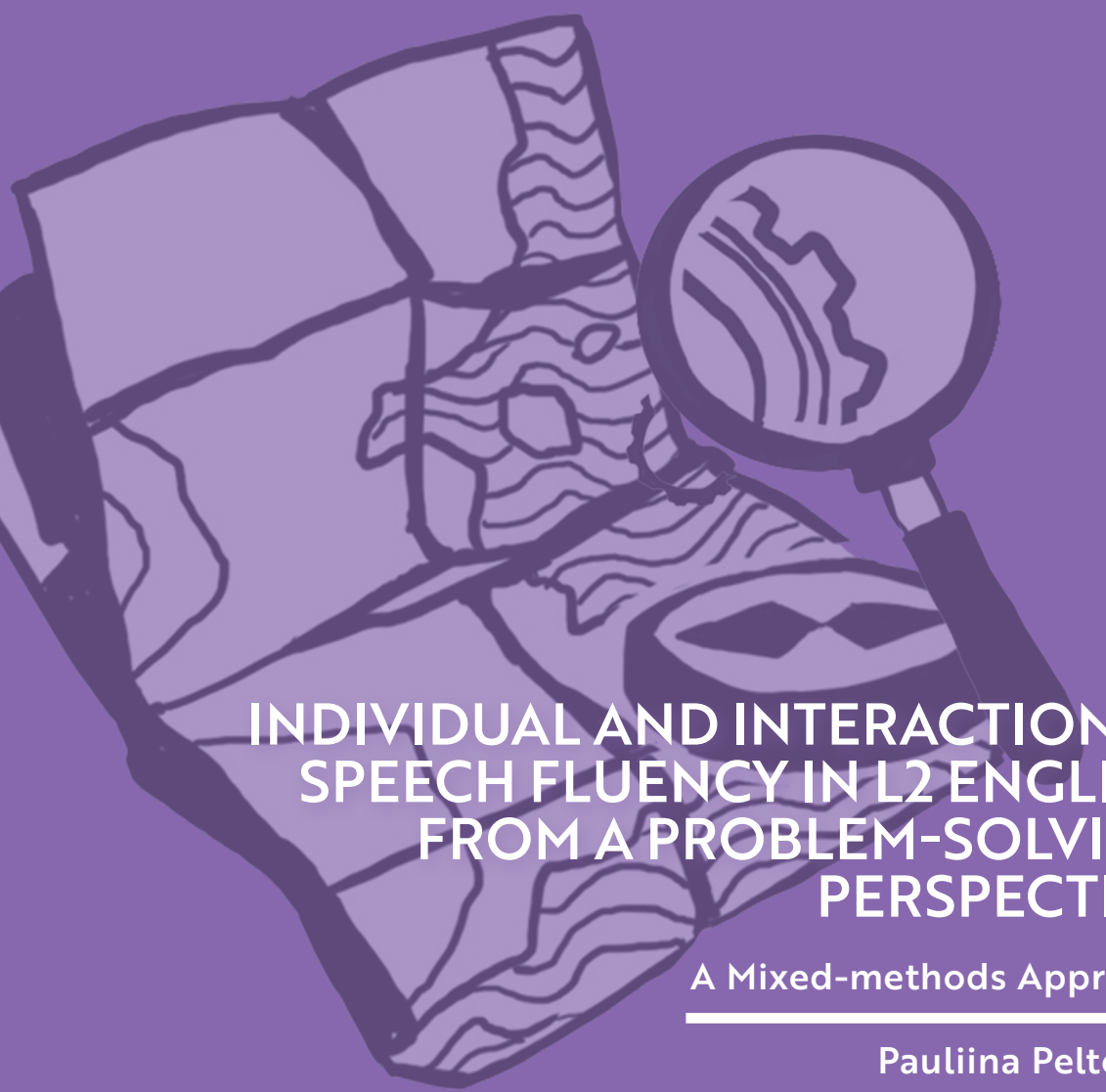




UNIVERSITY
OF TURKU

A stylized illustration in white and light purple on a purple background. It depicts a map with various geographical features like mountains and rivers. A magnifying glass is positioned over the map, focusing on a specific area. The magnifying glass's lens shows a detailed view of a mountain range. The handle of the magnifying glass is also visible.

