2 knowledge is not yet automatic. This lack of proceduralisation is perceived by the interlocutor as disfluency: by definition, speech is slower when it contains disfluencies and interruptions, such as pauses and repairs, which often signals access issues (e.g. Nakatsuhara, Tavakoli & Awwad 2019, 7).

However, whether pauses and repairs simply subtract from fluency can be questioned. Studies published in recent years have shown that pauses and repairs may be more difficult to categorise as either adding or subtracting from fluency, and as such, may not work as one-dimensional temporal variables alone. Fluency was long studied quantitatively as de facto temporal fluency, i.e. the sum of speed fluency and certain quantifiable breakdown phenomena, with qualitative methods perhaps receiving less

attention. In fact, Peltonen & Lintunen (2016) combined quantitative and qualitative methods in their study of Finnish-speaking and Swedish-speaking learners of English at two school levels and found that only qualitative analysis could reveal differences in the use of filled pauses and repairs.

Furthermore, it has been shown that certain measures of fluency may be more reliable predictors of proficiency than others, and ideas of reliable measures have changed with the growing body of research. In addition, many factors may hinder fluent performance: not all dysfluencies result from low proficiency. In the past, studies have attempted to connect language fluency with personality. In a classic example, Dewaele and Furnham (2000) found moderate correlations ($r = 0.40$ to 0.55) between extraversion and several fluency measures, speech rate first among them. They also found that stressful conditions affected the fluency of participants who had scored high on introversion more than those who had scored high on extroversion, leading to more lexical and structural search and disfluency (in Ortega 2009, 197-198). A more recent example of how dysfluencies are not necessarily caused by low proficiency is a study by Olkkonen (2017), who studied the relationship between fluency of lexical access and second language proficiency, distinguishing between inefficiency of lexical access and control of attention. Olkkonen found that not all types of access and retrieval difficulty related to proficiency but rather, control of attention (Olkkonen 2017, 37).

In sum, while CAF and fluency studies are abundant today, and more measures and dimensions of fluency are attached to the growing body of research as time progresses, there is still some disagreement on precisely how different quantitative fluency measures and second language proficiency interact. It may be safest to assume that some features of fluency are more difficult to quantify than others: particularly pause and repair phenomena are problematic. In contrast, links between speed fluency measures and proficiency, as well as utterance length and fluency, are well-established.

In the case of pauses and repairs, research tends to focus on their frequency, duration, and location as opposed to how pause and repair opportunities are used by the speaker. Speech is online processing and requires on one hand a vast amount of strategic planning and on the other, proceduralisation, automatization and immense flexibility: complex processes that should be proceduralized enough that they conceal any language formulation difficulty. Regardless, it becomes apparent that measure and variable selection is of utmost importance in operationalising fluency, and that a mixed-methods approach is required to address the issue of pauses and repairs in L2 speakers' fluency.

2.2.3 Emerging research

In the recent years, fluency studies have focused on for example suprasegmental features, such as prosody and rhythm and their proceduralisation, and related them to fluency. On the other hand, mixed methods studies are an increasing trend in fluency studies: we should move away from simply keeping score of pauses and repairs. While language proficiency test scales describe frequent interruptions on the flow of speech as affecting fluency adversely (e.g. University of Jyväskylä: YKI speaking scale), pause and repair use are not one-dimensional in the sense that they simply subtract from fluency (e.g. Peltonen & Lintunen 2016). Instead of simply describing their number, frequency and duration in learner speech on different levels (utterance fluency) we should note that our perception of fluent speech comprises inferences of the speakers' cognitive fluency and strive to make these inferences visible. One way to do this is to categorise pause and repair use based on their inferred purpose.

Indeed, when extending the scope of study beyond temporal fluency features, qualitative methods should be utilized in the study of pause and repair phenomena to avoid simply describing L2 speakers' fluency and instead, making inferences about their cognitive fluency. One interesting line of inquiry is pause and repair use. In a recent study, Nakatsuhara, Tavakoli & Awwad (2019) distinguished three different pause use categories and ten subcategories based on the results of a qualitative inquiry into how pause use distinguishes between B2 and C1 levels of the APTIS speaking test. Based on a qualitative analysis, the researchers were able to relate L2 learners' pauses with i) facilitating access and retrieval of lexical and structural items, ii) reformulating previously produced units, ii) improving communicative effectiveness (Nakatsuhara, Tavakoli & Awwad 2019, 2). These categories are as follows:

- 1) **Pauses related to access and retrieval difficulty**, containing instances of
 - a) mid-clause pauses for lexical/structural search, followed by more sophisticated language (Lexical structure)
 - b) mid-clause pauses for lexical/structural search resulting in erroneous utterances or in generic expressions (Lexical structure)
 - c) pauses in the middle of / after producing sophisticated language (Lexical structure)
 - d) pauses to recall items from long-term memory (Memory)
- 2) **Pauses related to reformulations**, containing instances of
 - a) Mid/end-clause pauses occurring during / before reformulating ideas and utterances, and making self-corrections (Reformulating)
 - b) Mid-clause pauses in the middle of ungrammatical structures in the attempt of restructuring sentences (Rescuing)
- 3) **Pauses related to effective speech delivery**, containing instances of
 - a) Pauses before adding more information, examples and justifications (Topic development)
 - b) Mid-clause pauses before making evaluative comments and before expressing feelings (especially after an intensifier) (Attracting listeners' attention)
 - c) End-clause pauses before topic shift (Topic shift)
 - d) Turn-initial pauses before dispreferred responses (Dispreference)

Nakatsuhara, Tavakoli & Awwad 2019, 26

As per the analysis, L2 speakers used pauses for access and retrieval of lexical and structural items (and monitoring them), reformulations, and making their communicating and delivery more effective. Attempts to search for lexical and structural items were associated with mid-clause pauses as opposed to clause-border or end-clause pauses. How successful this search was differentiated between B2 and C1 levels of the APTIS speaking test: C1 speakers were able to more successfully use pause opportunities for access and retrieval of lexical and structural items, producing sophisticated language more often than B2 speakers, whose pause opportunities relating to access and retrieval often resulted in erroneous or generic language. In addition, pauses and repairs were found to be linked to the need to monitor and repair speech (Nakatsuhara, Tavakoli & Awwad 2019). Finally, Nakatsuhara, Tavakoli & Awwad found that participants on C1 level used their pause opportunities for making their delivery more effective and that B2 speakers were less effective in this. In sum, while both groups used pauses for access and retrieval, and reformulation and repair, speakers at the C1 level proved more successful in producing correct language and used pause opportunities for making their speech more effective as

opposed to monitoring and correcting minor errors (Nakatsuhara, Tavakoli & Awwad 2019, 37-38). Furthermore, similar results were reported for repairs as pause and repair were found to be connected: pauses often precede repairs, and it has been noted that repairs are common even at higher levels of proficiency (ibid.). Repairs and repetitions may also imply that the speaker aims towards accuracy and avoids pausing. As such, it follows that the relationship between pause and repair use and fluency may be non-linear. While some temporal measures, such as speech and articulation rates and mean length of run, are proven to be reliable predictors of proficiency, operationalisations of fluency in proficiency testing criteria should be updated to reflect the multidimensional role of pauses and repairs in L2 speakers' speech.

To summarise, fluency studies are still aiming for comprehensive definitions, descriptions, and operationalisations of fluency. Most operationalisations of fluency include temporal fluency measures relating to the speed and density of speech and the frequency, density, location and type of pause and repair phenomena in some form or capacity. Still, a body of research moving beyond temporal fluency is developing, and new lines of inquiry are being opening towards the relationship between fluency and suprasegmentals, and on the other hand, social dimensions of fluency, including interaction, conversation strategies and problem-solving (Peltonen 2017). On the other hand, there is a need for mixed-methods studies tapping into L2 speakers' cognitive fluency and the causes of disfluency. Finally, results by Nakatsuhara, Tavakoli & Awwad (2019) indicate a need for a mixed-methods study combining a quantitative analysis temporal fluency features and a qualitative analysis of intermediate and advanced L2 speakers' pause and repair use.

2.3 Fluency and the Finnish National Certificates of Language

Proficiency

Before moving on to fluency and the Finnish National Certificates of Language Proficiency, some attention must be given to the Finnish L2 context. In Finland, English is undeniably the lingua franca at many a workplace, and a basic proficiency in English is often assumed, if not taken for granted. For example, Pietilä and Lintunen (2014) describe English “an almost essential means for international communication” and also predict that “[f]or future job markets, English may be taken for granted much like mathematical or IT skills, and so the importance of proficiency in other languages will increase” (Pietilä & Lintunen 2014, 9). English is typically the first foreign language a

child encounters in formal education, the language most widely studied, and the foreign language most used (Leppänen et al. 2011, 20). For decades, formal instruction in the English language started at nine years of age. Since then, the age of onset has been lowered to seven. Studies have shown that earlier age of onset leads to better language ability (e.g. Flege et al. 1999), but this is yet to be reflected in any fluency studies. Finns usually study English for a minimum of 7 years, and additional courses are available in upper secondary and vocational education. Additionally, there is some compulsory English instruction in higher education. Being the most widely studied language, and the foreign language most commonly used, English in Finland appears to have a special status somewhere between a second and a foreign language. A survey by Leppänen et al. (2011) confirms the robust presence of English in Finland. The results indicate that given the positive attitudes and active use of English among certain groups, “Finland might, even now, be considered a country in which English has the status of a second language [...] or of a “third national language” (Leppänen et al. 2011, 168). Thus, English appears to have a special status somewhere between a second and a foreign language – but only for “certain groups” (ibid.) Interestingly, the survey revealed some socioeconomic differences in English use, as well as attitudes towards the language between respondent groups. Differences were especially prominent between the old and relatively low-educated, and the urban, relatively well-educated younger generations. The latter use English deftly in their day-to-day communication functions (Leppänen et al. 2011, 124-125).

On the other hand, language testing in Finland is strongly associated with national curriculums for basic education, the Finnish Upper Secondary school, and compulsory foreign language studies in higher education. The education system uses comparatively few standardised tests: the national examinations on the ninth grade of compulsory education, and the matriculation examinations at the end of Upper Secondary School measure the learners’ achievements in their curriculums. In contrast, the Finnish National Certificates of Language Proficiency, or the YKI test, is a language proficiency test by the Finnish National Agency for Education and is intended for measuring “the functional language proficiency of adults [...] [Functional language proficiency] refers to how appropriately and fluently an individual is expected to manage various situations and tasks that require language comprehension and production” (Finnish National Agency for Education 2011, 7). Test scores may be of interest to possible employers, may be used to prove proficiency for international work assignments or education, or in the case of

Finnish and Swedish, to demonstrate the level of language proficiency required for acquiring Finnish citizenship. The certificate may also be used to determine and demonstrate language proficiency in the four language subdomains of language proficiency, i.e. reading comprehension, listening comprehension, writing and speaking, independent of curricula and regardless of candidates' age, level of formal education or language instruction. Other suggested uses include using the test scale in planning curriculums and for candidates' self-assessment.

In the YKI test, candidates are assessed on a six-point proficiency scale. Their receptive and productive language are assessed according to specific assessment criteria, and the certificate contains a separate assessment for each skill. The test itself comprises classic pen-on-paper tasks (listening and reading comprehension) and performance tasks (speaking and writing). Prior to testing, candidates must choose between basic, intermediate, and advanced level tests. This choice is based on self-assessment with the help of proficiency level descriptions provided by the Finnish National Agency for Education.

Candidates may only take the test at one proficiency level at a time. This effectively means that candidates can only score at the level of their choice (1 or 2 at Basic level; 3 or 4 at Intermediate level; and 5 and 6 at Advanced level). However, the Certificate also indicates cases in which the candidates' achievement fails to meet the criteria of their chosen test: "In the case of the basic level test, the language proficiency assessment can be less than 1, 1 or 2; at Intermediate level less than 3, 3 or 4; and at Advanced level less than 5, 5 or 6. If the test-taker has totally omitted a particular subtest, or has only partly completed it, this will be indicated on the certificate as 'cannot be assessed'" (Finnish National Agency for Education 2011, 12). Furthermore, the six-point proficiency scale used in the YKI test corresponds to the CEFR scale as follows: on the basic level, YKI levels 1 and 2 correspond to CEFR A1 and A2; on the intermediate level, YKI levels 3 and 4 correspond to CEFR levels B1 and B2; and finally, on the advanced level, YKI levels 5 and 6 correspond to CEFR levels C1 and C2 (Finnish National Agency for Education 2011, 22).

In connection with the proficiency test, samples and background data from the participants are collected into the YKI corpus, a dynamic corpus compiled by the Centre for Applied Language Studies (CALS) at the University of Jyväskylä. The corpus contains both quantitative and qualitative data in several target languages, includes three written and one oral performance for each candidate, accompanied by the candidates'

proficiency assessments for each of the four subtests (reading comprehension, listening comprehension, writing and speaking) as well as self-assessments and background information.

While the YKI speaking scale aims for an analytical treatment of proficiency, a brief analysis of the YKI speaking scale reveals that the descriptors overlap: while there is a separate descriptor for the fluency criterion, linguistic features which may also be categorised under fluency are found in both the general proficiency level description and descriptors for other criteria. The criteria with most overlap include the criteria for flexibility and coherence/cohesion. For example, the flexibility criterion on level 6 (CEFR C2) states that “[the participant] can flexibly reformulate their thoughts using different linguistic forms to [1] eliminate ambiguities, [2] indicate emphasis, and [3] vary their speech according to the interlocutor and situation” (University of Jyväskylä 2011: Speaking scale). While the flexibility criterion is mostly about adequacy and using the correct register, the notion of flexible reformulation of thoughts also fits the fluency criterion.

As above, strictly analytical treatment of natural language can be difficult and, in some cases, impractical and even impossible. Our interpretation of speech is by nature holistic, and fluency features may be difficult to isolate. As such, many of the YKI speaking scale level descriptors may be interpreted to contain descriptions of fluency. Next, these descriptions shall be analysed for references towards fluency and hints of its operationalisation, with special attention to pause and repair use.

On the intermediate level, the general description describes the speech of candidates assessed at the intermediate proficiency level three (3) of the YKI scale “fairly slow” with “few unnatural interruptions”. As described earlier in Section 2.2.1, fluency features related to speed are well-established indicators of fluency and proficiency. Speed fluency features are often operationalised as for example speech rate (syllables per minute), articulation rate (syllables per minute with pauses removed) and phonation-time ratio (“the percentage of time spent speaking as a proportion of the total time taken to produce the speech sample” (De Jong & Perfetti 2011, 538)). On the other hand, the YKI scale mention of “few unnatural interruptions” refers to the number of interruptions: whether this includes filled pauses in addition to unfilled pauses is unclear. It is also unclear whether repairs are included as interruptions of speech here. Furthermore, the YKI speaking scale also describes candidates’ speech at level 3 “comprehensible”. Comprehensibility is a concept that has been studied in connection with fluency, and it

seems that these two are closely related. For example, O'Brien (2014) was interested in learners' perceptions of fluency, accentedness and comprehensibility in native and non-native speech and found that learner-raters found less fluent speech less comprehensible and more accented than more fluent speech (O'Brien 2014, 734-741). In addition, the coherence/cohesion descriptor also describes that candidates assessed at level 3 can "combine expressions into coherent and cohesive speech even though the use of connectors may be incomplete and repetitive". While this coherence/cohesion descriptor mostly describes the use of connectors, *incompleteness* and *repetitiveness* may hint towards pauses and repairs. Finally, according to the fluency descriptor, candidates assessed at level 3 of the YKI speaking test can "express themselves comprehensibly and relatively easily without help" and that "pauses related to difficulties in speech formulation are common, especially in longer, continuous utterances" (University of Jyväskylä 2011, own translation). Again, the fluency descriptor mentions comprehensibility. In addition, ease or effortlessness are words used to describe fluent speech in many fluency definitions: at level 3, candidates may struggle to express themselves, but not so much that it affects the speaker's comprehensibility or the perceived ease of their self-expression overly much. Pauses related to speech formulation are common at intermediate level 3.

As for proficiency level 4, according to the general description, candidates are "obliged only rarely to use circumlocutions in everyday communication because of inadequate language proficiency" and according to the fluency descriptor, "long pauses are rare even though hesitation may occur when searching for structural and lexical items" (University of Jyväskylä 2011, own translation). According to the general description, circumlocutions related to inadequate language proficiency and long pauses are rare, but candidates have long pauses and hesitation related to lexical and structural access and retrieval difficulty in their speech. What is meant by hesitation here (hesitation devices, all pausing phenomena, filled pauses) is not specified.

In contrast, on the advanced level, the general description states that candidates assessed at proficiency level five (5) of the YKI scale speak "fluently without frequent obvious need to search for an expression", and that their "delivery [is] characterised by naturalness, coherence and appropriate length". They can also use "idiomatic and everyday expressions and are able to express nuances fairly well, even though the use of less common vocabulary and complex sentence structures may cause difficulties". According to the fluency descriptor, candidates at this level are capable of "expressing

themselves somewhat naturally, fluently and spontaneously”, and “linguistic hesitation may only occur in terminologically difficult topics” (University of Jyväskylä 2011, own translation). According to these descriptions, advanced candidates assessed at level 5 should not have “frequent obvious need” for lexical and structural search, i.e. pauses and repairs related to access and retrieval difficulty. Their speech is characterised as natural and coherent, and they can express themselves, but less common vocabulary and complex sentence structures may cause difficulties. From a pause and repair use point of view, “difficulties” may imply rescue and reformulation pauses and repairs as well as access and retrieval difficulty pauses and repairs. In addition, naturalness, fluency, and spontaneity are mentioned – of these, spontaneity implies flexibility, apparent lack of planning and effortlessness. Finally, linguistic hesitation only occurring in connection with terminologically difficult topics suggests that pauses and repairs related to access and retrieval difficulty might occur in connection with more sophisticated ideas and language.

Finally, in addition to the flexibility criterion discussed above, at proficiency level six (6), descriptors stress fluency, near-absence of non-native features, ability to vary speech linguistically and with regard to the content, and expressing oneself “fluently, naturally and without hesitation even in long speeches [...] only occasionally pause to search for the correct word in order to express their thoughts or to find a fitting example or explanation” (University of Jyväskylä 2011, own translation). This suggests a degree of flexibility as opposed to access and retrieval difficulty.

In sum, due to the nature of speech, linguistic features are difficult to isolate. However, regarding fluency and pause and repair use, the descriptors suggest that as proficiency, fluency, flexibility and coherence/cohesion increase, interruptions and hesitation related to formulation of speech, access and retrieval difficulty and reformulations decrease. Given a growing body of research suggesting that the role of pauses and repairs in cognitive fluency is far from one-dimensional, we now move on to Section 3, the present study.

3 The present study

This section introduces the data and methodology of the present thesis. First, the study design, along with the research questions and hypotheses, will be presented in section 3.1. Section 3.2 introduces data and methodology: participants will be introduced in section 3.2.1 and transcriptions of their speech samples in 3.2.2. The data collection process for RQ1 is presented in 3.2.3, as well as the extreme case sampling for research question 2. Finally, Section 3.2.4. presents data collection process for research questions 2 and 3.