INDIVIDUAL AND INTERACTIONAL SPEECH FLUENCY IN L2 ENGLISH FROM A PROBLEM-SOLVING PERSPECTIVE

A Mixed-methods Approach

Pauliina Peltonen



UNIVERSITY
OF TURKU

INDIVIDUAL AND INTERACTIONAL SPEECH FLUENCY IN L2 ENGLISH FROM A PROBLEM-SOLVING PERSPECTIVE

A Mixed-methods Approach

Pauliina Peltonen

University of Turku

Faculty of Humanities
School of Languages and Translation Studies
Department of English
Doctoral programme in Languages and Translation Studies (Utuling)

Supervised by

Dr. Pekka Lintunen
University of Turku, Finland

Reviewed by

Professor Judit Kormos
Lancaster University, UK

Professor Parvaneh Tavakoli
University of Reading, UK

Opponent

Professor Judit Kormos
Lancaster University, UK

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

Cover design by Mikko Tirronen, cover illustration by Ella Kurki

ISBN 978-951-29-8136-6 (PRINT)
ISBN 978-951-29-8137-3 (PDF)
ISSN 0082-6987 (Print)
ISSN 2343-3191 (Online)
Painosalama Oy, Turku, Finland 2020

UNIVERSITY OF TURKU

Faculty of Humanities

School of Languages and Translation Studies

Department of English

PAULIINA PELTONEN: Individual and Interactional Speech Fluency in L2

English from a Problem-solving Perspective: A Mixed-methods Approach

Doctoral Dissertation, xii + 230 pp.

Doctoral Programme in Languages and Translation Studies (Utuling)

August 2020

ABSTRACT

This study examines second language (L2) speech fluency and resources for maintaining it based on monologue and dialogue data. The study extends previous L2 speech fluency research by making three theoretical–methodological contributions: adopting a mixed-methods approach (including quantitative and qualitative analyses), developing a problem-solving framework for fluency analysis that incorporates *fluency resources* (stalling mechanisms and communication strategies), and measuring and assessing fluency based on L2 dialogue data. The study is situated within L2 oral proficiency and second language acquisition research and has implications for L2 teaching and assessment.

The study is based on four original publications (Articles I–IV). In the articles, Finnish learners’ monologue and dialogue speech fluency is examined with two data sets including L2 English samples from ninth graders, upper secondary school students, and university students of English. L1 Finnish samples are also examined in Article III. Fluency is analyzed using measures related to temporal (individual) fluency, interactional fluency, and problem-solving mechanisms, with an added perceived fluency perspective (raters’ assessments) in Article IV.

The results demonstrate that the qualitative analysis complements the quantitative analysis in various ways, notably highlighting the role of stalling mechanisms in compensating for local disfluencies. Based on the analyses, more advanced students also speak more fluently. L2 fluency is also linked to some extent to L1 fluency. In L2 dialogue data, (between-)turn pause length in particular distinguishes between learners at different levels. Based on the study, interactional fluency can be viewed as consisting of two main components: temporal fluency (turn pauses) and cohesive devices (other-repetitions and collaborative completions).

The study demonstrates the benefits of triangulating quantitative and qualitative methods in fluency research and applying a problem-solving perspective to fluency analysis. The approach to interactional fluency that is developed in the study can also be used as a basis for further research. Based on the results, the influence of L1 speaking style on L2 fluency and the joint nature of fluency in L2 dialogue contexts should be more widely acknowledged in L2 assessment and teaching.

KEYWORDS: fluency, interaction, mixed-methods, oral proficiency, problem-solving, second language acquisition

TURUN YLIOPISTO

Humanistinen tiedekunta

Kieli- ja käännöstieteiden laitos

Englannin kieli

PAULIINA PELTONEN: Individual and Interactional Speech Fluency in L2

English from a Problem-solving Perspective: A Mixed-methods Approach

Väitöskirja, xii + 230 s.

Tohtoriohjelma Utuling

Elokuu 2020

TIIVISTELMÄ

Väitöstutkimukseni käsittelee vieraskielisen puheen sujuvuutta ja sujuvuuden ylläpitämistä monologi- ja dialogiaineistoissa. Tutkimus laajentaa aiempaa sujuvuustutkimusta kolmella teoreettis–metodologisella kontribuutiolla: yhdistämällä laadullisia ja määrällisiä menetelmiä, kehittämällä *sujuvuuden resurssit* (viivytyksmekanismit ja kommunikaatiostrategiat) huomioivan ongelmanratkaisukeinojen viitekehysten sekä mittaamalla ja arvioimalla sujuvuutta dialogiaineistosta. Suullisen kielitaidon ja vieraan kielen oppimisen tutkimuksen aloille sijoittuvalla tutkimuksella on implikaatioita erityisesti kielten opetukselle ja arvioinnille.

Tutkimus koostuu neljästä alkuperäisjulkaisusta (Artikkelit I–IV). Artikkeleissa tutkin suomalaisten englannin oppijoiden puheen sujuvuutta monologeissa ja dialogeissa kahdella aineistolla. Tutkimusaineistot sisältävät englanninkielisiä puhenäytteitä yhdeksäsluokkalaisilta, lukion opiskelijoilta ja englannin yliopisto-opiskelijoilta. Artikkelissa III tarkastelen myös suomenkielisiä puhenäytteitä. Analysoin sujuvuutta temporaalisen (yksilön) ja vuorovaikutuksen sujuvuuden mittareilla sekä ongelmanratkaisukeinoihin perustuen. Artikkelissa IV analysoin sujuvuutta myös havaitun sujuvuuden eli kuulijoiden arvioiden näkökulmasta.

Tulokset osoittavat, että laadullinen analyysi täydentää määrällistä analyysistä monin tavoin, erityisesti korostamalla viivytyksmekanismien merkitystä epäsujuvuuksien kompensoinnissa. Analyysien perusteella edistyneemmät oppijat myös puhuvat sujuvammin. Vieraskielisen puheen sujuvuus liittyy kuitenkin osittain äidinkielen sujuvuuteen. Dialogeissa erityisesti vuorojen välisten taukojen pituus erottelee eritasoisia oppijoita. Tutkimuksen perusteella vuorovaikutuksen sujuvuus koostuu temporaalisesta sujuvuudesta (vuorojen välisistä tauoista) sekä koheesiota luovista elementeistä (parin puheen toistamisesta ja vuorojen täydentämisestä).

Tutkimus osoittaa monimenetelmäisen lähestymistavan sekä ongelmanratkaisunäkökulman hyödyllisyyden sujuvuuden analyysissä, ja tutkimuksessa kehitetty lähestymistapa vuorovaikutuksen sujuvuuteen soveltuu pohjaksi jatkotutkimukselle. Tulosten perusteella kielten opetuksessa ja arvioinnissa tulisi huomioida äidinkielen vaikutus vieraan kielen sujuvuuteen sekä sujuvuuden vuorovaikutteinen luonne dialogikonteksteissa.

ASIASANAT: monimenetelmätutkimus, ongelmanratkaisu, sujuvuus, suullinen kielitaito, vieraan kielen oppiminen, vuorovaikutus

Acknowledgments

During the period of writing this dissertation, I have been extremely fortunate in receiving the help and support of many people. I would like to start by thanking the external examiners of this PhD dissertation, Professor Judit Kormos from Lancaster University and Professor Parvaneh Tavakoli from the University of Reading. I am greatly indebted to them for their constructive, detailed, and encouraging reviews, which helped me clarify my own thinking and improve the argumentation in the dissertation during the finalization process. I also wish to express my gratitude to Professor Kormos for agreeing to act as my opponent in the public examination of the dissertation.

My gratitude to my supervisor, Dr. Pekka Lintunen, goes beyond what can be expressed here; I could not have asked for a more encouraging and dedicated supervisor. Not only has Pekka helped me with the PhD writing process by providing critical and constructive feedback, but he has also consistently supported my growth as a researcher by providing guidance and opportunities to participate in various academic activities. Thanks to Pekka, I feel confident of taking the next steps toward postdoctoral research. My second supervisor, the late Professor Päivi Pietilä, tragically passed away before the completion of this PhD. From the very beginning of my advanced studies for my master's degree and the initial stages of my doctoral studies, Päivi's research, teaching, and feedback have shaped my work and inspired me, and will continue to do so in every step moving forward. She was a most wonderful and supporting supervisor, and I am grateful to Päivi for all her encouragement and assistance.

While conducting this PhD study, especially during the data collection, preparation, and analysis phases, I received help from numerous people, and I consider myself very lucky for this. I would first like to thank the school principals for granting me permission to organize data collection in their schools, the English teachers for their assistance with the practical matters and for dedicating their time for this project, and the students for participating in this study. Many thanks are also in order for Ella Kurki for creating the illustrations for the problem-solving and picture description tasks used in the data collection. Additionally, I am grateful to research assistants Emily Green, Tytti Kim, and Tomi Rautaoja for helping me with

data collection and transcription. I would also like to thank Jouko Katajisto for being available for consultations regarding the statistical analyses conducted for the substudies.