3.1 Study design

The present study is a post-hoc study of intermediate (N = 30) and advanced (N =30) participants taking the YKI speaking test and utilizes both quantitative and qualitative research methods. As described in Section 2, the need for a mixed-methods approach in studying fluency, and particularly pause and repair phenomena, is well-established. As such, the present study combined a quantitative analysis of fluency variables with a qualitative analysis of pauses and repairs to answer the following research questions:

- 1) Which quantitative measures of fluency distinguish between intermediate and advanced levels of proficiency of the Finnish National Certificates of Language Proficiency speaking test?
- 2) How do pause and repair use distinguish between intermediate and advanced proficiency levels of the Finnish National Certificates of Language Proficiency speaking test?
- 3) Are pause and repair use interdependent?

First, as for question 1, I hypothesised that the speed fluency measures, measures relating to pause duration and frequency, and repair frequency distinguish between intermediate and advanced levels of the YKI speaking test (De Jong et al. 2013, De Jong 2016, Kahng 2014). Considering pause and repair location, I expected mid-clause pauses and repairs to distinguish between the two proficiency levels. As for research question 2, I hypothesised that intermediate and advanced participants use their pause opportunities for different purposes. As per Nakatsuhara, Tavakoli & Awwad 2019, I expected both groups to use pause opportunities on pauses related to access and retrieval difficulty, but

I expected advanced participants to be more successful in their search for lexical and structural units. I also hypothesised that intermediate participants use more of their pause opportunities on pauses related to reformulation than advanced participants. I also expected advanced participants to use more of their pause opportunities on pauses related to effective speech delivery than the intermediate participants. Furthermore, I expected extreme case sampling for qualitative analysis based on quantifiable fluency features to distinguish between pause and repair use between the two levels of proficiency, potentially leading to interesting findings on proficiency level thresholds. Finally, as for research question 3, I expected pauses and repairs to be interconnected, as per Nakatsuhara, Tavakoli & Awwad (2019, 37).

In order to approach these three research questions and to compare oral L2 fluency on intermediate and advanced levels of proficiency, a total of 60 speech samples (30 from intermediate and 30 from advanced candidates) were selected from the Finnish National Certificate for Language Proficiency Corpus, or the YKI corpus (the candidates, from here on referred to as *participants*, are described in detail in Section 3.2.1). Next, the speech samples were transcribed and annotated for pauses and repairs. This process, the quantitative fluency measures selected, and the following statistical analysis, revealing which quantitative fluency measures best distinguish between the two proficiency levels, are described in detail in Section 3.2.3. Next, based on the results of the quantitative analysis, 6 extreme case samples from each group were selected (12 in total, discussed in detail in 3.2.4). Transcriptions of the participants speech samples were then revisited, and their pauses and repairs encoded and analysed based on a categorisation proposed by Nakatsuhara, Tavakoli & Awwad (2019, 26). Finally, Section 4 presents results of the quantitative and qualitative analyses. Next, the study design described above is presented in Figure 1:

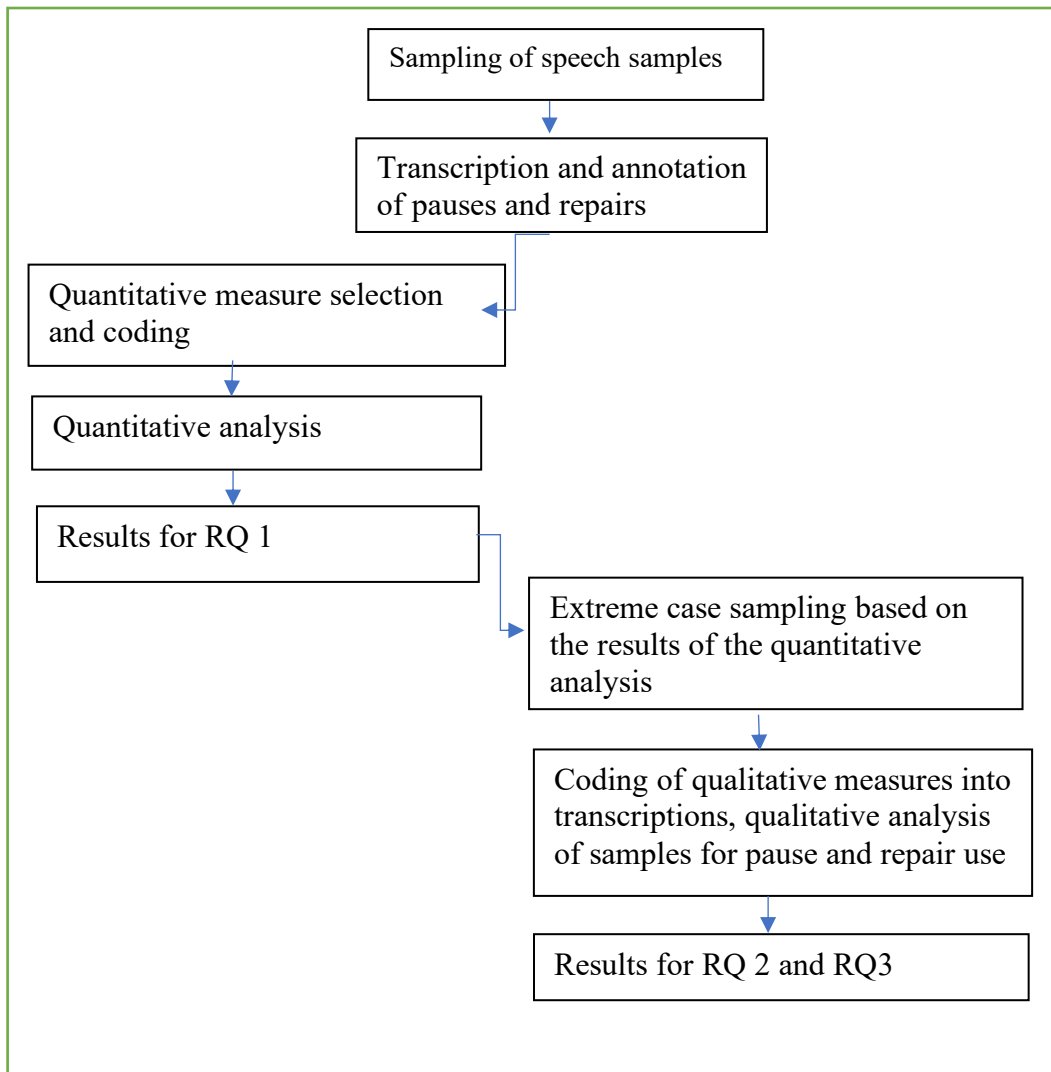


Figure 1 Study design

3.2 Data and methodology

This section introduces the data and methodology of the present thesis, starting with a detailed description of participants. Materials and transcriptions are introduced after this, including the coding of quantitative variables. We then move on to extreme case sampling and finally, qualitative data.

3.2.1 Participants

The participants (N=60, male N=30, female N=30) were Finnish-speaking adults taking the Finnish National Certificate of Language Proficiency, or the YKI test, in English at either intermediate level (proficiency assessment 3 and 4, N = 30) or advanced level (proficiency assessment 5 and 6, N = 30) after the YKI test was redesigned in 2011. Differences in subtests and background information forms between the old and new tests discouraged the use of material acquired before 2011. For the purposes of this study, participants were controlled for gender, age and level of education; namely, the groups were expected to be similar in age and gender distribution, and to have at least secondary education or higher.

The intermediate group (N = 30) were 14 males and 16 females, their ages ranging between 22 and 57 years. In this group, 2 participants reported having studied English for 4-6 years, 9 for 7-9 years and 19 ten years or more. In this group, 8 reported upper secondary education, 3 vocational education, 7 polytechnic or university of applied sciences and 12 university as their highest level of education.

The advanced group (N = 30) were 16 males and 14 females, their ages ranging between 24 and 53 years. In this group, 6 participants reported having studied English for 7-9 years, 9 for 7-9 years and 24 for ten years or more. In this group, 4 reported upper secondary education and 26 university as their highest level of education.

Despite careful sampling, there were small differences between the two participant groups. First, participants in the intermediate group were slightly younger on average (mean = 33.13, median = 30, SD = 9.001, minimum 22, maximum 57, range = 35) than advanced participants (mean = 35.9, median = 32.5, SD = 9.718, minimum = 24, maximum = 53, range = 29). This small difference in the groups' mean ages was not statistically significant. Second, on average, advanced participants reported having studied English longer than participants at the intermediate level. In addition, while sampling ensures that all participants had received at least secondary education, it should be noted that university education is more common in the advanced group. Still, Finnish

adults are expected to have similar levels of language proficiency at university and polytechnic, and at both branches of secondary education, i.e. upper secondary school and vocational school. Whether this is de facto true is more difficult to assess, and beyond the scope of this study. In sum, it can be concluded that the groups are similar in age, gender distribution, and education.

Furthermore, as some attention must be paid to ethical considerations even in post-hoc studies such as the present thesis, the participants' identities had been anonymised by providing them numerical identifiers before their data were encoded in the YKI corpus, and while the present study cannot be held responsible for the participants' ethical treatment during testing or the appropriate encryption, encoding or processing of their data with regard to European data protection laws, it can be concluded that appropriate data permits were granted to the author of this study by the University of Jyväskylä, and that applicable participant data were handled appropriately and with utmost care.

3.2.2 Material and transcriptions

As the present study is a post-hoc analysis of speech samples extracted from one task in the Finnish National Certificates of Language Proficiency speaking test, the test and task design are beyond the scope of this post hoc thesis. However, a brief description of the YKI test is included here for the sake of clarity:

The YKI test has been designed with minimal variation between different proficiency levels, but their “degree of difficulty, including vocabulary, topics and language functions is tailored to fit the test level” (Finnish National Agency for Education 2011, 9). As such, the required level and variety of expression (and thus, the level of proficiency) varies by test level (beginner, intermediate, and advanced). There are ten topic categories present across all three levels: personal identification, home and living, retail and services, culture, travel, health and wellbeing, work, environment, and society (Finnish National Agency for Education 2011, 9). Likewise, the tasks across the three tests draw from a pool of six communication functions, namely “giving and asking for factual information, expressing opinions and attitudes, expressing and enquiring about emotions, dealing with transactional activities, acting according to social norms and customs, and communication strategies (Finnish National Agency for Education 2011, 10)

As for the task for which the speech samples used in this study were produced, participants had two minutes to plan their speech before speaking for two minutes on a predetermined topic. The task was similar on both intermediate and advanced levels of proficiency, and some participants indicate in their samples that they were given a choice between topics. Four topics appear across samples used in the current study, two on the intermediate and two on the advanced level.

As for the transcriptions, a total of 60 speech samples (~2 minutes each) were transcribed and annotated for speed, pause and repair phenomena. The transcription process was partially computer-aided: the speech-to-text software Sonocent Audio Notetaker was used together with Dragon NaturallySpeaking to turn audio into rough transcription drafts. This was then followed a more in-depth analysis and annotation on Praat, which was used to verify any uncertain sounds, junctures, and interference. To aid the analysis, a script by De Jong & Wempe (2009) was used to automatically detect syllable nuclei in recordings and waveforms analysed on Praat (sample transcription in appendix 1). Praat was also used to identify the location of and measure the duration of unfilled pauses in the recordings at the silence threshold of -25 dB and minimum silent interval duration of 300 ms. This particular interval was chosen in order to avoid annotating e.g. plosive occlusion phases as unfilled pauses. Lehtonen (1979) examined pauses at or over 200 milliseconds, but in the present study, utterances between ≥ 300 ms unfilled pauses and pause-hesitation clusters are treated as uninterrupted, unbroken runs.

For the quantitative analysis, the transcriptions were annotated for:

- 1) Unfilled pause duration (where pauses were longer than 300 ms. Pauses shorter than 300 ms were treated as micropauses and were excluded from analysis but annotated to describe the subject's speech rhythm (see appendices 1 and 2). Audible breathing, coughs, sniffles, and laughter were included in the unfilled pauses, as well as any non-verbal hesitation within unfilled pauses (i.e. filled pauses and pause clusters).
- 2) Pause location (mid-clause / end-clause).
- 3) Repair location (mid-clause / end-clause), including repetitions, reformulations, false starts, and comparable repair phenomena.

Furthermore, repairs, repetitions, reformulations, lexical hesitation (e.g. *well*, *yeah* and *okay*) and non-lexical monosyllabic hesitation (e.g. *er*, *um*, *mm*) appearing mid-run (as opposed to directly before, after or within unfilled pauses) contributed to the syllable count (and therefore articulation rate and mean lengths of run). Breathing

(annotated `{*h*}` or `{*hh*}` depending on intensity and duration), dental clicks (`{*pt*}`), sniffles (`{*sf*}`), chuckles (`{*heh*}`), and coughs (`{*c*}`) were excluded from analysis where not included in unfilled pauses, and punctuation was transcribed only where made obvious by intonation. Colons after words in the transcription denote that the preceding sound has an atypically long duration (>300 ms).

3.2.3 Quantitative data

Quantitative data in the present study includes 7 quantitative fluency measures operationalised as 7 variables. These quantitative operationalisations of fluency are presented in Table 1.

Fluency features	Variables	Explanation
Articulation rate	Articulation rate (spm)	Number of syllables per minute, unfilled pauses removed
Mean length of run	Mean length of run (syllables)	Mean length of unbroken run, i.e. syllables between unfilled pauses or pause clusters
Pause duration	Total pause duration per minute	Total unfilled pause durations, standardised for one minute
Pause frequency	Pause frequency per minute	The frequency of unfilled pauses, standardised for one minute
Pause location	Mid-clause pauses per minute	The frequency of unfilled pauses in mid-clause position, standardised for one minute
Repair frequency	Repairs per minute	The frequency of repairs per minute
Repair location	Mid-clause repairs per minute	The frequency of mid-clause repairs per minute

Table 1 Quantitative fluency measures selected for the present study

Next, the fluency measures presented in Table 1 are explained. In the present study, speed fluency was examined through one quantifier, *articulation rate*. Fluency studies associate higher speech rates and higher rates of articulation with more fluent speech, but the problem with speech rate is that it is not independent of pauses. As such, in this study, speed was measured by articulation rate alone. Articulation rate is associated with the motor skill of the speaker, as well as the automatization of speech patterns (De Jong & Perfetti 2011). Participants' articulation rates were extracted by dividing the total number of syllables by the total duration of articulated sample, i.e. the sample with unfilled pauses removed, before multiplying the result by 60.

Likewise in fluency studies, breakdown fluency concerns the frequency, duration and placement of unfilled pauses which interrupt the continuous flow of speech, causing disfluency (e.g. Bosker et al. [2012] 2013, Tavakoli & Skehan 2005). In the present study,

breakdown fluency was examined through three quantitative measures, *total pause duration*, *pause frequency* and *pause location*. Individuals with less frequent unfilled pauses, and conversely, longer unbroken runs, are perceived more fluent and more proficient than speakers with shorter unbroken runs: short unbroken runs lend an atypical ‘staccato’ quality to the speech (e.g. Lehtonen 1979). In the present study, silences longer than 300 milliseconds were treated as unfilled pauses. However, often these unfilled pauses contained hesitation syllables (e.g. er, um, eh). The unfilled/filled pause clusters were treated as units and their total lengths were recorded in the transcriptions. Furthermore, as described on page 13, pause location has proved a reliable measure of fluency and proficiency: mid-clause pause frequency tends to be higher in the speech of less fluent and less proficient language speakers. For clarity, total pause duration is the total unfilled pause duration in seconds, standardised for one minute, pause frequency the number of unfilled pauses, standardised for one minute, and pause location the number of mid-clause pauses per one minute.

In addition, individuals with a lower number of repairs are often considered more fluent speakers. In the present study, two measures quantify repair fluency: repair frequency and repair location. Repair frequency is operationalised as the number of repairs, standardised per one minute, and repair location as the frequency of mid-clause repairs per one minute. In the present study, repairs include all cases of partial or complete lexical repetitions and reformulations.

In addition to speed, breakdown and repair fluency measures used in the present study, participants’ mean length of run was also measured. Mean length of run is a composite measure comprising elements of speed, breakdown and repair fluency and is operationalised as the total number of syllables divided by the number of unbroken runs.

The measures described above translated to scale variables which can be analysed statistically on SPSS to test whether any differences found between the two groups are statistically significant. The two proficiency groups’ mean values were then compared using either independent samples T-tests or Mann-Whitney U tests, depending on whether the variables are normally distributed in both groups (as per e.g. Larson-Hall 2016).

Extreme case sampling. Given that qualitative analysis of 60 speech samples is time-consuming and well beyond the scope of this study, the sample needed to be narrowed down prior to analysis. As such, extreme case sampling was chosen as the sampling method. While results acquired through extreme case sampling are perhaps less generalisable than representative or random samples, Dörnyei (2007, 153) suggests that

the sampling method works for singling out cases to reveal new information, which in turn may lead to new conceptualisations and findings, befitting the objectives of the present study.

Therefore, after the 60 speech samples were analysed quantitatively, the quantitative variable which proved the most effective at distinguishing between intermediate and advanced levels of proficiency, i.e. produced the largest difference, was chosen as the basis of extreme case sampling. Six samples were chosen from both proficiency groups (three with the lowest value, and three with the highest of the chosen quantitative variable), producing 12 speech samples for qualitative analysis of pause and repair use. In other words, in selecting three high-performing and three low-performing participants from both proficiency groups, two six-participant samples comprising participants who may be perceived the least fluent, and participants who may be perceived the most fluent are selected with the hopes that this may lead to new insights especially at the proficiency level threshold.

3.2.4 Qualitative data

After the 60 transcriptions were analysed quantitatively and an extreme case sample of 12 speech samples were chosen on the basis of the best-distinguishing quantitative variable, the 12 sample transcriptions were reworked and analysed qualitatively for pause use, as well as annotated based on the typology by Nakatsuhara, Tavakoli & Awwad (2019), presented on page 15. Pauses selected for analysis were expected to meet the following criteria: i) they must be a minimum of 300 ms long, and ii) they must be categorizable. The categorisation requires making inferences of the speaker's cognitive fluency, and clearly ambiguous pauses, such as pauses resulting from involuntary coughing, should be excluded from analysis. In addition, participants' repair use were also analysed and annotated into the transcriptions. Again, in the event of ambiguity, cases affected were excluded from analysis.

In Section 3, I hypothesised that pauses and repairs co-occur and are interconnected. As such, I expected that repairs may also be categorised as repairs related to access and retrieval, reformulations, and effective speech delivery to see whether there is an association between pauses and repairs. As such, following Nakatsuhara, Tavakoli & Awwad (2019, 26), repairs are categorised as repairs related to access and retrieval difficulty (1a-d), as repairs related to reformulations (2a-b), and repairs related to effective delivery (3a-d).

It should be noted that Nakatsuhara, Tavakoli & Awwad (2019, 15) also used a classification of lexis into frequent and less frequent words, which determined whether an expression was treated as sophisticated or generic. No such references were used for the present study in an effort to limit the scope of the study. Instead, language was evaluated on a case-by-case basis and related to the linguistic context, i.e. surrounding language. Once the 12 transcriptions were once more annotated for pause and repair use categories, the transcriptions were analysed for patterns, providing results for research question 2. It is interesting to see whether pause and repair use are related to individual speaking styles (e.g. Kahng 2014) or whether speakers at different levels of proficiency use pause and repair opportunities to aid their speech in different ways. Finally, by combining quantitative and qualitative analysis, results are acquired for research question 3, i.e. whether pause and repair use are interdependent, as shown by Nakatsuhara, Tavakoli & Awwad (2019, 26). The next Section presents these results.