Being able to collaborate with different people has greatly supported the writing of this dissertation and provided the important balance with the lengthy periods of solitary work. I am grateful to Professor Sandra Götz for her encouragement and for supporting my work; in particular, I would like to thank her for providing valuable comments on the manuscripts for Articles III and IV included in this dissertation. Many thanks to Dr. Sanna Olkkonen: our collaboration has expanded my thinking on fluency and greatly helped in writing this PhD dissertation. Dr. Maarit Mutta has supported my work in different roles throughout my years as a PhD student; I would particularly like to thank her for the enjoyable collaboration on fluency-related projects and publications. I am also grateful to Associate Professor Sanna-Kaisa Tanskanen from the University of Helsinki for providing me with the opportunity to join the teaching team for the course *L2 Spoken Interaction* in 2018 and 2019; teaching a course with a topic so closely related to this PhD greatly facilitated the writing process. Thank you also to Lasse Ehrnrooth for making the team-teaching experience so enjoyable.

The Department of English at the University of Turku provides a wonderful working environment for a PhD student; I would like to thank the whole staff collectively for creating such a supportive atmosphere. The second language acquisition (SLA) colloquium, organized by Päivi, played an important role in providing collegial support during my PhD studies. I would like to thank the colloquium members, my fellow SLA PhD students, for the fruitful discussions and useful feedback on my article drafts. Dr. Aleksi Mäkilähde: thank you for sharing various office spaces with me throughout these years, for all the discussions, support, and your friendship. Many thanks also to the other Philological Colloquium members of the English Department, namely Dr. Aino Liira, Dr. Hanna Salmi, Johanna Rastas, Dr. Mari-Liisa Varila, Sara Norja, Sirkku Ruokkeinen, and Susanna Mäkinen for the much-needed peer support at lunch and coffee breaks as well as for all the fun escape room sessions.

I am thankful to all the people who have commented on my work on various occasions at conferences, seminars, and other events in Finland and abroad—being able to discuss the work in progress has been invaluable to me in the writing process. Thank you to Kaisa Korpinen for all the inspiring lunch meetings—I am happy that our paths crossed in Louvain-la-Neuve in 2017. The Utuling thematic group for language learning and the Centre for Language Learning Research (Leala) also deserve thanks for providing forums for receiving feedback and discussing SLA matters across the different language departments at the University of Turku.

I gratefully acknowledge the funding provided by the Doctoral Programme Utuling of the University of Turku Graduate School from 2014 to 2017: I consider myself extremely fortunate having had the opportunity to work on my dissertation full-time for four years as a salaried Utuling doctoral candidate. In addition, I am grateful to the Alfred Kordelin Foundation for providing a 6-month grant that enabled me to finalize this dissertation. I would also like to thank the Finnish Association for Applied Linguistics (AFinLA), Langnet, Utuling, Turku University Foundation, and the English Department for providing financial support for attending conferences.

Thank you also to the proofreader at Scribendi for checking the language of the compilation part of this dissertation; all remaining errors are my own.

My friends and family have supported me throughout this PhD process and provided much-needed breaks from work. Thank you to my dear friend Heidi: I am extremely lucky to have a friend with whom I have been able to share the ups and downs of PhD life and also to forget all about academia at times and to focus on other important things. My parents: thank you for your support and encouragement in every step of my academic journey.

To my boyfriend and partner-in-crime, Mikko: thank you for sticking by me all these years. Your support means the world to me.

In Turku, August 2020
Pauliina Peltonen

Table of Contents

Acknowledgments	v
Table of Contents	viii
List of Original Publications	xi
Abbreviations	xii
1 Introduction	1
1.1 Background of the study.....	1
1.2 Research aims and outline of the dissertation	3
2 Fluency as an aspect of second language (L2) oral proficiency	8
2.1 L2 speech fluency definitions	9
2.2 L2 speech fluency measurements.....	12
2.3 Factors affecting L2 speech fluency	18
2.3.1 General L2 proficiency	18
2.3.2 L1 speech fluency	21
3 Broadening the scope of L2 speech fluency analysis	23
3.1 A mixed-methods approach	23
3.2 Combining fluency and problem-solving: The Fluency Resources Framework	24
3.3 From individual to interactional fluency.....	29
4 Data and methods	36
4.1 Participants	36
4.2 Data and procedure	40
4.3 Measures	46
4.4 Analyses	53
4.4.1 Quantitative analyses.....	53
4.4.2 The mixed methodology and qualitative analyses	55
5 Results	58
5.1 Results of Article I	60
5.2 Results of Article II	62
5.3 Results of Article III	67
5.4 Results of Article IV.....	71