4 Results

This section describes the results of the present mixed methods study, addressing all research questions and hypotheses. These results are complemented by several figures and tables, some of which are found in Appendices 2-3. First, Section 4.1 presents the quantitative results and research question 1 and addresses the extreme case sampling for qualitative analysis. Second, section 4.2. presents the qualitative results and research question 2. Finally, Section 4.3 presents the results to research question 3.

4.1 Quantitative fluency measures distinguishing between intermediate and advanced levels of YKI

This section introduces the results of the quantitative analysis, answering research question 1, i.e. which quantitative fluency measures distinguish between intermediate and advanced levels of proficiency of the Finnish National Certificates of Language Proficiency in English? As for my initial hypothesis, I expect the advanced group to have higher articulation rates, mean lengths of run, lower pause frequencies and total pause lengths, higher mean lengths of run, lower pause frequencies and total pause durations as well as lower mid-clause pause frequencies and higher repair frequencies than the intermediate group.

In the 118,15 minutes of samples analysed, participants spoke on average 116.8 seconds (SD = 8.1) on the intermediate and 119.5 seconds 5 (SD = 1.06) on the advanced level of proficiency. Shapiro-Wilk tests were run on all 8 fluency variables to test for normality before statistical analyses on SPSS. A normal distribution ($p > .05$) was found for the following fluency variables for both intermediate and advanced proficiency groups:

- Articulation rate
- Pause duration
- Pause frequency
- Mid-clause pause frequency

For these, independent samples t-tests were run in SPSS. For measures for which no normal distributions were found, Mann-Whitney U tests were run instead in SPSS. These variables were:

- Repair frequency
- Mid-clause repair frequency

Furthermore, Levene's tests for equality of variances were run on each normally distributed measure to ensure equal variances. All normally distributed variables were found to have equal variances.

Next, the results of the quantitative analysis are presented. These quantitative results are summarised in Table 2 below (statistically significant results highlighted in blue).

	INT (N=30)		ADV (N=30)		Finding	T-test		Mann-Whitney U					
	Mean	SD	Mean	SD		t	df	Sig. (2-tail.)	MWU	Rank (INT)	Rank (ADV)	SE	Sig.
AR	215	34.3	259	29.8	ADV 20.5% > INT	-5.3	58	p < .01					
Pause duration	22.1	1.2	16.4	0.9	ADV 25.8% < INT	3.75	58	p < .01					
Pause freq.	23.2	0.7	20.1	0.8	ADV 13.4% < INT	2.93	58	p < .05					
MC pause freq.	14.1	3.7	12.0	3.3	ADV 14.3% < INT	2.50	58	p < .05					
MLR	5.9	2.2	8.7	2.9	ADV 47.5% > INT				721.5	21.47	39.53	67.6	p < .01
Repair freq.	4.9	3.1	5.8	3.4	No sig. diff.				545.0	27.32	33.68	67.6	p > .05
MC repair freq.	3.4	2.3	4.4	2.4	ADV 29.4% > INT				598.0	25.57	35.43	67.6	p < .05

AR=articulation rate, MC pause freq.=mid-clause pause frequency, MLR=mean length of run, MC repair freq.=mid-clause repair frequency

Table 2 Statistical analysis: summary of quantitative results.

As per Table 2, normal distributions were found for articulation rates, total pause durations, pause frequencies and mid-clause frequencies of intermediate and advanced participants, and subsequent independent samples t-tests revealed statistically significant differences between the intermediate and advanced participants for these fluency variables (equal variances assumed). On the other hand, mean length of run, repair frequency and mid-clause repair frequency were not normally distributed. As such, Mann-Whitney U tests were used to determine whether there was a statistically significant difference between the mean lengths of run, repair frequencies and mid-clause repair frequencies of the two proficiency groups. These results are then discussed further in Sections 5.

To summarise the findings of the quantitative analysis and to answer research question 1, i.e. which quantitative measures of fluency distinguish between intermediate and advanced levels of proficiency of the Finnish National Certificates of Language Proficiency speaking test, it was found that articulation rate, mean length of run, pause duration, pause frequency, mid-clause pause frequency and mid-clause repair frequency distinguish between intermediate and advanced levels of proficiency of the Finnish National Certificates of Language Proficiency speaking test. My hypotheses were

confirmed, but while repair frequency distinguishes between the two proficiency levels, the result is not statistically significant.

Extreme case sampling. Based on the quantitative analysis, the fluency measure mean length of run produced the largest statistically significant difference between the two groups (47.5%). As such, this measure was selected as the basis of extreme case sampling for qualitative analysis. Three participants with the lowest and highest mean lengths of run were selected from both intermediate and advanced proficiency groups. This produced 12 speech samples for qualitative analysis. Their transcriptions were revisited, analysed qualitatively and annotated for pause and repair use to identify possible patterns in the speakers' use of pauses related to access and retrieval difficulty pauses, reformulation, and effective speech delivery, as per Nakatsuhara, Tavakoli & Awwad (2019), to see whether speakers at different levels of proficiency use pause and repair opportunities emphasising different aspects of L2 communication, or, in other words, to aid their speech in different ways, as well as what else examples and excerpts from L2 speakers reveal of pause and repair use. The following section, Section 4.2, presents the results of the qualitative analysis, answering research question 2.

4.2 Pause and repair use on the intermediate and advanced levels of YKI

This section introduces the results of the qualitative analysis, answering research question 2, i.e. how do pause and repair use distinguish between intermediate and advanced proficiency levels in the Finnish National Certificates of Language Proficiency speaking test? As for my initial hypotheses and in line with the findings of Nakatsuhara, Tavakoli & Awwad (2019), I expect to find that while both advanced and intermediate advanced and intermediate participants use pause and repair opportunities for access and retrieval difficulty, advanced participants are expected to be successful in their lexical and structural search more often than intermediate participants. Furthermore, I expect the advanced group to utilize fewer pause and repair opportunities for pauses and repairs related to reformulations (i.e. pauses and repairs related to reformulating ideas and utterances and rescuing ungrammatical utterances), and more pause and repair opportunities to improve the effectiveness of their speech (i.e. adding information, examples and justifications, offering opinions and comments, shifting topics and signalling dispreference for the topic). Conversely, I expect the intermediate group to use more pause and repair opportunities for pauses and repairs related to access and retrieval

difficulty and reformulations, and fewer pause and repair opportunities to improve the effectiveness of their speech. As such, I expect the two groups to compensate for different areas of their speech, namely that the intermediate group is more concerned with self-monitoring and correcting, and the advanced group with the effectiveness of their speech and their overall communicative prowess. Next, section 4.2.1 discusses results related to pause use, and Section 4.2.2 results related to repair use.

4.2.1 Pause use

This qualitative analysis is based on 421 pauses categorised based on their type or function in the participant’s speech. To answer Research question 2, the qualitative analysis of unfilled pauses led to several findings, presented as a table in Appendix 2. Below, Figures 2 and 3 visualise intermediate and advanced participants’ pause use by pause category and subcategory. The results presented in the figures below are discussed in more detail in sections 4.2.1.1 through 4.2.1.3.

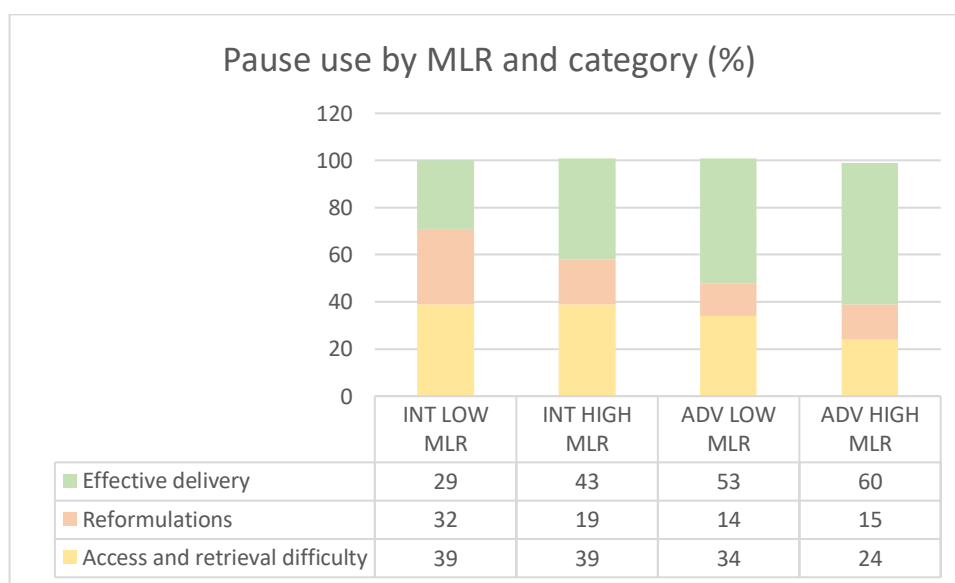


Figure 2 Participants’ pause use by MLR and pause category (%).

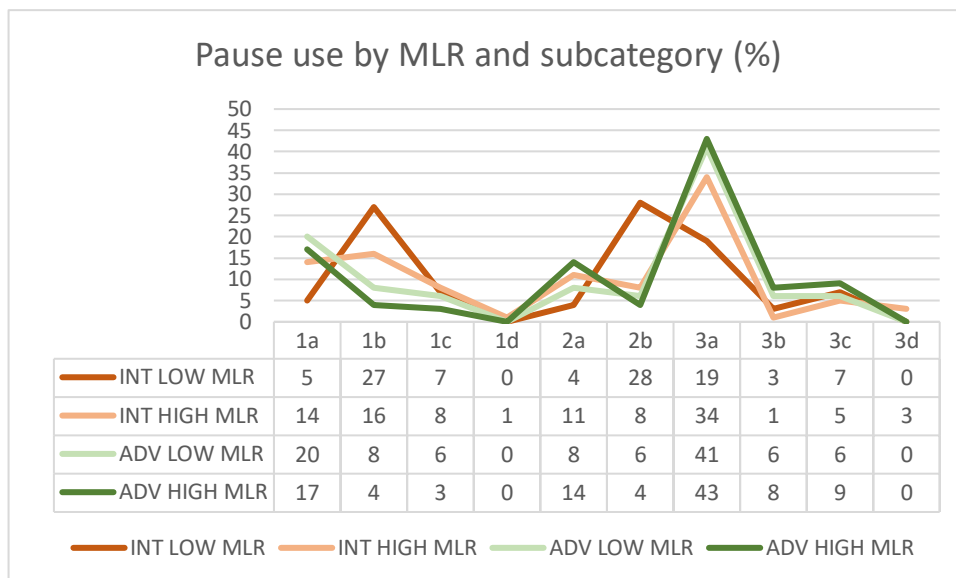


Figure 3 Participants’ pause use by MLR and pause subcategory (%).

Next, these findings are presented in the order of pauses related to access and retrieval difficulty, pauses related to reformulations, and pauses related to effective delivery, along with examples from participants’ samples.

4.2.1.1 Pauses related to access and retrieval difficulty

Based on the present analysis, intermediate participants use slightly more of their pause opportunities on pauses related to access and retrieval difficulty than advanced participants, as shown by Figure 2. 39% of the pauses produced by intermediate participants are related to access and retrieval difficulty regardless of whether the participants had had low or high mean lengths of run. In contrast, 34% of the pauses in the speech of advanced participants with low MLR and 24% of the pauses in the speech of advanced participants with high MLR are related to access and retrieval difficulty. Based on the present qualitative analysis, advanced participants are successful in their lexical and structural search more often than intermediate participants: as per Figure 3, 5% of all pauses in the speech of intermediate participants with low MLR and 14% of pauses in the speech of intermediate participants with high MLR are successful lexical and structural search pauses, i.e. pauses preceding more sophisticated language (pause use subcategory 1a), while advanced participants with low MLR used 20% and advanced participants with high MLR 17% of their pause opportunities on successful lexical and structural search.

As mentioned above, advanced participants have mid-clause pauses related to lexical/structural search, followed by more sophisticated language (1a) in their speech

than intermediate participants. However, that is not to say intermediate participants never succeed in lexical or structural search. In fact, even though mid-clause pauses related to successful lexical/structural search were the least common in the speech of intermediate participants with low mean lengths of run, (1) is a good example of a more sophisticated lexical unit preceded by pauses of this category in the speech of an intermediate participant with low MLR. In this example, the intermediate participant is searching for a word to describe the topic of the task, Finnish road safety, and the search results in the word *complicated*:

(1) ID 106042, intermediate participant with low MLR

```
:: uh finnish road safes [(2b_MC.5) uh (.)] [MC_repla_is:  
{1a_MC*hh*_1.6} is: (1a_MC1.9)] complicated ::
```

Still, mid-clause pauses related to successful lexical/structural search were more common in the speech of advanced participants. Example 2 below is from an advanced participant with a low MLR. Here, a longer pause of this category precedes the word *appraising* and a shorter one the word *integrity*. Here, a 1a pause is followed by a micropause of <300 ms, which, had it been longer, would have been categorised as a pause in the middle of / after producing sophisticated language (1c). While end-clause pauses cannot be categorised as pauses related to successful lexical/structural search according to the pause use typology by Nakatsuhara, Tavakoli & Awwad (2019, 26), the fact that there is a would-be 1c pause following directly after supports the first pause being categorised 1a despite its proximity to the preceding clause border.

(2) ID 68024, advanced participant with low MLR

```
:: and (1a_MC2.884) appraising (.) the candidates' (1a_MC.59) integrity  
is difficult of course ::
```

Finally, mid-clause pauses related to successful lexical/structural search sometimes also co-occurred with repairs, especially in the speech of advanced participants. Most often a word is repeated to mask lexical or structural search. This is more common in the speech of advanced participants than the speech of intermediate participants, and more common in the speech of participants with high mean length of run than participants with low mean length of run. In Example 3, the advanced participant with a high mean length of run has two mid-clause pauses related to successful lexical/structural search: a short pause followed by a hesitational filled pause *er* precedes the word *promote*, and a slightly longer

pause the word *council*. What is notable here is that the second mid-clause pause related to successful lexical/structural search is also part of a repetition, where the participant attempts to partially mask the search for the word *council* by repeating the word *various*. This is typical of advanced repair use.

(3) ID 70000, advanced participant with high MLR

```
:: it's very politically correct to try to [(1a_MC.5) er] promote  
multiculturality there are several different themes and projects going  
on in [MC_rep1_various (1a_MC.57) various councils] [...]
```

In contrast, Figure 3 also shows how mid-clause pauses for lexical/structural search resulting in erroneous utterances or in generic expressions (1b) are more common in the speech of intermediate than advanced participants: as per figure 3, on intermediate level, participants with low MLR used 27% and participants with high MLR 16% of their pause opportunities on unsuccessful lexical and structural search, while on the advanced level, participants used 8% (low MLR) and 4% (high MLR) of their pause opportunities on unsuccessful lexical and structural search. As for pauses occurring after lexical and structural search (1c), intermediate participants used slightly more of their pause opportunities on pauses related to 1c pauses occurring after lexical and structural search: intermediate participants with low MLR used 7%, intermediate participants with high MLR 8%, advanced participants with low MLR 6% and advanced participants with high MLR 3%.

Based on the present analysis, mid-clause pauses for lexical/structural search resulting in erroneous utterances or generic expressions are one contributing factor to the staccato cadence and halted speech characteristic of language learners: A typical example is (4), which is from an intermediate participant with a low MLR. In this example, the participant is likely looking for the expression *because there are so many children*, resulting in several long unfilled mid-clause pauses (and pause clusters) for lexical/structural search, first in search for the correct alternative of the *there is / there are* structure, and then for the expression *so many children*. The lexical and structural search finally results in the erroneous utterance *because there is: er because there is(h) so much child(hh)*. Notably, this example also includes the use of a repetition for access and retrieval, i.e. the participant attempts to buy processing time by repeating *because there is*.

(4) ID 105571, intermediate participant with low MLR

it's: quite safe [because {1b_MC*hh*_1.45} [MC_rep1b_ there is:
{1b_MC*pt*_2.867} er (MC1.048) because {1b_MC*hh*_1.624} there is(h)]
so much child(hh) ::

In contrast, advanced participants, especially those with high MLR, only rarely have access and retrieval pauses followed by erroneous or generic expressions, and any such are often unsuccessful attempts of searching for idiomatic expressions (such as in (5)). In this example, the participant has an unfilled-filled-unfilled pause cluster while searching for the idiomatic expression *on the other hand*, resulting in an unidiomatic expression *on the other side*. Other examples included mid-clause pauses being followed by erroneous expressions which the participant attempted to reformulate into better expressions directly after (e.g. in (6)). Based on the analysis, these kinds of reformulations are typical of intermediate participants with high MLR as well as advanced participants.

(5) ID 70415, advanced participant with high mean MLR

[...] [(1b_MC*hh*_768) um {*hh*_52}] | on the other side uh
multicultural teams [MC_ref3a_>can be< or they have been found to be]
[MC_rep3_very very] effective/ [...]

(6) ID 70000, advanced participant with high MLR

[...] [uh (1b_MC.4)] governmental sections but [uh
(2a_MC.33) [MC_ref2a_sect- sectors- (2a_MC.3) sections]] [...]

Furthermore, interesting pause use was also observed in the speech of advanced participants with low MLR. In (7), an advanced participant with a low MLR is using pauses for access and retrieval difficulty. The first pauses in this excerpt are clearly related to lexical and structural search: the participant is searching for words like kindergarten and (most likely) immigrant. In contrast, the pause before “*of different colour*” can also be interpreted as the speaker signalling that they are not quite sure how to describe someone being of different ethnicity in an appropriate way. As such, this pause facilitates effective speech delivery and was therefore categorised as a mid-clause pause before making evaluative comments and before expressing feelings (3b).

(7) ID 68431, advanced participant with low MLR

{3a_EC*h*_1}] um (1)] in (1a_MC.5)] the kindergarten where my children
went to (.) there were (MC_1b_MC.8) some [(MC_1b_MC.4) um {*pt*_1}]
people that had [(1b_MC.6) uh (0.6)) eh (0.6)] come to finland/
{1c_EC*hh*_6} and they were (3b_MC0.5) of different colour
[{1c_EC*h*_1}]

In their 2019 study, Nakatsuhara, Tavakoli & Awwad found that participants had pauses during or after producing sophisticated language (1c), reflecting the need to monitor language (Nakatsuhara, Tavakoli & Awwad 2019, 37-38). However, based on the present qualitative analysis, participants often paused after erroneous utterances and reformulations as well, which led to a reworking of this pause use subcategory to include pauses during or after unsuccessful searches and reformulations. As such, in the present study, this subcategory is titled *pauses in the middle of / after producing monitored sophisticated, erroneous or generic language, and after producing reformulations*. This amended category is discussed further in Section 5.