5.5	Summary of the results	73
6	Discussion.....	77
6.1	Theoretical and methodological considerations for L2 fluency research.....	77
6.1.1	Qualitative analysis	78
6.1.2	A problem-solving perspective on L2 speech fluency ..	81
6.1.3	Interactional fluency	85
6.1.4	A classification of L2 speech fluency components from a problem-solving perspective	87
6.2	Limitations and future directions	92
6.3	Implications for L2 assessment and teaching	96
6.4	Concluding remarks	100
	List of References.....	103
	Appendices	116
	Appendix 1: Monologue picture prompts	116
	Appendix 2: Dialogue task instructions	117
	Appendix 3: Rating sheet	120
	Original Publications	125

Tables

Table 1.	The research aims and their connections to the original publications.	5
Table 2.	Fluency measures, their operationalizations, their use in different articles, and links to Skehan's dimensions and the Fluency Resources Framework categories.	47
Table 3.	Focus of the articles and connections to the research aims.	59
Table 4.	Statistically significant differences across the groups based on a one-way ANOVA in Article I (for all comparisons, see Table 2 in Article I, p. 223).	60
Table 5.	Statistically significant differences between the groups based on Mann-Whitney U-tests in Article II (for all comparisons, see Table 1 in Article II, p. 6).	63
Table 6.	Statistically significant correlations across problem-solving mechanisms and temporal/interactional fluency in Article II.	65
Table 7.	Simple regression models predicting L2 fluency measures based on L1 fluency measures in Article III (Table 5 in Article III, p. 685).	69

Figures

Figure 1.	The traditional approaches, gaps, and the present study's contributions to L2 speech fluency research.	3
Figure 2.	Skehan's fluency dimensions and common L2 speech fluency measures.	13
Figure 3.	The Fluency Resources Framework (originally published in Article II, p. 3).	26
Figure 4.	Phases of data collection (Articles II–IV).	42
Figure 5.	Main results based on the three research aims and their contributions to the general aim.	78
Figure 6.	Components and measures of L2 speech fluency from a problem-solving perspective in monologue and dialogue.	90

List of Original Publications

This dissertation is based on four original publications. The original publications are referenced in the dissertation as follows:

- Article I Peltonen, P., & Lintunen, P. (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, 209–238.
- Article II Peltonen, P. (2017). Temporal fluency and problem-solving in interaction: An exploratory study of fluency resources in L2 dialogue. *System*, 70, 1–13.
- Article III Peltonen, P. (2018). Exploring connections between first and second language fluency: A mixed methods approach. *The Modern Language Journal*, 102, 676–692.
- Article IV Peltonen, P. Connections between measured and assessed fluency in L2 peer interaction: A problem-solving perspective. Manuscript.

The original publications have been reproduced with the permission of the copyright holders.

Abbreviations

ANOVA	Analysis of variance
AR	Articulation rate
AS-unit	Analysis of Speech Unit
CA	Conversation analysis
CAF	Complexity, accuracy, and fluency
CEFR	The Common European Framework of Reference for Languages
CLIL	Content and language integrated learning
CS	Communication strategy
FDR	False discovery rate
FP	Filled pause
IC	Interactional competence
L1	First language
L2	Second language
LexTALE	Lexical Test for Advanced Learners of English
MLR	Mean length of run
MLT	Mean length of turn
NNS	Non-native speaker
NS	Native speaker
PTR	Phonation–time ratio
SA	Study abroad
SLA	Second language acquisition
SP	Silent pause
SR	Speech rate
TP	Turn pause

1 Introduction

One of the most important goals for second language¹ (L2) learners is to be able to communicate efficiently in the L2. To ensure efficient communication, learners should be able to speak with relative ease and effortlessness. When encountering different types of problems during communication, learners also need various resources for maintaining fluent speech: For instance, when forgetting a word in the target language, a paraphrase can be used instead of the target word. This doctoral dissertation examines L2 speakers' communication from these perspectives: the fluency of L2 speech and how well speech fluency is maintained. The focus is on L2 learning; yet, the topic is particularly timely and interesting from the perspectives of L2 teaching and assessment. In the past decade, the teaching of L2 speaking skills, including fluency, has received more emphasis in Finland, especially in upper secondary schools, with the introduction of a national course focusing on spoken language in 2010. Moreover, spoken components of the exams for language subjects that are included in the national school-leaving exam, the Matriculation Examination, are currently under preparation (Ministry of Education and Culture, 2017). By focusing on fluency, the present study provides insights into the skills related to maintaining the flow of speech across learner groups. The background of the study will be presented in more detail in the following section.