As for this subcategory, it was found that intermediate participants used slightly more of their pause opportunities on pauses during or after producing sophisticated, erroneous or generic language, or reformulations, than advanced participants: as per Figure 3, intermediate participants with low MLR used 7% of their pause opportunities on 1c pauses, intermediate participants with high MLR 8%, advanced participants with low MLR 6% and advanced participants with high MLR 3%. While the differences are minor, based on the analysis, intermediate participants paused more frequently after producing erroneous, monitored language and advanced participants after producing sophisticated language. (8) is a good example of pauses in the middle of/ after producing monitored sophisticated, erroneous or generic language, and after producing reformulations (1c) in the speech of intermediate participants. In this example, an intermediate participant with a low MLR has two pauses of this category. The first pause opportunity is likely used to search for an utterance like *alcohol* or *alcoholic beverages*, but the participant produces **alcoholic* after a mid-clause pause for lexical/structural search resulting in erroneous utterances or in generic expressions. This erroneous lexical item is followed by an evaluative or reflective .5 second pause after producing erroneous or generic language. The second pause is a similar pause monitoring or evaluating the erroneous lexical item *happenings*. Based on the present qualitative analysis, pauses evaluating the results of unsuccessful lexical or structural search are typical of the intermediate level, particularly the intermediate participants with low MLR, after producing a monitored erroneous utterance.

(8) ID 106052, intermediate participant with low MLR

:: nowadays we are offered (1b_MC.552) alcoholic (1c_MC.5) in many (1b_MC.5) happenings {1c_MC*hh*_1.091} [...]

In contrast, (9) is an example of how pauses are used for evaluating the results of lexical and structural search on the advanced level. While the utterance *answer* is not erroneous, it is generic in contrast to the surrounding language (and thus categorised as a pause preceding erroneous language, or 1b), and the participant was likely searching for a more appropriate word, like the synonym *reply*.

(9) ID 68024, advanced participant with low MLR

```
:: and they have to (1b_MC.5) answer (1c_MC.5) immediately without notes  
[...]
```

Finally, the last subcategory of pauses related to access and retrieval difficulty, pauses to recall items from long-term memory (1d), was found to be the rarest of access and retrieval difficulty pauses, as well as the most difficult category to analyse, as identifying whether or not a particular item was in fact recalled from long-term memory proved difficult. As such, this category was limited to search for examples and ideas as opposed to search for linguistic units, effectively avoiding the issue. The only examples of this were found in the samples of an intermediate participant with high MLR. Neither intermediate participants with low mean lengths of run or advanced participants with low or high mean lengths of run appeared to use pauses for this purpose. As such, one example of this pause subcategory is discussed. In (10), an intermediate participant with a high MLR appears to use a pause to recall an item from long-term memory and recovers an example, *India*, from their long-term memory.

(10) ID 106561, intermediate participant with high MLR

```
:: it's not for free like in (1d_MC.6) India where you (.) practically  
have to only give your {1a_MC*hh*_1} personal id to get a driver's  
licence for you ::
```

4.2.1.2 Pauses related to reformulation and rescue

Based on the present analysis, intermediate participants use more of their pause opportunities on pauses related to reformulations than advanced participants: as per Figure 2, intermediate participants with low mean lengths of run used 32% and intermediate participants with high mean lengths of run 19% of their pause opportunities on pauses related to reformulations, while advanced participants with low mean length of run used 14% and advanced participants with high mean lengths of run 15% of their pause opportunities on pauses related to reformulations.

Furthermore, as shown in Figure 3, it was found that intermediate participants with high MLR (11%) as well as advanced participants with low MLR (8%) and high MLR (14%) used more of their reformulation pauses on mid/end-clause pauses occurring during / before reformulating ideas and utterances, and making self-corrections (2a), as opposed to intermediate participants with low MLR, who used only 4% of their pauses on these 2a pauses. In contrast, intermediate participants with low MLR used most of their reformulation pauses (28% of their total pauses) on mid-clause pauses in the middle of ungrammatical structures in the attempt of restructuring sentences (2b), whereas intermediate participants with high MLR used only 8%, advanced participants with low MLR 6%, and advanced participants with high MLR as few as 4%. In fact, many pauses related to rescues were found in the speech of intermediate participants with low mean lengths of run, while intermediate participants with high mean lengths of run were closer to the advanced groups with their use of reformulation (2a) and rescue (2b) pauses.

As stated above, intermediate participants with high mean lengths of run approach advanced participants in their use of reformulation pauses, or mid/end-clause pauses occurring during / before reformulating ideas and utterances, and making self-corrections. In Example 11, the intermediate participant with a high mean length of run has three mid/end-clause pauses occurring during / before reformulating ideas and utterances, and making self-corrections, all occurring within repairs. The first pauses of this category are part of an idea being reformulated, which based on the present analysis is typical of advanced participants' reformulation pause use, while the third reformulation pause occurs while a determiner is being reformulated for accuracy.

(11) ID 106204, intermediate participant with high MLR

```
:: and where the laws are not restricting [rep2_certain {2a_MC*h*_.608}
uh certain (2a_MC.4) uh] let's say where the laws are not so restrictive]
[MC_ref2a_for the (2a_MC.4) for their- their] [{1a_MC*hh*_.6} uh]
production ::
```

As for mid-clause pauses in the middle of ungrammatical structures in the attempt of restructuring sentences or rescuing them (2b), this subcategory mostly applies to the intermediate participants, where the speech of intermediate participants with low MLR had the majority of pauses of this category. In contrast, mid-clause pauses in the middle of ungrammatical structures in the attempt of restructuring sentences or rescuing them are in the speech of advanced participants. As such, example 12 is a typical example of rescue pause use for intermediate participants with low MLR. In this example, the

participant has three rescue pauses as part of a short description of road safety being affected by there being an insufficient police presence on the road.

(12) ID 105571, intermediate participant with low MLR

:: and the road is [{2b_MC*h*_1.675} so [(2b_MC3.177) uh] there is (2b_MC.91) not: enough /police ::

In addition, based on the present analysis, rescue pauses were used by intermediate participants in mid-clause and even mid-phrase positions in grammatical and ungrammatical utterances in ways that may suggest that the proceduralisation of the speech rhythm and/or linguistic chunks of the target L2, such as phrasal verbs, is incomplete. For example, in examples 13 and 14 below, intermediate participants with low MLR have multiple rescue pauses.

(13) ID 105571, intermediate participant with low MLR

:: so (.) it(h) also problem (2b_MC1) with (2b_MC.5) drivers :: {3a_EC*h*_1.3} many think (2b_MC.461) about (2b_MC.81) it so (2b_MC.7) I'm drink (.) just only (2b_MC.4) few peers (2b_MC.5) well I can drive :: {3a_EC*hh*_.9} this is {2b_MC*hh*_2.833} the road (2c_MC.824) that I drive every day ::

(14) ID 105466, intermediate participant with low MLR

[...] create our cars (2b_MC2.3) to (2b_MC.6) keep (2b_MC1.5) in road [(3a_EC1) or: (.7)] [EC_rep3a_maybe we can {3a_*pt*_1.7} maybe we can] (1b_MC4.5) [MC_rep1b_addition- (2b_MC.4) maybe we can addition]{2b_MC*hh*_.5} police (2b_MC.4) in police department (3a_MC.6) so we have more people (2b_MC.4) to watching {2b_MC*hh*_.5} that people (2b_MC1.2) drive safely ::

Finally, it was found that intermediate participants used more of their reformulation pauses on restructuring their utterances, self-correcting grammar and pronunciation, and reformulating utterances, whereas advanced participants focused on reformulating ideas as opposed to correcting minor errors. This is discussed further in Section 5. Next, we move on to pauses related to effective delivery.

4.2.1.3 Pauses related to effective delivery

Advanced participants use more of their pause opportunities on pauses related to effective delivery than intermediate participants. The number of effective delivery pauses increases with mean length of run: intermediate participants with low MLR used 29%, intermediate participants with high MLR 43%, advanced participants with low MLR 53% and

advanced participants with high MLR 60% of their pause opportunities on pauses related to effective delivery. This confirms my initial hypothesis.

Most pauses related to effective delivery were pauses before adding more information, examples, and justifications (3a), and these pauses were more common in the speech of advanced participants (low MLR: 41%, high MLR: 43%) than intermediate participants (low MLR: 19%, high MLR: 34% of total pauses). In example 15 below, an intermediate participant with a high mean length of run has two examples of end-clause pauses before adding more information, examples, and justifications. The first pause before adding more information, examples, and justifications co-occurs with a repair and precedes an example of why Finnish roads are safe, while the second offers additional information relating to this example.

(15) ID 106561, intermediate participant with high mean length of run

```
(3a_EC.5) [EC_rep3a we have <we have] also clear rules which we obey>  
:: {3a_EC*hh*_.5} finnish people are known for obeying rules very  
properly ::
```

In contrast, advanced participants often had pauses of this category in clause-medial positions. In example 16, an advanced participant has a pause before adding more information, examples, and justifications in mid-clause position between two repairs. In this example, the participant gives an example of how staying abroad for extended periods of time benefits a person, along with a personal justification.

(16) ID 70000, advanced participant with high mean length of run

```
and sometimes [[MC_rep3_even (3a_MC*h*_.5) even [MC_rep3_make make]] you  
appreciate your own culture even more I noticed myself when I lived  
abroad that I had become more of a patriot there than I ever was in  
Finland
```

On the other hand, as per Figure 3, mid-clause pauses before making evaluative comments and before expressing feelings (3b), are rare in the speech of intermediate participants (low MLR: 3%, high MLR: 1%) but more common in advanced participants' speech (low MLR: 6%, high MLR 8%). Example 17 below is taken from an advanced participant with a high mean length of run. In this example, the participant is making an evaluative comment, *definitely an advantage*. A mid-clause pause before making evaluative comments and before expressing feelings precedes this assessment.

(17) ID 70415, advanced participant with high mean length of run

:: err I believe it's (3b_MC.472) definitely an advantage to work and live in a multicultural environment ::

End-clause pauses before a topic shift (3c) were the second most common pause type in the speech of both intermediate (low MLR: 7%, high MLR: 5%) and advanced participants (low MLR: 6%, high MLR 9%). In example 3c-1, an intermediate participant with a low mean length of run shifts topics in a way typical of intermediate participants: a long 3c pause in end-clause position followed by the co-ordinator *and*.

(18) ID 105571, intermediate participant with low MLR

so (.) they don't drive overspeed :: {3c_EC*hh*_1.3} and in finland(h) (.) we have a problem with alcohol ::

In contrast to Example 18, advanced participants, regardless of their mean length of run, most often shifted topics by explicitly introducing the next topic (examples 19 and 20).

(19) ID 68024, advanced participant with low MLR

::[{3c_EC*sf*_1.2} err] the topics that I'm interested in [...]

(20) ID 70006, advanced participant with high mean MLR

:: [(3c_EC.7) uh] the most interesting topics from my point of view [...]

Finally, turn-initial pauses before dispreferred responses (3d), were rare, the only occurrences in one intermediate participant's speech (1%). In other words, out of twelve samples analysed, only one participant had turn-initial pauses before dispreferred responses in their speech, their three 3d pauses amounting to 1% of the total pauses analysed from intermediate participants. In Example 21, the intermediate participant with high MLR has three turn-initial pauses before dispreferred responses in their speech.

(21) ID 104656, intermediate participant with high mean length of run

question is if I trust finnish politicians [{3d_EC*hh*_1 um {h*_.5}][EC_rep3_it's: it's] definitely not an easy question [{3d_EC*h*_1.3} uh] 'cause there are so many politicians and [uh] most of them(hh) I don't know {3d_EC*h*_1} or actually I don't know any one of them personally (.)

In example 21 above, the last pause could also be categorised a reformulative 2a pause but has been categorised as a 3d pause due the participants' clear use of pausing to plan their response, hedge their answer and express dispreference. The next section discusses the results related to repair use.

4.2.2 Repairs

This qualitative analysis is based on 143 repairs categorised based on their inferred function in the L2 learner’s speech. A qualitative analysis and categorisation of repairs into access and retrieval repairs, repairs related to reformulations and effective delivery repairs has led to the following findings, visualised in Figure 4 and Figure 5. A detailed analyses of repair use, along with examples, can be found in sections 4.2.2.1 to 4.2.2.3.

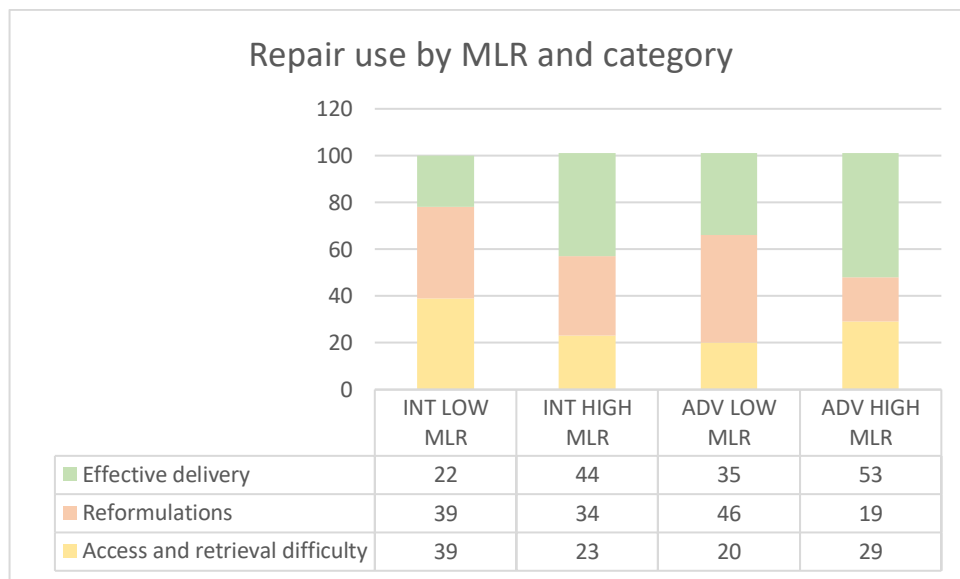


Figure 4 Participants’ repair use by MLR and repair use category.

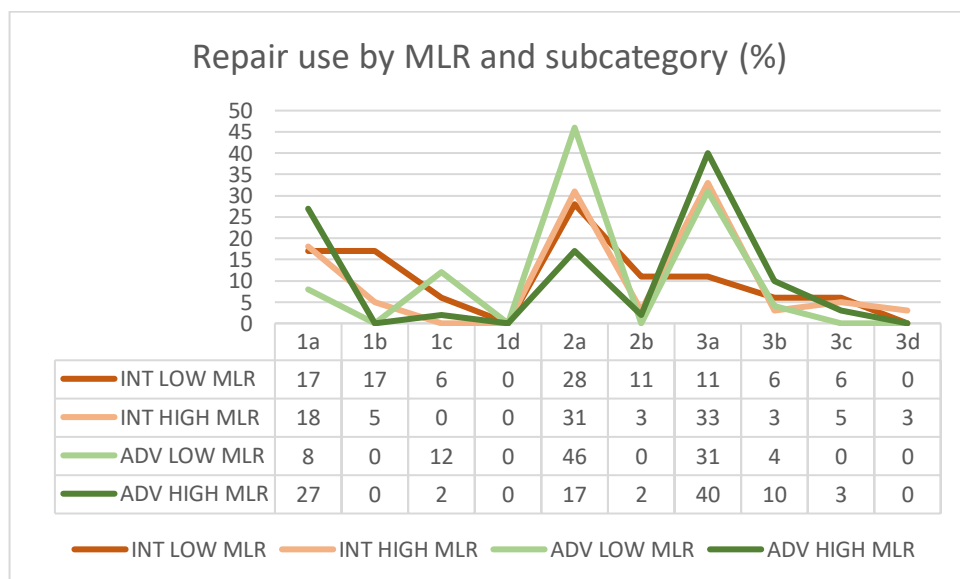


Figure 5 Participants’ repair use by MLR and repair use subcategory (%).

Next, the results are presented, along with examples from participants’ samples.

4.2.2.1 Repairs related to access and retrieval difficulty

As per Figure 4, intermediate participants with low MLR used 39% and intermediate participants with high MLR 23% of their repairs on repairs related to access and retrieval, while advanced participants had slightly fewer repairs of this category (low MLR: 20%, high MLR 29%). Based on the present qualitative analysis, repairs related to access and retrieval difficulty often either co-occur with pauses related to access and retrieval difficulty or are used to replace a pause of the same subcategory. Based on the present analysis, such repairs were often repetitions of lexical and structural items or parts thereof.

As per Figure 5, repairs occurring in mid-clause position in search for lexical or structural items and followed by more sophisticated language (1a), co-occurred or replaced pauses of the same category by both intermediate (low MLR: 17%, high MLR: 18%), and advanced participants (low MLR: 8%, high MLR 27%). Participants with high mean lengths of run were more effective in using repairs occurring in mid-clause position in successful search for lexical or structural items to replace or shorten the duration of pauses of the same category. In example 22, the intermediate participant with a high mean length of run repeats the determiner *the* while searching for the word *well-being*.

(22) ID 106204, intermediate participant with high MLR
...[MC_rep1a_the (MC_1a.5) the[(r) (.) er the]] well-being (1c_MC.456)
'cross the globe ...

Similarly, repairs occurring in mid-clause position in search for lexical or structural items and resulting in erroneous utterances or generic expressions (1b), (INT low MLR: 17%, INT high MLR: 5%, as per Figure 5), also co-occurred with and replaced pauses of the same category. Interestingly, these repairs were only present in the samples of intermediate participants. In Example 23, an intermediate participant with a low mean length of run uses repetitions along with several pauses of the same category to allow for processing time during a search for the phrasal verb *care about*, and directly after, the noun *traffic*.

(23) ID 106052, intermediate participant with MLR
:: and then we drive fast (3a_EC1.4) and (1b_MC.5) the (1b_MC1) animals
[MC_rep1b_don't (1b_MC1.2) uh (1b_MC2.4) don't (1b_MC2.7) don't]
(1b_MC.5) care about [MC_rep1_the- the] traffic ::

Again, repairs occurring in the middle of / after producing monitored (sophisticated, erroneous or generic) language (1c), often either co-occur with or replace 1c pauses. As

per Figure 5, repairs of this category in the speech of intermediate participants with low MLR amount to 6%, in the speech of advanced participants with low MLR to 12%, and finally, in the speech of advanced participants with high MLR to 2%. The speech of intermediate participants with high MLR had no repairs of this category. In Example 24 on page 52, the advanced participant with a low MLR repeats parts of a less-frequent word *advertising*, the repetition co-occurring with a pause of the same category.

(24) ID 68024, advanced participant with low MLR

[...] but (.) of course there are also laws concerning (1a_MC.4) the
[MC_rep1c_ adver- adver(1c_MC.4)tising]

No repairs to recall items from long-term memory (1d) were identified in the participants' speech. Given their absence from the present analysis, we shall move on to repairs related to reformulations next.

4.2.2.2 Repairs related to reformulation and rescue

As per Figure 4, advanced participants with low MLR had the most reformulation repairs. At 46%, they used more repair opportunities on repairs related to reformulations than intermediate participants, who used over a third of their repair opportunities on reformulations, with repairs in the speech of participants with low MLR amounting to 39%, and repairs in the speech of participants with high MLR to 34%. Finally, advanced participants with high MLR had the fewest reformulation repairs at 19%. Based on the present qualitative analysis, repairs related to reformulations mostly co-occur with or replace pauses related to reformulations. While participants' reformulation repairs mostly consist of repairs occurring in mid/end-clause positions during / before reformulating ideas and utterances, and making self-corrections (2a, reformulation repairs) (INT, low MLR: 28%, INT, high MLR: 31%, ADV, low MLR: 46%, ADV, high MLR: 17%) as opposed to repairs occurring in mid-clause position in the middle of ungrammatical structures in the attempt of restructuring sentences (2b, rescue repairs) (INT, low MLR: 11%, INT, high MLR: 3%, ADV, low MLR: 0%, ADV, high MLR: 2%, as per Figure 5), intermediate participants have more repairs occurring in mid-clause position in the middle of ungrammatical structures in the attempt of restructuring sentences than advanced participants. Furthermore, based on the present analysis, intermediate participants' reformulation repairs tend to co-occur with and replace rescue pauses as well as reformulation pauses, while advanced participants' reformulation repairs co-occur with

and replace only reformulation pauses, and more specifically reformulation pauses where ideas are being reformulated for greater accuracy or for cultural or strategic compliance. In other words, intermediate participants used reformulation and rescue repairs most often for rescuing ungrammatical utterances and self-correction, and advanced participants for reformulating ideas for greater accuracy and cultural and strategic sophistication. An example of intermediate reformulation repair use is in Example 25, where an intermediate participant with a high mean length of run has two repairs preceding a pause occurring in mid-clause position in the middle of ungrammatical structures in the attempt of restructuring sentences: they are attempting to rescue an utterance while searching for the best way to express what they want to say:

(25) ID 106561, intermediate participant with high mean length of run
 [EC_ref2b_we are- have kind of used] [MC_rep2a_al- also]
 {2b_MC*hh*_.648} only the (1a_MC.472) automatic control of the traffic
 and that's not good because it only takes the speeding and not
 motorcycles and stuff like that ::

In contrast, in example 26, an advanced participant with a high mean length of run has two reformulation repairs, one reformulating for grammar (in-into) and one reformulating for better accuracy (multicultural- cross-cultural). It should be noted that repairing ungrammatical utterances was atypical for advanced participants, who mostly repaired and reformulated ideas instead of language.

(26) ID 70415, advanced participant with high mean length of run
 :: [EC_rep3_they- (.) the society and our educational system and our
 companies] (.) should (.) put more effort [MC_rep2a_in {*pt*2a_MC.5} in-
 into] learning and teaching about [MC_rep2a_multicultural- cross-
 cultural] competences and behaviour at work or in studies ::

4.2.2.3 Repairs related to effective delivery

Finally, as per Figure 4, advanced participants had more repairs related to effective delivery than intermediate participants. However, intermediate participants with low MLR had the fewest effective delivery repairs (22%), followed by advanced participants with low MLR (35%), whereas participants with high MLR used more repair opportunities on effective delivery repairs (intermediate, high MLR: 44%, advanced, high MLR: 53%). At the threshold of intermediate level with low mean lengths of run and advanced level with high mean lengths of run, advanced participants with low mean lengths have an increase in reformulation repairs at the expense of access and retrieval

difficulty repairs and effective delivery repairs. Based on the present qualitative analysis, repairs related to effective delivery often co-occur with or replace pauses related to effective delivery. According to the analysis, the most common effective delivery repairs were repairs occurring before adding more information, examples and justifications (3a), (INT, low MLR: 11%, INT, high MLR: 33%, ADV, low MLR: 31%, ADV, high MLR: 40%). Intermediate participants mostly had repairs of this category on clause borders, and the most common repair was repeating the co-ordinator *and* at clause borders. While this behaviour was more common of intermediate participants, it is exemplified here in Example 27 by an advanced participant with a high mean length of run. Here, a repair to include additional information about their multicultural circumstances.

(27) ID 70415, advanced participant with high MLR

[...] and nowadays I also have lots of colleagues who come from different countries :: [EC_rep3a_and- (.) and] maybe half of my students come from other countries than Finland

In contrast to intermediate participants, advanced participants mostly have repairs occurring before adding more information, examples and justifications in mid-clause positions, often offering more information and aiming for more accuracy, as exhibited by both Examples 28 and 29 (*we (I) in my age group (.) we (...); <can be> or they have been found to be (...)*). These repairs appear to be reformulation repairs at first glance, but instead of reformulating an utterance or idea, they instead function to offer volunteer information to the interlocutor. Furthermore, Example 29 also has a repair occurring in mid-clause position before making evaluative comments or expressing feelings (3b) (INT, low MLR: 11%, INT, high MLR: 33%, ADV, low MLR: 31%, ADV, high MLR: 40%). In Example 29, the intensifier *very* is repeated as part of expressing an opinion or making a comment:

(28) ID 68431, advanced participant with low MLR

[EC_ref3a_we (3a_EC*h*_1) as in my age group (.) we] haven't had that many opportunities to get to know people from other countries [...]

(29) ID 70415, advanced participant with high MLR

on the other side uh multicultural teams [MC_ref3a_<can be> or they have been found to be] [MC_rep3b_very very] effective/

Furthermore, end-clause repairs related to topic shifts (3c) were more common in the speech of intermediate participants (low MLR: 6%, high MLR: 5%) than advanced

participants (low MLR 0%, high MLR 3%). In the speech of intermediate participants, these often took the form of repeating the co-ordinator *and*:

(30) ID 105571, intermediate participant with low MLR

```
:: [EC_rep3c_and-(3c_EC1.6) and and [EC(.5) oh(.4) well*hh*_.5]]  
on the road {2b_MC*hh*_1.15} out of the /town [...]
```

In contrast, turn-initial repairs before dispreferred responses (3d), only occurred in the speech of one intermediate participant with high MLR (3% of total repairs occurring in the speech of intermediate participants with high MLR).

Finally, while the lines between pauses and repairs were often blurred, this was especially true with effective delivery repairs. In Example 31 below, an advanced participant with a high mean length of run uses many repairs in positions where most other participants would have unfilled pauses:

(31) ID 70000, advanced participant with high mean length of run

```
:: there are benefits and challenges [MC_repla_of- of- of] multicultural  
environments if I first think of the challenges[EC_rep3a_there- there]  
are of course (.) [MC_repla_many- (.) many]] cultural differences  
depending on which cultures but if you think about the Nordic cultures  
versus [MC_repla_the- the uh (1a_MC.5) Muslim Arabic cultures for  
instance] there are great differences especially between the role of the  
male and the female [(3a_EC*hh*_.8) and uh (.3)] those naturally stem  
[rep2_from very often from very] different religious backgrounds and [...]
```

This kind of behaviour was typical of intermediate and advanced participants with high mean lengths of and contributed to their high mean lengths of run. Next, findings related to research question 3 are presented below.

4.3 The interdependence of pause and repair use

To answer Research question 3, i.e. whether pause and repair use is interdependent, the present qualitative analysis of unfilled pauses and repairs has resulted in the following findings:

- 1) Pauses and repairs of the same category often co-occur: participants often have pauses and repairs of the same category co-occur on both clause borders and in mid-clause positions.

- 2) Pauses are often replaced or substituted by repairs, blurring the lines between pauses and repairs (pause-repair replacement). This behaviour is typical of both advanced and intermediate participants, particularly those with higher mean lengths of run, i.e. more fluent participants, and is especially common with pauses and repairs related to access and retrieval difficulty and pauses and repairs related to effective delivery. As per Example 31 above, the highly fluent participant (advanced level, high mean length of run) employed repetitions in positions where most participants would have unfilled pauses (in bold). In addition, this participant's sample exhibits pause-repair co-occurrence.
- 3) As pauses and repairs related to effective delivery are one-dimensional categories in the sense that all her subcategories are typical of more advanced and more fluent L2 speakers, they could be correlated for the purposes of triangulation. A strong positive correlation was found between pauses and repairs related to effective delivery ($r_s = .69649$, $df = 11$, $p < .05$): participants who had proportionally more pauses related to effective delivery also had proportionally more repairs related to effective delivery.

In contrast, pauses and repairs related to access and retrieval difficulty have subcategories that are typical of both advanced and intermediate pause and repair use (1a and 1b, respectively). The same is true for pauses and repairs related to reformulations: intermediate participants have more rescue pauses (2b) than advanced participants, but both intermediate and advanced participants have reformulation pauses (2a) – what participants reformulate distinguishes between the two groups as it was found that advanced participants reformulate ideas whereas intermediate participants focus on self-correction by reformulating ungrammatical utterances.

In sum, pauses and repairs of the same category often co-occur, and repairs are used by especially advanced participants to avoid the need for pausing. Furthermore, pauses and repairs of the same category are used to achieve the same end: some speakers repair where others pause, and vice versa. Finally, participants who had proportionally more pauses related to effective delivery also had proportionally more repairs related to effective delivery ($r_s = .69649$, $df = 11$, $p < .05$). Next, these findings are discussed in more detail in Section 5, Discussion.

5 Discussion

Section 5 revisits the research questions and hypotheses in the order presented in this study, further discussing, interpreting and explaining the results to research question 1 in Section 5.1, research question 2 in Section 5.2, and research question 3 in 5.3, relating the results to earlier studies where possible, with special attention to the significance and generalisability of the results of the present study. Section 5.4 discusses the present study's implications for the YKI speaking scale, while the study is evaluated in 5.5. Finally, Section 5.6 sets out to give recommendations for further research.

5.1 Quantitative fluency measures distinguishing between the intermediate and advanced levels of YKI

The first objective of this study was to study which quantitative measures of fluency distinguish between intermediate and advanced levels of the Finnish National Certificates of Language Proficiency in English. Based on the quantitative analysis of 60 speech samples and to answer Research Question 1, it was found that i) advanced participants had on average 20.5% higher articulation rates than intermediate participants, confirming my initial hypothesis; ii) advanced participants had on average 47.5% longer mean lengths of run than intermediate participants, confirming my initial hypothesis; iii) advanced participants had on average 25.8% shorter total pause durations than intermediate participants, confirming my initial hypothesis; iv) advanced participants had 13.4% fewer pauses per minute than intermediate participants, and 14.3% fewer mid-clause pauses per minute than intermediate participants, confirming both hypotheses; and v) advanced participants had 29.4% more mid-clause repairs per minute than intermediate participants, confirming my hypothesis. As such, the quantitative fluency measures distinguishing between intermediate and advanced levels of the Finnish National Certificate of Language Proficiency in English are articulation rate, mean length of run, total unfilled pause duration, pause frequency, mid-clause pause frequency and repair frequency. Regarding research question 1, all hypotheses but one are confirmed. However, while advanced participants had higher repair frequency than intermediate participants, the difference was not statistically significant: here, the null hypothesis stands. As mean length of run produced the largest difference between the two groups, it was chosen as the extreme sampling measure for the qualitative part of this study.

As introduced in Section 3, speed fluency measures have been found to be good predictors of proficiency: higher speech rates and higher articulation rates are thought to be a result of more automatised and proceduralised top-down and bottom-up processes underlying L2 processing (e.g. Kormos 2006). While speech rate was not used as a quantitative measure in this study due to the measure overlapping with pause and repair phenomena, speech rate remains a good standalone measure that correlates with other temporal fluency measures and predicts proficiency.

In addition, mean length of run is one of the most oft-cited fluency measures to distinguish between proficiency levels and to predict proficiency. However, mean length of run is a measure that overlaps with other fluency measures, being a comprehensive measure combining elements of speed, breakdown and repair fluency depending on which temporal features are treated as interruptions and which included in fluent runs. Still, the present study also revealed that mean length of run was a somewhat reliable measure in grouping participants by their language assessment (YKI levels 3 and 4 for intermediate level and 5 and 6 for advanced level): intermediate participants with low mean lengths of run had been assessed at proficiency level 3 in their speaking skills, while intermediate participants with high mean lengths of run had been assessed at proficiency level 4 in their speaking skills. This was also true for advanced participants with low mean lengths of run: all three had been assessed at level 5 in their speaking skills. However, mean length of run did not as accurately distinguish between advanced participants, i.e. proficiency levels 5 and 6 (Appendix 2).

Furthermore, unfilled pause frequency, particularly mid-clause unfilled pause frequency, are commonly found to distinguish between different levels of proficiency, However, as opposed to De Jong 2016, the total duration of unfilled pauses was also found to distinguish between different proficiency levels (De Jong 2016, 206). As expected, they also distinguished between intermediate and advanced levels of proficiency in the current sample. On the other hand, while repair frequency did not produce a statistically significant difference between the two levels of proficiency as expected, mid-clause repair frequency did: interestingly, advanced participants had more mid-clause repairs in their speech than intermediate participants even though pauses and repairs have been thought to subtract from fluency. In contrast, Nakatsuhara, Tavakoli & Awwad (2019) found that repairs may in fact be related to aiming towards accuracy, something that is characteristic of higher levels of fluency. This is in line with the findings

of the present study and is echoed by other studies examining repairs (e.g. Olkkonen 2017, Peltonen 2017).

5.2 Pause and repair use on the intermediate and advanced levels of YKI

The second objective of this study was to study how pause and repair use distinguish between intermediate and advanced levels of the Finnish National Certificates of Language Proficiency speaking test. The categories used in this study are based on a pause use typology proposed by Nakatsuhara, Tavakoli & Awwad (2019, 26), but the typology had to be amended based on the present qualitative analysis. First, as for pauses related to access and retrieval, it was found that participants, regardless of their proficiency level, paused not only after having produced sophisticated language (1a) but also after producing erroneous or generic language (1b), and sometimes after reformulations (2a). As such, subcategory 1c, or pauses in the middle of / after producing sophisticated, generic or erroneous, or reformulated language, was amended to reflect this finding.

As for pauses related to reformulations, it was found that pauses related to rescuing, i.e. mid-clause pauses in the middle of ungrammatical structures in the attempt of restructuring sentences, also occurred within grammatical structures in learner speech, in positions where there was no evidence of lexical or structural search or aiming towards efficient delivery, suggesting that these pauses may not have been pauses related to rescuing (2b), but rather pauses related to incomplete proceduralisation of e.g. speech rhythm, reduced forms, or syntax, or linguistic chunks. These pauses would occur mid-clause and even mid-phrase, such as within phrasal verbs. In sum, there is evidence of a fourth category of pause and repair use, but given the limited scope of the present study, it was not operationalised as a separate pause use category. Instead, pauses arousing such suspicions were categorised as 2b “rescue” pauses. However, in future studies, qualitative analysis of pause and repair use could include a fourth category: mid-clause pauses resulting from incomplete proceduralisation of speech rhythm and/or linguistic chunks. However, as suprasegmentals are beyond the scope of the current study, subcategory 2b was simply amended to include both grammatical and ungrammatical structures. As such, in the current study, 2b pauses contain pauses likely related to incomplete proceduralisation of speech rhythm or linguistic chunks. For the sake of clarity, the amended typology is presented on page 62 (amendments in italics).

Pause use typology (Nakatsuhara, Tavakoli & Awwad 2019, 26), amended with the findings of the present study:

- 1) **Pauses related to access and retrieval**, containing instances of
 - a) mid-clause pauses for lexical/structural search, followed by more sophisticated language (language structure)
 - b) mid-clause pauses for lexical/structural search resulting in erroneous utterances or in generic expressions (language structure)
 - c) *pauses in the middle of / after producing sophisticated, generic or erroneous, or reformulated language*
 - d) pauses to recall items from long-term memory (memory)
- 2) **Pauses related to reformulations**, containing instances of
 - a) Mid/end-clause pauses occurring during / before reformulating ideas and utterances, and making self-corrections (reformulations)
 - b) *Mid-clause pauses in the middle of ungrammatical or grammatical structures in the attempt of restructuring or monitoring sentences, as well as mid-clause pauses related to incomplete proceduralisation of suprasegmentals (rescuing)*
- 3) **Pauses related to effective speech delivery**, containing instances of
 - a) Pauses before adding more information, examples and justifications (topic development)
 - b) Mid-clause pauses before making evaluative comments and before expressing feelings (especially after an intensifier) (Attracting listeners' attention)
 - c) End-clause pauses before topic shift (Topic shift)
 - d) Turn-initial pauses before dispreferred responses (Dispreference)

Pauses. Based on the qualitative analysis of 12 extreme cases, and to answer Research Question 2, it was found that while intermediate participants had only slightly more access and retrieval difficulty pauses in their speech than advanced participants, advanced participants were successful in their lexical and structural search more often than the intermediate participants. This is in line with the findings of Nakatsuhara, Tavakoli & Awwad (2019, 37-38). Advanced participants with high mean lengths of run also had the lowest frequency of access and retrieval pauses, suggesting that the number of access and retrieval pauses may decrease as fluency and proficiency increase. However, further research is required given the small sample size. Furthermore, some pauses categorised as relating to access and retrieval difficulty may not actually signal difficulty in accessing lexis or grammar: sometimes more fluent participants would also pause before and after culturally or socially difficult concepts, signalling their hesitation in using loaded expressions. For example, in (7) on page 41, an advanced participant with a low mean length of run was interpreted to signal that by pausing before and after a sensitive concept, they are not quite sure how to appropriately or acceptably describe someone being of

different ethnicity.) Pausing before and after places the spoken equivalent of quotes around sensitive concepts and/or loaded expressions, exhibiting cultural and strategic competence and flexibility instead of lack of proficiency. This demonstrates that not all mid-clause pauses signal lack of proficiency. Pauses related to hedging for social or cultural reasons were typical of advanced participants, who exhibited ability to hold the attention of their (imaginary) interlocutor even in a test situation. This also shows that in typologies such as these, an element of inference and interpretation is ever-present: without having access to the participant's thoughts on what they did at the given moment, full certainty is impossible.

Intermediate participants also used more of their pause opportunities on pauses related to reformulations (category 2) than advanced participants. However, there was a difference in how the pauses were distributed between reformulation (2a) and rescue pauses (2b): Intermediate participants used more of their reformulation pauses on rescuing (2b), and on the other hand, reformulating ungrammatical utterances (2a). Surprisingly, it was found that advanced participants in fact had more reformulation pauses (2a) than intermediate participants: however, advanced participants focused on reformulating ideas instead of minor grammatical inaccuracies. While Nakatsuhara, Tavakoli & Awwad (2019) did not explicitly report the frequency of pauses in different subcategories, the present findings are in line with theirs.

On the other hand, advanced participants had more pauses related to effective delivery than intermediate participants: advanced participants were able to use more of their pause opportunities for more effective delivery as opposed to lexical and structural search or reformulations. This feature increased as fluency increased: effective delivery pauses were least common in the speech of intermediate participants with low mean lengths of run and most common in the speech of advanced participants with high mean lengths of run. As for the subcategories of pauses related to effective speech delivery, pauses related to topic development (3a) and justifying and evaluating points in discussion (3b) increased with fluency. Advanced participants used more descriptive language to describe feelings, and more quantifiers than intermediate participants, and offered their opinion more explicitly using phrases like *I think* and *definitely*. As for pauses relating to topic shifts, pauses related to topic shifts (subcategory 3b) and surrounding language revealed differences in how intermediate and advanced participants shifted between topics: in the speech of even highly fluent intermediate participants, these pauses often co-occurred with the co-ordinator *and*, while advanced participants often

explicitly introduced a new topic after a 3c pause, having more cohesion and coherence in their speech and signalling strategic competence. In line with the findings of the present study, advanced participants studied by Nakatsuhara, Tavakoli & Awwad (2019) used pauses to produce more effective language, i.e. “to provide opportunities for topic development, to justify and evaluate points of discussion, to indicate topic shift, to intensify feelings, and to adhere to rules in conversation” (Nakatsuhara, Tavakoli & Awwad 2019, 37).

Based on the results, the typology used in the present thesis could also be criticized for partly overlapping or obscure categories, which is also reflected in the amendments made by the author of the present study. In addition, it was found that categories 2a (reformulating) and 3a (adding new information) partially overlap. In addition, the subcategory 1d, or pauses to recall items from long-term memory, was found to be difficult to analyse due to difficulties in identifying whether or not a particular item was in fact recalled from long-term memory. This is a nigh-impossible feat in a post hoc analysis. As such, in the present study, pauses to recall items from long term memory wear limited to search for examples and ideas.

Repairs. In line with the findings related to pauses related to access and retrieval difficulty, intermediate participants had more repairs related to access and retrieval difficulty than advanced participants. In addition, both intermediate and advanced participants had comparable numbers of reformulation repairs. However, reformulation repairs were more common in the speech of intermediate and advanced participants with low mean lengths of run (proficiency levels 3 and 5). Less fluent or less proficient speakers may have more need for reformulation repairs. Furthermore, repairs have been thought to subtract from fluency. Surprisingly, as stated in Section 5.1, the quantitative analysis revealed that the advanced group had more mid-clause repairs in their speech than the intermediate group. Based on the present qualitative analysis, advanced participants were more focused on reformulating ideas and aiming towards accuracy of expression or thought than correcting their utterances.

As expected, advanced participants used more repairs related to effective speech delivery than intermediate participants. Based on the analysis, intermediate participants mostly had repairs related to effective delivery on clause borders, and the most common repair in the speech of intermediate participants was repeating the co-ordinator *and* at clause borders, which fits the coherence/cohesion descriptor of the YKI speaking scale

(level 3): “the use of connectors may be incomplete and repetitive” (University of Jyväskylä: YKI speaking scale).

5.3 The interdependence of pause and repair use

The third objective of this study was to examine whether pause and repair use are interdependent. The results showed that i) pauses and repairs of the same category often co-occur: for example, participants have unfilled pauses, pause clusters, and repetitions on clause borders and especially in mid-clause positions when searching for structural and lexical items; ii) repairs, particularly repetitions, are used to avoid and shorten pauses; and iii) a strong correlation was found between pauses and repairs related to effective delivery.

Based on the present qualitative analysis, speakers with higher mean lengths of run employ more mid-clause repairs and speakers with lower mean lengths of run more mid-clause pauses. The present findings suggest that using repairs as opposed to unfilled pauses may not be related to individual speaking style but may in fact have something to do with proficiency level: as hypothesised, advanced speakers may use repairs to maintain speech and to avoid unfilled pauses. In dialogue, such repairs, often manifesting as repetitions, signal to the interlocutor that the speaker is still holding the floor. Notably, more fluent and more proficient participants used repairs to buy time for lexical and structural retrieval, reformulation and efficient delivery in the same way as pauses are used to allow for processing time. Research has shown that too-frequent and too-long pauses have adverse effects on perceived fluency and comprehensibility as the speaker risks losing the interlocutor’s attention (e.g. O’Brien 2014). Following this, maintaining speech through repairs and repetitions may in fact be a better fluency strategy. Peltonen (2017) distinguishes repairs from temporal fluency, categorising them into stalling strategies together with drawls.

In sum, not all mid-clause pauses have to do with lack of proficiency. Still, in the speech of intermediate participants, the frequency of pauses does lend a staccato cadence to their speech. As presented in Section 2, Lehtonen (1979) described the speech of Finnish-speaking L2 speakers “staccato”. Based on the present study, this characterisation describes especially intermediate-level L2 speech, and based on the present qualitative analysis, can be attributed to 1b+1c pause clusters as well as frequent 2b rescue pauses permeating the speech of intermediate participants. Emergent research points out that there is a correlation between L2 speakers’ speech rhythm and their perceived fluency.

5.4 Implications for the Finnish National Certificates of Language Proficiency speaking scale

As presented in Sections 2.3 and 5.2, the YKI speaking scale does include some inferences of L2 speakers' cognitive fluency. However, based on the results, I tentatively propose including short descriptors of pause and repair use in level descriptors 3 to 6. The following descriptors are suggested as additions to the fluency criterion:

- 3 Frequent interruptions preceding lexical and structural search often result in the production of unsophisticated or erroneous language. Pauses are often used in connection with reformulating or rescuing ungrammatical utterances. Pause and repair use is characterised by focus on grammar and lexical and structural search even though pauses and repairs seldom also relate to effective speech delivery, providing opportunities for topic development, evaluation, topic shifts and hedging.
- 4 Despite self-monitoring, interruptions preceding lexical and structural search often result in the production of unsophisticated or erroneous language as well as reformulation of utterances, thoughts, and ideas. Pauses and repairs are sometimes used in connection with reformulating or rescuing ungrammatical utterances. Pause and repair use is characterised by focus on grammar and lexical and structural search even though pauses and repairs sometimes also relate to effective speech delivery, providing opportunities for topic development, evaluation, topic shifts and hedging.
- 5 Due to self-monitoring, interruptions mostly precede successful lexical and structural search and the production of sophisticated language, successful reformulations of ideas and thoughts where necessary, and efforts to increase the perceived effectiveness of delivery. Pauses and repairs are only seldom used in connection with reformulating or rescuing ungrammatical utterances, or unsuccessful or unsophisticated language. Instead, pause and repair use is characterised by flexibility and successful strategy, providing opportunities for topic development, evaluation, topic shifts and hedging.
- 6 Interruptions almost always precede successful lexical search and production of sophisticated language, successful reformulations of ideas and thoughts where necessary, and successful efforts to increase the perceived effectiveness of delivery. Pauses and repairs are almost never used in connection with reformulating or rescuing ungrammatical utterances, or unsuccessful or unsophisticated language.

Instead, pause and repair use is characterised by a high level of flexibility and successful strategy, providing opportunities for topic development, evaluation, topic shifts and hedging.

It should be noted that the descriptors proposed are only tentative and further research is required. However, they serve as a point of reflection: instead of simply focusing on temporal aspects of pause and repair phenomena, attention should also be paid to what an interruption achieves, as opposed to what it takes away. Next, the present study is evaluated in Section 5.5.

5.5 Evaluating the present study

Originally, the purpose of the present study was to study the differences in pause and repair use between intermediate and advanced L2 speakers of English, forgoing the division of speakers into four proficiency groups by their speaking assessment (3, 4, 5 and 6; CEFR B1, B2, C1 and C2) and instead, studying speakers by the test they had taken (intermediate vs. advanced). As such, the purpose of the extreme sampling was to distinguish the most fluent and the least fluent in intermediate and advanced groups, not the most proficient, and to avoid comparing levels 3 and 4 and 5 and 6 as fluency was only one component of their speaking assessment. However, the four groups unintentionally re-emerged with the extreme case sampling by mean length of run: intermediate speakers with the lowest mean lengths of run chosen for the qualitative analysis had all been assessed at speaking level 3, intermediate speakers with highest mean lengths of run at speaking level 4, and advanced speakers with lowest mean lengths of run at speaking level 5, which further triangulates that mean length of run is a reliable measure of both fluency and proficiency. However, only one of the advanced speakers with highest mean lengths of run had been assessed at speaking level 6, which may suggest that mean length of run alone is not enough to distinguish between highly proficient users' proficiency levels. Mean length of run is a good predictor of proficiency although distinguishing between highly proficient users by one measure alone is difficult. Perhaps a comprehensive fluency score should have been used instead of a single best distinguishing measure.

Furthermore, reliability and validity should be considered. As for the validity and reliability of the present study, they are inevitably linked to the authenticity, validity and reliability of the YKI speaking test. Even though the YKI speaking test is a well-designed

speaking test, the language it elicits is not completely authentic. As such, the texts produced by participants are not authentic texts as “an authentic text is a stretch of real language, produced by a real speaker or writer for a real audience and designed to convey a real message of some sort” (Morrow 1977, 13). Furthermore, the present study is a post-hoc study and the speaking assessments it relies on have been provided by graders with access to material that the author of the present study does not have access to, including possible rater rubrics that complement the speaking scale. While the grading process aims for analytical treatment of speech, the participants only receive one holistic grade for each subskill of language ability, including speaking.

In addition, the present study has been designed to take reliability and validity into account and the author has attempted to triangulate references and make defensible choices. The methodology is based on a body of quantitative studies studying temporal fluency features, mixed-methods studies offering new perspectives in addition to temporal fluency, as well as emerging results and a pause and repair use typology by Nakatsuhara, Tavakoli & Awwad (2019, 26). The mixed methods chosen complemented each other and increased the reliability of the present study through triangulation and while the results of the present study cannot be generalised due to its limited scope, they are part of an expanding and developing field of study.

While the typology by Nakatsuhara, Tavakoli & Awwad used in this study is not perfect and there is overlap based on the present analysis, such typologies are examples of how second language pause and repair uses can be operationalised. Having many different perspectives on fluency is especially important in the context of language testing, where it is a common criterion. While operationalising Nakatsuhara, Tavakoli and Awwad’s typology was difficult, it should be noted that any *post hoc* analysis of second speakers’ pause and repair use is based on inferences as the inner workings of the second language speakers’ minds are not known to the researcher. The internal reliability of such studies could be increased by using multiple evaluators and raters in the qualitative analysis stage and thus provide a more comprehensive idea of second language speakers’ use of pause and repair opportunities as well as their effects on cognitive and perceived fluency. Still, the the results of the present study are in line with previous findings (Nakatsuhara, Tavakoli & Awwad 2019, Peltonen 2017) and provide a foundation for future studies.

In conclusion, the present study recommends a new perspective on second language speakers’ pause and repair use. Furthermore, the present study points out a

possible link between learners' pause and repair use and the incomplete proceduralisation of second language prosody and/or rhythm. In any case and regardless of its limitations, this study suggests further research into second language speakers' pause and repair use as well as tentatively recommends their inclusion into conceptualisations and operationalisations of second language fluency.

Next, recommendations for further research are discussed in more detail.

5.6 Recommendations for further research

Fluency is an abstract concept and needs to be carefully defined and operationalised to ensure validity and reliability. While many proficiency scales and rater rubrics share an affinity to CAF or parts thereof, fluency is defined and operationalised in several different ways. As such, it would be interesting to compare and contrast definitions and operationalisations of fluency and pause and repair use across proficiency scales, rater rubrics and grading processes to ensure inter-item reliability and intra- and inter-rater reliability where available. Furthermore, based on the present study, a tentative proposal could be made on including pause and repair use in the fluency criterion of the YKI speaking scale, but any serious proposal would require further research into pause and repair use and their effects on cognitive and perceived fluency. Both Nakatsuhara, Tavakoli & Awwad (2019, 37-38) and the present study distinguish patterns in how intermediate and advanced second language users use their pause and repair opportunities. Lexical and structural search, rescues and reformulations are common in learner speech and are thought to be related to Levelt's Formulation Stage (Levelt 1989 in Nakatsuhara, Tavakoli & Awwad 2019, 37). However, native speakers also sometimes pause to search for words, have occasional dysfluencies in their speech and repair their speech: not all dysfluencies are related to low proficiency. While the native speaker is only one standard for fluent speech, comparing native and L2 speaker pause and repair use might lead to interesting insights into second language users' use of pause and repair opportunities. For example, such an enquiry might include a qualitative analysis of L2 speakers' and native speakers' pause use on four levels and then quantitatively analysing pause and repair use, looking for relationships and associations between different pause and repair use categories.

In addition, researchers such as Segalowitz (2016) have called for expanding the study of fluency features into the social dimension of fluency. In the future, pause and repair use should be studied in interaction to see how different kinds of pauses and repairs

are used in relation to turn-taking, problem-solving and negotiating meaning. Based on the present study, inferences of L2 speakers' flexibility and strategic competence can be made based on their pause and repair use, and pauses and repairs often relate to effective delivery, particularly at the advanced level. However, such inferences can only be confirmed after observing the same effects in interaction.

Finally, although beyond the scope of the present study, the reciprocal relationship between speech rhythm and fluency cannot be dismissed. The present qualitative analysis found that mid-clause (and even mid-phrase) pauses related to rescuing ungrammatical utterances also occurred within completely grammatical segments, suggesting that these pauses may not have been pauses related to rescuing (2b), but rather pauses related to incomplete proceduralisation of e.g. speech rhythm, reduced forms, or syntax, or linguistic chunks. Regardless, the placement, duration and frequency of interruptions has an undeniable effect on speech rhythm, and speech rhythm has been shown to correlate strongly with fluency (e.g. Salomaa 2019). As such, a study combining pause and repair use with suprasegmentals such as speech rhythm would also work in extending fluency dimensions beyond temporal fluency.

Next, the present thesis concludes with an overview of the present study.

6 Conclusion

In conclusion, the present thesis combines a quantitative analysis of 30 intermediate and 30 advanced ESL speakers' temporal fluency features in Finnish National Certificates of Language Proficiency speaking test with a quantitative analysis of the pause and repair use of an extreme case sample. Quantitative and qualitative research methods were combined to answer the following research questions: "Which quantitative measures of fluency distinguish between intermediate and advanced levels of proficiency?"; "how do pause and repair use distinguish between intermediate and advanced levels of proficiency?"; and "are pause and repair use interdependent?"

The present thesis found that the quantitative fluency measures of articulation rate, mean length of run, pause duration, pause frequency, mid-clause pause frequency and mid-clause repair frequency distinguished between intermediate and advanced levels of the Finnish National Certificate of Language Proficiency speaking test, confirming most initial hypotheses, and that the largest difference between the two groups was in the mean lengths of run, on the basis of which extreme cases were sampled for qualitative analysis of pause and repair use. The qualitative analysis these speech samples shows that intermediate participants had slightly more pauses and repairs related to access and retrieval difficulty than advanced participants and that advanced participants were successful in their lexical and structural search more often than intermediate participants, confirming my initial hypothesis. Furthermore, not all pauses categorised as relating to access and retrieval difficulty are necessarily such: sometimes participants would also pause before and after culturally or socially difficult concepts, signalling their hesitation in using loaded expressions. In addition, it was found that in addition to pausing after having produced sophisticated language, participants have pauses also after having produced monitored erroneous or generic speech, as well as reformulated speech. In contrast, intermediate participants had more pauses related to reformulations than advanced participants, confirming my initial hypothesis. In specific, intermediate participants used more of their reformulation pauses on rescuing and reformulating ungrammatical utterances, whereas advanced participants focused on reformulating ideas instead of rescue pauses or reformulating minor grammatical inaccuracies. Finally, it was found that pauses also occur within grammatical utterances where no lexical or structural search, restructuring or reformulation or improvement of delivery is attempted. As for repairs related to reformulations, intermediate and advanced participants with low MLR

had the most reformulation repairs, but together, intermediate participants had more reformulation repairs than advanced participants. Furthermore, in using both pauses and repairs, advanced participants and intermediate participants with high mean lengths of run were more focused on reformulating ideas and aiming towards accuracy of expression or thought than correcting their utterances.

In addition, the results showed that advanced participants used more of their pause and repair opportunities for effective delivery. As for repairs however, intermediate participants with low MLR had the fewest effective delivery repairs, followed by advanced participants with low MLR, whereas participants with high MLR used more repair opportunities on effective delivery repairs regardless of their proficiency level. At the threshold of intermediate level with low mean lengths of run and advanced level with high mean lengths of run, advanced participants with low mean lengths have an increase in reformulation repairs at the expense of access and retrieval difficulty repairs and effective delivery repairs. Furthermore, based on the current analysis, intermediate participants had most of their effective delivery pauses and repairs on clause borders. Finally, the present study showed that there is pause-repair co-occurrence in the speech of intermediate and advanced participants; that repairs are used by especially advanced participants to avoid the need for pausing; that pauses and repairs can be used by learners to achieve the same end; and that there is a strong correlation between pauses and repairs related to effective speech delivery.

Based on the results, pause and repair use are integral in making inferences of L2 speakers' cognitive fluency and should not be overlooked in fluency studies or proficiency testing. The present study tentatively suggests that pause and repair use descriptors should be added into the fluency criterion of the YKI scale while the author advises emphasising caution against using the results in any definitions or operationalisations before they have been extensively studied and the methodology validated. In the future, it is suggested that definitions and operationalisations of fluency and pause and repair use be compared and contrasted across proficiency scales, rater rubrics and grading processes. In addition, native and non-native speakers' pause and repair use could be studied using a similar study design. Furthermore, pause and repair use should be studied in interaction in relation to problem-solving and turn-taking. Researchers, such as Segalowitz (2016), are calling for expanding the study of fluency features into the social dimension of fluency. Based on the present study, inferences of L2 speakers' flexibility and strategic competence can be made based on their pause and

repair use. However, such inferences can only be confirmed after observing the same effects in interaction. Finally, it is suggested that pause and repair use be combined with the study of suprasegmentals to extend fluency dimensions beyond temporal fluency.

List of References

- Bosker, Hans Rutger, Anne-France Pinget, Hugo Quené, Ted Sanders, and Nivja H. de Jong. [2012] 2013. "What makes speech sound fluent? The contributions of pauses, speed and repairs." *Language Testing* 30 (2):159–175.
- Chambers, Francine 1997. "What do we mean by fluency?" *System* 25 (4): 535–544.
- Council of Europe 2011. *Common European Framework of Reference for Languages: Learning, teaching and assessment*. Cambridge University Press.
- De Jong, Nel, and Charles A. Perfetti. 2011. "Fluency training in the ESL classroom: An experimental study of fluency development and proceduralisation." *Language Learning* 61 (2): 533–568.
- De Jong, Nivja H. 2016. "Fluency in second language assessment." In *Handbook of Second Language Assessment*, edited by Tsagari Dina and Banerjee Jayanti, 203–218. de Gruyter Mouton.
- De Jong, Nivja H., Margarita P. Steinel, Arjen Florijn, Rob Schoonen, and Jan H. Hulstijn. 2013. "Linguistic skills and speaking fluency in a second language." *Applied Psycholinguistics* 34 (4): 893–916.
- De Jong, Nivja H., and Ton Wempe. 2009. "Praat script to detect syllable nuclei and measure speech rate automatically." *Behaviour Research Methods* 41: 385–390.
- Dewaele, J.-M., and A. Furnham (2000). Personality and speech production: A pilot study of second language learners. *Personality and Individual Differences*, 28 (2): 355–365.
- Dörnyei, Zoltan. 2007. *Research Methods in Applied Linguistics: Quantitative, Qualitative, and Mixed Methodologies*. Oxford: Oxford University Press.
- Fillmore, C. J. (1979). "On fluency." In *Individual differences in language ability and language behavior*, edited by C. J. Fillmore, D. Kempler and W. S. Y. Wang, 85–102. New York: Academic Press.
- Finnish National Agency for Education. 2011. *The framework of the Finnish National Certificates of Language Proficiency 2011*. Order 24/011/2011. Finnish National Agency for Education.
- Flege, Jim. 1999. "Age of Learning and Second Language Speech." In *Second Language Acquisition and the Critical Period Hypothesis*, edited by D. Birdsong, 101–132. Hillsdale, NJ: Lawrence Erlbaum.
- Fulcher, Glenn. 2003. *Testing Second Language Speaking*. Pearson Education.
- 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. [Available at <http://ebookcentral.proquest.com/lib/kutu/detail.action?docID=1040791>. Accessed 17 February 2021.]
- Housen, Alex, Folkert Kuiken, and Ineke Vedder, eds. 2012. *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*. John Benjamins Publishing Company.
- Kahng, Jimin. 2014. "Exploring Utterance and Cognitive Fluency of L1 and L2 English Speakers: Temporal Measures and Stimulated Recall." *Language Learning* 64 (4): 809–854.
- Kormos, Judit. 2006. *Speech production and second language acquisition*. Lawrence Erlbaum Associates Publishers.
- Larsen-Freeman, Diane. 2009. "Adjusting expectations: The study of complexity, accuracy and fluency in second language acquisition." *Applied Linguistics* 30 (4): 579–589.
- Larson-Hall, Jenifer. 2016. *A Guide to Doing Statistics in Second Language Research Using SPSS and R*. Routledge London.

- Lehtonen, Jaakko. 1979. "Speech rate and pauses in the English of Finns, Swedish-speaking Finns, and Swedes." In *Perception and Production of English: Papers on Interlanguage*, edited by R. Malmberg, AFTIL 6, 35–75. Department of English, Åbo Akademi.
- Lennon, Paul. 1990. "Investigating fluency in EFL: A quantitative approach." *Language Learning* 40 (3): 387–417.
- . 2000. "The lexical element in spoken second language fluency." In *Perspectives on fluency*, edited by H. Riggenbach, 25–42. Ann Arbor: The University of Michigan Press.
- Leppänen, Sirpa, Anne Pitkänen-Huhta, Tarja Nikula, Samu Kytölä, Timo Törmäkangas, Kari Nissinen, Leila Kääntä, Tiina Räisänen, Mikko Laitinen, Heidi Koskela, Salla Lähdesmäki, and Henna Jousmäki. 2011. "National Survey on the English Language in Finland: Uses, meanings and attitudes." *Studies in Variation, Contacts and Change in English*, Volume 5. (Accessed 27 August 2017).
- Liddick, Anthony. 2011. *An Introduction to Conversational Analysis*. London: Continuum Books.
- Lintunen, Pekka, Maarit Mutta, and Pauliina Peltonen, eds. 2019. *Fluency in L2 Learning and Use*. Multilingual Matters.
- McNamara, Tim. 2000. *Language Testing*. Oxford University Press.
- Morrow, K. 1977. "Authentic texts and ESP." In *English for Specific Purposes*, edited by Holden S., 13–17. London: Modern English Publications.
- Nakatsuhara, Fumiyo, Parvaneh Tavakoli, and Anas Awwad. 2019. *Towards a Model of Multi-dimensional Performance of C1 Level Speakers Assessed in the APTIS Speaking Test*. British Council.
- O'Brien, Mary Grantham. 2014. "L2 learners' assessments of accentedness, fluency, and comprehensibility of native and nonnative German speech." *Language Learning* 64 Vol. 4, 715–748.
- Olkkonen, Sanna. 2017. "Processing limitations in L2 fluency: Analysis of inaccuracies in lexical access." *Apples – Journal of Applied Language Studies* Vol. 11 (1): 19–41.
- Olkkonen, Sanna and Pauliina Peltonen 2017. "Mitä on toisen kielen sujuvuus? Näkökulmia kognitiivisen ja puhetuoksen sujuvuuden tutkimuksesta." [What is second language fluency? Insights into researching cognitive and utterance fluency.] In *Näkökulmia toisen kielen puheeseen – Insights into second language speech*, edited by Kuronen, M., P. Lintunen & T. Nieminen. AFinLA-e. Soveltavan kielen tutkimuksia. 2017, Vol. 10, 234-257.
- Ortega, Lourdes. 2009. *Understanding Second Language Acquisition*. London: Routledge.
- Peltonen, Pauliina. 2017. "Temporal fluency and problem-solving in interaction: An exploratory study of fluency resources in L2 dialogue." *System* Vol 70: 1–13.
- Peltonen, Pauliina, and Pekka Lintunen. 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.
- Pietilä, Päivi, and Pekka Lintunen. 2014. "Kielen oppiminen ja opettaminen." In *Kuinka kieltä opitaan* [How Languages are Learned], edited by Päivi Pietilä and Pekka Lintunen. Turku: Gaudeamus.
- Revesz, Andrea. 2016. "The Effects of Complexity, Accuracy, and Fluency on Communicative Adequacy in Oral Task Performance." *Applied Linguistics* 37 (6): 828–848.
- Robinson, Peter. 2001. "Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA." In P. Robinson (Ed.). *Cognition and second language instruction*, edited by Peter Robinson, 287– 318. Cambridge: Cambridge University Press.
- . 2003. "The Cognition Hypothesis, task design and adult task-based language learning." *Second Language Studies* 21 (2): 45–107.

- . 2005. “Cognitive complexity and task sequencing: Studies in a Componential Framework for second language task design.” *International Review of Applied Linguistics in Language Teaching* 43 (1), 1–32.
- Salomaa, Emma. 2019. *The Development of Speech Rhythm and Fluency of Advanced English Learners: A Mixed Methods Study of the Correlation Between Native Speaker Evaluations and Acoustic Measures*. MA Thesis. University of Turku.
- Segalowitz, Norman. 2010. *Cognitive bases of second language fluency*. New York: Routledge.
- . 2016. “Second language fluency and its underlying cognitive and social determinants.” *International Review of Applied Linguistics in Language Teaching* 54 (2): 79–95.
- Skehan, Peter. 1998. *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.
- . 2000. “Choosing and Using Tasks: Research Contributions.” *TESOL in Context*. Volume 10 (2): 18–26.
- . 2009. “Modelling Second Language Performance: Integrating Complexity, Accuracy, Fluency, and Lexis.” *Applied Linguistics* 30 (4), 510–532.
- Tavakoli, Parvaneh, and Peter Skehan. 2005. “Strategic planning, task structure and performance testing.” In *Planning and task performance in a second language*, edited by Ellis Rod, 239–277. Amsterdam: John Benjamins Publishing Company.
- Tominaga, Yuko. 2011. “An Analysis of English Pronunciation of Japanese Learners: From the Viewpoint of EIL.” *Journal of Pan-Pacific Association of Applied Linguistics* 15 (2): 45–57. Available at <https://eric.ed.gov/?id=EJ97991> (accessed 2 March 2020).
- Towell, Richard. 2012. “Complexity, Accuracy and Fluency from the perspective of psycholinguistic Second Language Acquisition research.” In *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*, edited by Alex Housen, Folkert Kuiken, and Ineke Vedder, 47–70.
- University of Jyväskylä 2011. *The Finnish National Certificates of Language Proficiency Speaking Scale*. [Available in Finnish at https://www.jyu.fi/hytk/fi/laitokset/solki/yki/yleista/tietoakielitutkinnoista/puhumisen_arviointikriteerit.pdf]. Accessed 30 March 2020.
- VanPatten, Bill, and Alessandro G. Benati. 2010. *Key terms in Second Language Acquisition*. London: Bloomsbury.
- Wright, CEM, and Parvaneh Tavakoli. 2016. “New directions and developments in defining, analyzing and measuring L2 speech fluency.” *International Review of Applied Linguistics in Language Teaching*, 54 (2): 73–77.

Appendix 1. Example transcription

104656

1 question is if I trust finnish politicians [{3d_EC*_hh*_1
2 um {*_h*_6}][EC_rep3_it's: it's] definitely not an easy
3 question [{3d_EC*_h*_1.272} uh] 'cause there are so many
4 politicians and [uh] most of them(hh) I don't know
5 [{3d_EC*_h*_1}] or actually I don't know any one of them
6 personally (.) [EC_rep2_which is kind of- which is really]
7 hard because {3a_MC*_h*_9} to trust someone
8 {3a_MC*_h*_6} you would really need to know (.)
9 someone personally (.) because only after that you really
10 know [MC_rep2_that {2a_MC*_h*_1} what] the person is from
11 his or her background [{3a_EC*_h*_6} uh (.6)] what is
12 the family ties (3a_MC.3) behaviour models personality traits
13 {3a_MC*_h*_5} education {3a_MC*_h*_5} uh experience
14 [{3a_EC*_h*_4} um (.4)] really kind of the behaviour with
15 people how do they treat [uh uh] other people how do they
16 treat environment {3a_MC*_h*_5} animals: weakest in the
17 society and so on {EC*pt*_1.2} but eh naturally every one
18 of us {1d_MC*_hh*_1.1} has to vote someone so I'm trying to
19 look [MC_rep3_at the {1a*_hh*_2.6} um: for example] uh the]
20 expertise and education on a person {3a_EC*_h*_6} and uh try
21 to read and uh listen interviews from that particular person
22 to understand what he or she is alike [{3c_EC*_h*_9} uh]
23 it's pretty much about communication [(3a_EC.6)
24 uh[EC_rep3_and- and]] then seeing that politician
25 communicating with each other {3a_EC*_h*_4} how does that
26 happen [{EC*_h*_6} uh] what kind of dialogue there is and
27 so on {3a_EC*pt*_1.1} also their symbolic {1b_MC*_h*_8}
28 acts (1c_MC.4) are needed (1b_MC.5) to create trust I mean
29 if a politician {1a_MC*_h*_5} turns out to be
30 [{1b_MC*_h*_7} uh (.3)] dirty(hh) in [MC_rep2_hi- in his or
31 her][1a_{MC*_h*_6} um {*_pt*_8}] personal life
32 [MC_rep3_the-then] it would be hard

Appendix 2.

Table 3. Participants' pause use.

Participants' pause use (%)																
Description	ID	MLR	LVL	1a	1b	1c	1d	Total	2a	2b	Total	3a	3b	3c	3d	Total
INT-low MLR	105571	3.2	3	5	28	10	0	43	0	31	31	15	3	8	0	26
	105466	3.3	3	5	28	8	0	41	5	23	28	21	3	8	0	32
	106052	3.7	3	4	25	4	0	33	6	31	37	21	4	6	0	31
INT-low MLR by subcategory (%)				5	27	7	0	39	4	28	32	19	3	7	0	29
INT-low MLR by category (%)				39					32			29				
INT-high	106204	9.1	4	16	19	9	0	44	22	6	28	25	0	3	0	28
	104656	10	4	12	16	4	0	32	4	8	12	44	0	0	3	47
	106561	11.2	4	16	13	10	3	42	6	6	12	35	3	10	0	48
INT-high MLR by subcategory (%)				14	16	8	1	39	11	8	19	34	1	5	3	43
INT-high MLR by category (%)				39					19			43				
INT LVL BY SUBCATEGORY (%)				8	22	7	0	37	7	20	27	25	2	6	1	34
INT LVL BY CATEGORY (%)				37					27			34				
ADV-low	68024	5.6	5	19	9	12	0	40	0	7	7	44	5	5	0	54
	68431	5.8	5	33	10	2	0	45	10	0	10	38	2	5	0	45
	68435	6	5	7	5	5	0	17	14	10	24	40	12	7	0	59
ADV-low MLR by subcategory (%)				20	8	6	0	34	8	6	14	41	6	6	0	53
ADV-low MLR by category (%)				34					14			53				
ADV-high	70006	14	5	18	4	0	0	22	14	0	14	50	4	11	0	65
	70000	15.5	6	22	4	0	0	26	22	4	26	33	4	11	0	48
	70415	15.5	5	10	5	10	0	25	5	0	5	48	19	5	0	72
ADV-high MLR by subcategory (%)				17	4	3	0	24	14	4	18	43	8	9	0	60
ADV-high MLR by category (%)				24					15			60				
ADV LVL BY SUBCATEGORY (%)				19	6	5	0	30	10	4	14	42	7	7	0	56
ADV LVL BY CATEGORY (%)				30					14			56				

Appendix 3.

Table 4. Participants' repair use.

Participants' repair use (%)																
Description	ID	MLR	LVL	1a	1b	1c	1d	Total	2a	2b	Total	3a	3b	3c	3d	Total
INT-low MLR	105571	3.2	3	0	20	20	0	40	0	40	40	0	0	20	0	20
	105466	3.3	3	14	14	0	0	28	29	0	29	29	14	0	0	43
	106052	3.7	3	33	17	0	0	50	50	0	50	0	0	0	0	0
INT-low MLR by subcategory (%)				17	17	6	0	39	28	11	39	11	6	6	0	22
INT-low MLR by category (%)				39				39			22					
INT-high	106204	9.1	4	26	7	0	0	33	26	4	30	37	0	0	0	37
	104656	10	4	0	0	0	0	0	29	0	29	43	14	0	14	71
	106561	11.2	4	0	0	0	0	0	60	0	60	0	0	40	0	40
INT-high MLR by subcategory (%)				18	5	0	0	23	31	3	34	33	3	5	3	44
INT-high MLR by category (%)				23				34			44					
INT LVL BY SUBCATEGORY (%)				18	9	2	0	29	30	5	35	26	4	11	2	43
INT LVL BY CATEGORY (%)				29				35			43					
ADV-low	68024	5.6	5	0	0	33	0	33	0	0	0	66	0	0	0	66
	68431	5.8	5	0	0	25	0	25	63	0	63	13	0	0	0	13
	68435	6	5	13	0	20	0	33	47	0	47	33	7	0	0	40
ADV-low MLR by subcategory (%)				8	0	12	0	20	46	0	46	31	4	0	0	35
ADV-low MLR by category (%)				20				46			35					
ADV-high	70006	14	5	31	0	0	0	31	10	0	10	48	7	3	0	58
	70000	15.5	6	30	0	4	0	34	22	4	26	22	13	4	0	39
	70415	15.5	5	0	0	0	0	0	25	0	25	63	13	0	0	76
ADV-high MLR by subcategory (%)				27	0	2	0	29	17	2	19	40	10	3	0	53
ADV-high MLR by category (%)				29				19			53					
ADV LVL BY SUBCATEGORY (%)				21	0	5	0	26	26	1	27	37	8	2	0	47
ADV LVL BY CATEGORY (%)				26				27			47					

Finnish summary

Johdanto ja teoria

Tämä pro gradu -tutkielma käsittelee englantia toisena kielenään keskitasolla ja edistyneellä tasolla puhuvien suomenkielisten aikuisten toisen kielen puheen sujuvuutta. Tutkielmassa tutkitaan määrällisin ja laadullisin tutkimusmetodein puheen temporaalisia piirteitä sekä sitä, miten keskitason ja edistyneen tason kielenkäyttäjät käyttävät taukoja ja korjauksia Yleistä kielitutkintoa (YKI) varten tallennetussa kohdekielisessä puheessa. Yleinen kielitutkinto tarjoaa luvanvaraisesti tutkimuskäyttöön paitsi aitoja puhenäytteitä, myös esimerkin arviointikriteerien huolellisen määrittelyn käytännön merkityksestä. Osallistujan YKI-kokeesta tai muusta kielitestistä saama arvosana voi vaikuttaa merkittävästi tämän asemaan ja mahdollisuuksiin työelämässä tai yhteiskunnassa, mikä entisestään korostaa sujuvuuden ja muiden testeissä arvioitavien kriteerien validiteetin ja reliabiliteetin arvioinnin sekä uusien tutkimustulosten huomioimisen tärkeyttä. Mittareiden valinta on erityisen tärkeää sujuvuuden arvioinnissa (esim. De Jong 2016, 206). Perinteisessä temporaaalisessa sujuvuustutkimuksessa kuitenkin keskitytään usein temporaalisiin mittareihin ja epäsujuvuuksien (dysfluency) määrään ja niiden sijaintiin puheessa sen sijaan, että pohdittaisiin, mihin puhuja esimerkiksi taukoja tai korjauksia oikeastaan käyttää.

Tutkimuksessa käytettyjä tärkeimpiä viitekehyksiä ovat ensinnäkin temporaaლისen sujuvuuden puolella Skehanin (Skehan 1998 ja 2000) luoma CAF-viitekehys, joka määrittelee kielitaidon osa-alueiksi laajuuden, tarkkuuden ja sujuvuuden (Complexity, Accuracy, Fluency), toiseksi jo vuosikymmeniä vallalla ollut psykolingvistinen käsitys sujuvuudesta puheen suunnittelua ja tuottamista koskevana psykolingvistisena prosessina (mm. Lennon 1990), ja lisäksi käsitys kognitiivisen sujuvuuden vaikutuksista puhetuotoksen sujuvuuteen (utterance fluency) ja toisaalta havaittuun tai tulkittuun sujuvuuteen (perceived fluency) (mm. Segalowitz 2010). Kuten aiemmin mainittu, noin 2010-luvulle asti sujuvuutta onkin tutkittu eniten määrällisenä suurena. Perinteisesti sujuvuus on jaettu kolmeen alakategoriaan: nopeuteen (speed fluency), tauottamiseen (breakdown fluency) ja korjauksiin (repair fluency) (Skehan 2009), suomeksi esim. Olkkonen & Peltonen 2017, 242). Temporaaლისen sujuvuuden mittareita ovat esimerkiksi puhe- ja artikulaationopeus, puheessa ilmenevä tauotus eli hiljaiset ja täytetyt tauot, sekä toistot ja uudelleenmuotoilut. Joitakin mittareita yleisesti käytettyjä voi kuitenkin olla

vaikea sijoittaa Skehanin kolmijakoon. Yksi tällainen mittari on puhunnosten keskipituus (mean length of run), jossa on piirteitä useammasta kategoriasta.

Viimeaikaisissa tutkimuksissa (esim. Peltonen & Lintunen 2016, Nakatsuhara, Tavakoli & Awwad 2019) on osoitettu, että tauot ja korjaukset eivät välttämättä yksinomaan heikennä toisen kielen puhujien sujuvuutta. Lisäksi Peltonen (2017) on osoittanut, että puhujat voivat käyttää korjauksia paitsi oman, myös keskustelukumppaninsa sujuvuuden ylläpitämiseen. Esimerkiksi Nakatsuharan, Tavakolin ja Awwadin (2019) tutkimuksen mukaan taukoja ja korjauksia käytetäänkin eri tarkoituksiin: puhujat käyttävät puheen taukoja ja korjauksia leksikaaliseen ja rakenteelliseen hakuun ja palauttamiseen (access and retrieval), puhunnosten uudelleenmuotoiluun (reformulating) sekä tehokkaaseen sanalliseen viestintään (effective speech delivery) (Nakatsuhara, Tavakoli & Awwad 2019, 26). Nakatsuharan, Tavakolin ja Awwadin tutkimuksessa kahden eritasoisen (B2 ja C1) ryhmän välillä havaittiin eroja: edistyneemmät kielenkäyttäjät onnistuivat sanojen sekä rakenteiden haussa ja palauttamisessa ylemmän keskitason kielenkäyttäjiiä useammin. Lisäksi edistyneet kielenkäyttäjät käyttivät taukoja ja korjauksia tehokkaaseen viestintään useammin kuin ylemmän keskitason kielenkäyttäjät, jotka keskittyivät enemmän vähäisten virheiden korjaamiseen (Nakatsuhara, Tavakoli ja Awwad 2019, 37-38). Taukojen ja korjausten välillä havaittiin yhteys: korjauksia edeltää usein tauko, ja korjauksia esiintyi myös niiden tutkimushenkilöiden joukossa, joiden kielitaito oli korkea (*ibid.*).

Lisätutkimuksille on kuitenkin aihetta. Tässä tutkimuksessa sujuvuuden arviointiin käytettiin Opetushallituksen järjestämän Yleisen kielitutkinnon (YKI) suullisen kokeen puhenäytteitä. YKI-kokeen arviointi perustuu kuusiportaiseen, kriteeripohjaiseen arviointiin ja taitotasoasteikkoon, jossa on kuvaukset jokaiselle kielen osa-alueelle sekä näiden osatekijöille, kuten puheen sujuvuudelle, kullakin taitotasolla. YKI-tutkintojen suoritteista on koottu korpus tutkimuskäyttöön. Korpuksen sisältyvissä kokeissa testatuista henkilöistä on saatavilla joitain yleisiä taustatietoja (mm. ikä ja koulutustausta), jotka mahdollistavat esimerkiksi kirjoitettujen tai puhenäytteiden valikoimisen kulloisenkin tutkimuksen erityispiirteiden mukaan. Muilta osin näytteet ovat anonymisoituja.

Tässä tutkimuksessa hain vastausta kolmeen kysymykseen. Ensimmäkin, mitkä kvantitatiiviset tekijät erottavat keskitason ja edistyneen tason puhujia Yleisen kielitutkinnon puhumisen osakokeessa? Hypoteesini ovat, että edistyneiden puhujien

puhenopeus on keskitason puhujia nopeampi, edistyneiden puhujien puhunnokset ovat keskimäärin keskitason puhujien puhunnoksia pidempiä, edistyneiden puhujien puhe tauottuu harvemmin ja taukojen kokonaiskesto on lyhyempi verrattuna keskitason ryhmään (kuten esim. De Jong et al. 2013 ja Kahng 2014), ja että Peltosen (2017, 10) havaintojen mukaisesti edistyneen tason puhujat korjaavat puhettaan keskitason puhujia useammin. Toinen tutkimuskysymys on, miten keskitason ja edistyneen tason toisen kielen puhujien taukojen ja korjausten käyttötavat eroavat toisistaan. Hypoteesini on Nakatsuharan, Tavakolin & Awwadin (2019, 26-27) tulosten suuntaisesti, että sekä edistyneiden että keskitason puhujien näytteissä ilmenee leksikaaliseen tai rakenteelliseen hakuun liittyviä taukoja, jotka johtavat edistyneiden puhujien onnistumiseen keskitason puhujia useammin, ja että edistyneet puhujat käyttävät taukoja ja korjauksia puhunnosten uudelleenmuotoilujen sijaan tehokkaaseen viestimiseen (lisätiedoilla, esimerkeillä ja selityksillä, mielipiteillä ja kommentteilla, aiheen vaihtamisella sekä aiheiden välttämällä). Sen sijaan keskitason puhujien puhenäytteissä esiintyisi puhunnosten uudelleenmuotoiluihin liittyviä taukoja useammin kuin edistyneiden puhenäytteissä. Kolmas kysymys on, onko taukojen ja korjausten käytön välillä yhteyttä? Nakatsuharan, Tavakolin & Awwadin (2019, 37) tutkimuksen perusteella hypoteesini on, että kyllä on ja että niitä käytetään puheessa samanlaisiin tarkoituksiin joko toisensa yhteydessä tai asemesta – ja että edistyneen tason puhujat käyttävät puheessaan korjauksia välttääkseen sujuvuutta heikentäviä hiljaisia taukoja (Peltonen 2017, 10).

Aineisto ja tutkimusmenetelmät

Tutkimuksen kohteeksi valittiin 30 keskitasoista ja 30 edistynyttä, Yleisen kielitutkinnon englannin kielikokeen vuonna 2011 tai sen jälkeen suorittanutta aikuista kielenkäyttäjää. Osallistujat valikoitiin siten, että keskitason ja edistyneiden ryhmät olivat ikä- ja sukupuolijakaumaltaan samankaltaiset, ja kaikki ryhmiin kuuluvat olivat suorittaneet vähintään toisen asteen tutkinnon. Puhenäytteiden sujuvuutta mitattiin sekä määrällisillä että laadullisilla metodeilla sujuvuustutkimusten nykysuuntausten mukaisesti (esim. Peltonen & Lintunen 2016).

Tutkimuksen puhenäytteet analysoitiin akustisesti ja transkriptioitiin tietokoneavusteisesti. Analysoinnissa käytettiin puheanalyysiohjelma Praatia (esimerkkitranskriptio liitteessä 1). Näin näytteistä voitiin johtaa määrälliset taukojen, täytettyjen taukojen ja korjausten arvot puheen sujuvuuden analysointiin; tässä tutkimuksessa tauoksi määriteltiin äänenvoimakkuuden ≥ 25 dB:n alentuminen

≥300 ms:n ajaksi, jotta esimerkiksi klusiilien okklusiovaihetta ei rekisteröity tauoiksi.

Analysoinnissa käytettäviksi mittareiksi valikoituivat seuraavat:

- artikulaationopeus (articulation rate), eli puhuttujen tavujen määrä (hiljaiset tauot poistettuna) jaettuna näytteen kestolla ja kerrottuna 60:llä (tavua/ minuutti)
- puhunnosten keskipituus (mean length of run), puhunnosten keskimääräinen kesto tavuissa taukojen (≥300 ms:n hiljaisten taukojen tai taukoryppäiden) välillä
- hiljaisten taukojen kokonaiskesto (total pause duration), puheen taukojen kokonaiskesto jaettuna näytteen kestolla ja kerrottuna 60:llä (s/min)
- hiljaisten taukojen tiheys (pause frequency), puheen taukojen kokonaismäärä jaettuna näytteen kestolla ja kerrottuna 60:llä (tauoja/min)
- taukojen tiheys lauseen keskellä (mid-clause pause frequency), puheen hiljaisten taukojen lukumäärä lauseiden keskellä jaettuna näytteen kestolla ja kerrottuna 60:llä (tauoja/min)
- korjausten tiheys (repair frequency), kokonaisten tai osittaisten leksikaalisten yksiköiden toistojen ja uudelleenmuotoilujen lukumäärä jaettuna näytteen kestolla ja kerrottuna 60:llä (korjauksia/min)
- korjausten tiheys lauseen keskellä (mid-clause pause frequency), korjausten kokonaismäärä jaettuna näytteen kestolla ja kerrottuna 60:llä (korjauksia/min).

Mittarit suhteutettiin kunkin puhenäytteen keston, jotta näytteet ja havainnot olisivat keskenään vertailukelpoisia. Tilastollisesti suurimmat erot havaittiin siis puhunnosten keskipituudessa, minkä vuoksi näytteistä valittiin tällä perusteella poikkeavimmat tapaukset laadulliseen tarkasteluun. 60:n puhenäytteen kvalitatiivinen analyysi olisi ollut tutkimuksen raameissa haastavaa, minkä vuoksi keskitason ja edistyneiden puhujien ryhmistä valikoitiin poikkeavien tapausten analyysiin kolme näytettä, joissa puhunnosten keskipituus oli ryhmän sisällä lyhin, ja kolme, joissa se oli ryhmän pisin. Poikkeavien tapausten analyysin tulokset eivät välttämättä ole yleistettävissä yhtä suoraan kuin vaikkapa edustavan otoksen tai satunnaisotannan, mutta Dörnyein mukaan yksittäisten tapausten erottuminen voikin johtaa uusiin havaintoihin (Dörnyei 2007, 153). Tässä tutkimuksessa poikkeavien tapausten analyysi tarjosi mahdollisuuden erityisesti kielitaitotasojen rajapinnan tarkasteluun. Poikkeavien tapausten analyysiin valikoiduissa

yhteensä kahdessatoista näytteessä esiintyvät tauot ja korjaukset analysoitiin kvalitatiivisesti Nakatsuharan, Tavakolin & Awwadin (2019) kehittämän typologian avulla: tauot ja korjaukset jaoteltiin leksikaalisen ja rakenteellisen haun ja palauttamisen vaikeuksiin (access and retrieval difficulty), puhunnosten uudelleenmuotoiluun (reformulation) ja tehokkaaseen sanalliseen viestintään (effective speech delivery) sekä näiden alakategorioihin. Tämän kvalitatiivisen analyysin tuloksia vertailemalla saatiin vastaukset kahteen tutkimuskysymykseen: miten keskitason ja edistyneen tason toisen kielen puhujien taukojen ja korjausten käyttötavat eroavat toisistaan, ja onko taukojen ja korjausten käytön välillä yhteyttä.

Tulokset ja päätelmät

Näytteiden kvantitatiivisessa analyysissä keskitason ja edistyneen tason ryhmien välillä havaittiin tilastollisesti merkittäviä eroja kaikilla mittareilla korjausten kokonaistiheydettä lukuun ottamatta. Kvantitatiivisessa analyysissä keskitason ja edistyneen tason ryhmien välillä ilmeni eroja lähes kaikilla sujuvuuden mittareilla: edistyneen ryhmän artikulaationopeus oli keskimäärin 20,5 % nopeampi, puhunnosten keskipituus 47,5 % pidempi, taukojen kokonaiskesto 25,8 % lyhyempi ja heidän puheessaan oli 13,4 % harvemmin taukoja kuin keskitason ryhmällä. Vaikka korjausten kokonaistiheydessä ei havaittu tilastollisesti merkittäviä eroja, edistynyt ryhmä käytti korjauksia 29,4 % useammin lauseen keskellä kuin keskitason ryhmä. Suurimmat erot ilmenivät puhunnosten keskipituudessa: edistyneen ryhmän puhunnosten keskipituus oli 47,5 % pidempi kuin keskitason ryhmän, ja molemmista ryhmistä valittiin poikkeavimmat tapaukset kvalitatiiviseen analyysiin. Kummastakin ryhmästä valittiin kolme puhujaa, joiden puhunnokset olivat ryhmänsä sisällä keskimäärin pisimpiä ja kolme, joiden puhunnokset olivat ryhmän sisällä keskimäärin lyhimpiä, yhteensä siis 12 puhujaa.

Kvalitatiivisen analyysin tuloksista ilmeni, että keskitason puhujien puhenäytteissä esiintyi edistyneiden puhenäytteisiin verrattuna hieman enemmän leksikaaliseen ja rakenteelliseen hakuun ja palauttamiseen liittyviä taukoja: niin keskimäärin pisimpiä kuin lyhimpiäkin puhunnoksia tuottavien keskitason puhujien tauoista 39 % liittyi leksikaaliseen ja rakenteelliseen hakuun ja palauttamiseen, kun taas edistyneiden puhujien osalta keskimäärin lyhimpiä puhunnoksia tuottaneiden vastaava luku on 34 % ja pisimpiä tuottaneiden 24 %. Lyhimpiä puhunnoksia tuottaneiden keskitason puhujien puheessa tauot edelsivät onnistuneita leksikaalisia ja rakenteellisia

hakuja 5 %:ssa ja pisimpiä puhunnoksia tuottaneiden puheessa 14 %:ssa tapauksista, kun taas lyhimpiä puhunnoksia tuottaneiden edistyneiden puhujien puheessa saman kategorian hauista onnistui 20 % ja pisimpiä puhunnoksia tuottaneiden hauista 17 %.

Uudelleenmuotoiluja edeltäviä taukoja esiintyi useammin keskitason kuin edistyneen tason puhujien puhenäytteissä: keskipituudeltaan lyhimpiä puhunnoksia tuottaneiden keskitason puhujien tauoista 32 % ja pisimpiä puhunnoksia tuottaneiden tauoista 19 % edelsivät uudelleenmuotoilua, ja edistyneen tason puhujilla vastaavat luvut olivat 14 % ja 15 %.

Edistyneen tason puhujat käyttivät taukoja keskitason puhujia useammin tehokkaaseen viestintään, ja puhunnosten keskipituuden sekä tehokkaan viestinnän välillä havaittiin yhteys: lyhimpiä puhunnoksia tuottaneiden keskitason puhujien tauoista 29 % edelsi tehokasta viestintää ja pisimpiä puhunnoksia tuottaneiden keskitason puhujien tauoista 43 %, lyhimpiä puhunnoksia tuottaneiden edistyneiden puhujien tauoista 53 % ja pisimpiä puhunnoksia tuottaneiden keskitason puhujien tauoista 60 %.

Korjauksia analysoitaessa havaittiin, että lyhimpiä puhunnoksia tuottaneiden keskitason puhujien korjauksista 39 % ja pisimpiä puhunnoksia tuottaneiden keskitason puhujien 23 % liittyi leksikaaliseen ja rakenteelliseen hakuun ja palauttamiseen; edistyneen tason puhujien vastaavat osuudet olivat 20 % ja 29 %. Kvalitatiivisen analyysin mukaan leksikaaliseen ja rakenteelliseen hakuun ja palauttamiseen liittyviä korjauksia ilmenee vastaavien taukojen kanssa tai niiden asemesta.

Uudelleenmuotoiluihin liittyviä korjauksia esiintyi kaikkein eniten (46 %) lyhimpiä puhunnoksia tuottaneiden edistyneiden puhujien, toiseksi eniten lyhimpiä puhunnoksia tuottaneiden keskitasoisten puhenäytteissä (39 %). Keskimäärin pisimpiä puhunnoksia tuottaneiden keskitasoisten puhujien korjauksista 34 % ja pisimpiä puhunnoksia tuottaneiden edistyneen tason puhujien korjauksista 19 % liittyi uudelleenmuotoiluihin.

Tehokkaaseen viestintään liittyviä korjauksia esiintyi vähiten keskipituudeltaan lyhimpiä puhunnoksia tuottaneilla keskitason (22 %) ja lyhimpiä puhunnoksia tuottaneilla edistyneen tason (35 %) puhujilla. Pisimpiä puhunnoksia tuottaneiden puheessa tehokkaaseen viestintään liittyvät korjaukset olivat yleisempiä (keskitason ryhmässä 44 %, edistyneiden ryhmässä 53 %).

Tuloksista selvisi alkuperäisen hypoteesin mukaisesti myös, että taukojen ja korjausten välillä on yhteys. Samaan luokkaan kuuluvia taukoja ja korjauksia esiintyi usein yhdessä niin lauseiden välillä kuin niiden keskelläkin, ja taukojen korvaaminen

korjauksilla oli tavallista niin edistyneen kuin keskitasonkin puhujien puhenäytteissä, erityisesti keskimäärin pisimpiä puhunnoksia tuottaneiden puhujien tapauksessa. Taukojen korvaaminen korjauksilla oli yleistä varsinkin leksikaalisen ja rakenteellisen haun sekä tehokkaan viestinnän yhteydessä.

Tämän tutkimuksen tuloksilla voidaan ajatella olevan mahdollisia implikaatioita taukojen ja korjauksien operationalisoinnille kielitesteissä. Lisätutkimuksia kuitenkin tarvitaan. Tutkimuksessa käytetty Nakatsuharan, Tavakolin ja Awwadin (2019) typologia ei ole täydellinen, mutta se on esimerkki tällaisesta toisen kielen puheen taukojen ja korjausten käyttötapojen operationalisoinnista. Mitä tämän tutkimuksen arviointiin tulee, tutkimuksen validiteetti ja reliabiliteetti ovat väistämättä sidonnaisia Yleisen kielitutkinnon puhumisen osakokeen autenttisuuteen, validiteettiin ja reliabiliteettiin. Puhenäytteistä on tehty päätelmiä jälkikäteen, ei puhehetkellä tapahtuvien kognitiivisten prosessien perusteella. Tutkimuksessa käytettyjen YKI-kokeen puhenäytteiden arvioijilla on saattanut olla käytettävissään esimerkiksi arviointirubriikkeja tai muuta materiaalia, jota tähän tutkimukseen ei ollut saatavilla. Lisäksi tutkimuksessa käytetty Nakatsuharan, Tavakolin ja Awwadin (2019) typologia ei ole täydellinen, mutta se on esimerkki toisen kielen puheen taukojen ja korjausten käyttötapojen operationalisoinnista. Typologian kategorioiden rajat eivät ole täysin selkeitä, vaan analyysissa esiintyy tiettyä päällekkäisyyttä. Toisen kielen puhujien taukojen ja korjausten kategorisointi *post hoc* on myös väistämättä tulkinnanvaraista. Tutkimuksen sisäistä reliabiliteettia voitaisiinkin lisätä käyttämällä useita arvioijia osallistujien taukojen ja korjausten analyysivaiheessa, jolloin saataisiin myös luotettavampi käsitys toisen kielen puhujien taukojen ja korjausten käyttötapojen vaikutuksesta sekä kognitiiviseen että havaittuun sujuvuuteen (perceived fluency).

Lopuksi

Tässä tutkimuksessa käsiteltiin englantia toisena kielenä puhuvien suomenkielisten aikuisten toisen kielen puheen sujuvuutta niin määrällisin kuin laadullisinkin menetelmin. Ensimmäinen tutkimuskysymykseni oli, mitkä sujuvuuden mittarit erottavat keskitason ja edistyneen tason puhujat Yleisen kielitutkinnon kokeissa. Tilastollisesti merkittäviä eroja havaittiin erityisesti puhunnosten keskipituudessa, mutta myös kaikilla muilla mittareilla paitsi korjausten kokonaistiheydessä. Toiseksi halusin selvittää, miten taukojen ja korjausten käyttötavat poikkeavat Yleisen kielitutkinnon keskitason ja edistyneen tason puhujien välillä. Tuloksista ilmeni, että keskitason

puhujien tauot ja korjaukset painottuvat kieliopillisiin seikkoihin, edistyneiden taas tehokkaaseen puhuttuun viestintään. Kolmanneksi hain vastausta siihen, onko taukojen ja korjausten välillä yhteyttä hypoteesini mukaisesti. Tutkimuksen tulokset tukevat myös tätä hypoteesia: analysoiduissa puhenäytteissä samaan luokkaan kuuluvia taukoja ja korjauksia esiintyi usein samassa ympäristössä, puhujat käyttivät taukojen välttämiseen tai lyhentämiseen korjauksia, erityisesti toistoja, ja tehokkaaseen viestintään liittyvät tauot ja korjaukset korreloivat vahvasti.

Vaikka tutkimuksen rajoitteiden vuoksi tuloksia ei välttämättä voi yleistää, ne ovat osa laajempaa ja kehittyvää sujuvuustutkimusten joukkoa. Tulokset paitsi tukevat aiempia tutkimuksia, myös viitoittavat tietä mahdollisille jatkotutkimuksille. Sujuvuuden monipuolinen tarkastelu on erityisen tärkeää kielitestauksen näkökulmasta. Tutkimuksen tulosten perusteella toisen kielen puhujien taukojen ja korjausten käyttöön tulisikin kiinnittää huomiota monipuolisemmin. Tulokset herättävät myös pohtimaan, voisivatko jotkin tauot ja korjaukset johtua toisen kielen puhujien puheen prosodian epätäydellisestä proseduralisaatiosta. Rajoitteistaan huolimatta tutkimus joka tapauksessa antaa aihetta puhutun kielen taukojen ja korjausten käytön tarkempaan tutkimukseen sekä niiden sisällyttämiseen toisen kielen puheen sujuvuuden määritelmiin ja operationalisointeihin.