1.1 Background of the study

In the present study, fluency is understood as one component of L2 oral proficiency that refers to the smoothness and effortlessness of speech; it can be captured by measuring aspects such as speech rate (SR) and pausing from speech samples (e.g., Chambers, 1997). While there is a relatively long tradition of studying speech

¹ The term *second language* is used in a broad sense to refer to any language learned after the first language(s) (not necessarily the second language chronologically). The target L2 in the present study is English. However, considering the context of the present study, English is technically learned as a foreign language, rather than as a second language (cf. Swedish), in Finland, although the language is also widely used outside the classroom. (For further discussion, see Leppänen *et al.*, 2011.)

fluency, starting with Goldman-Eisler's (1968) pausological studies on first language (L1) speech in the 1950s and 1960s and expanding to temporal aspects of L2 speech in the 1980s (e.g., Raupach, 1980), it was not until the seminal studies by Lennon (1990) and Riggenbach (1991), among others, in the early 1990s that fluency as an aspect of L2 oral proficiency emerged as a systematic object of study. In the 21st century, with increasingly larger samples of speech data being examined with partially automatized analyses enabled by recent technological advancements, fluency has become a popular topic among researchers working in the fields of task-based research, second language acquisition (SLA), and L2 assessment, in particular. With more than 30 years of research on L2 speech fluency, researchers have recently started to look back at the research traditions (e.g., De Jong, 2018) while also contemplating the future directions of L2 speech fluency research (Foster, 2020; Lintunen, Mutta, & Peltonen, 2020b; Wright & Tavakoli, 2016).

Some of the challenges related to L2 fluency research, which are also addressed in the present study, relate to the theoretical and methodological approaches that are applied in empirical studies. Most L2 speech fluency research is quantitative in nature, aimed at discovering broad differences in fluency—for instance, among learners at different proficiency levels. The limitations of the quantitative approach have been acknowledged especially in the calls to expand fluency research from monologue to dialogue contexts (Segalowitz, 2016). While the majority of fluency research has focused on individual speakers' fluency, researchers have recently emphasized the importance of acknowledging the communicative and social aspects of fluency (Segalowitz, 2016; Wright & Tavakoli, 2016). These developments echo broader issues in SLA, with researchers drawing attention to the fact that L2 learning and competence should be approached not only from a cognitive perspective but also from a social one (Atkinson, 2019; Douglas Fir Group, 2016; Ortega, 2019).²

As fluency has traditionally been defined from a cognitive–individualistic perspective, expanding the scope of fluency research from L2 monologue to dialogue contexts poses challenges for definition and measurement. In the present study, a novel framework for monologue and dialogue fluency analysis that is suitable for both quantitative and qualitative approaches to L2 speech fluency will be presented (see Section 6.1.4). As such, the present study is among the first to systematically operationalize *interactional fluency* (referred to as *dialogue fluency* in Article II) in addition to examining individual aspects of fluency with more widely established measures. Interactional fluency refers to the co-construction of flow across turn boundaries (following McCarthy's 2010 definition of *confluence*; see further Section 3.3). In the following section, the general aim and the three specific research aims of

² See also Clark, 1996, on L1 use and Roever and Kasper, 2018, for a similar argument concerning L2 assessment.

the present study will be outlined, followed by brief introductions to the four original publications included in this dissertation (Articles I–IV). Section 1.2 concludes with an outline of the structure of the dissertation.

1.2 Research aims and outline of the dissertation

The general aim of the present study is to examine how learners maintain speech fluency in L2 English monologue and dialogue, using a mixed-methods approach. Related to different aspects of the general aim, the three central theoretical–methodological contributions of the dissertation are summarized in Figure 1. The contributions fill gaps in the research by a) broadening the methodological repertoire in L2 speech fluency research using a mixed-methods approach that involves complementing quantitative analysis with qualitative analysis; b) expanding the theoretical framework for studying L2 speech fluency with a problem-solving perspective (see Dörnyei & Kormos, 1998) that focuses on speakers’ resources for maintaining speech fluency (*the Fluency Resources Framework*); and c) expanding the scope of L2 speech fluency analysis to include L2 dialogic data, which enables the study of both individual and interactional fluency. These contributions will be presented in more detail in the following.

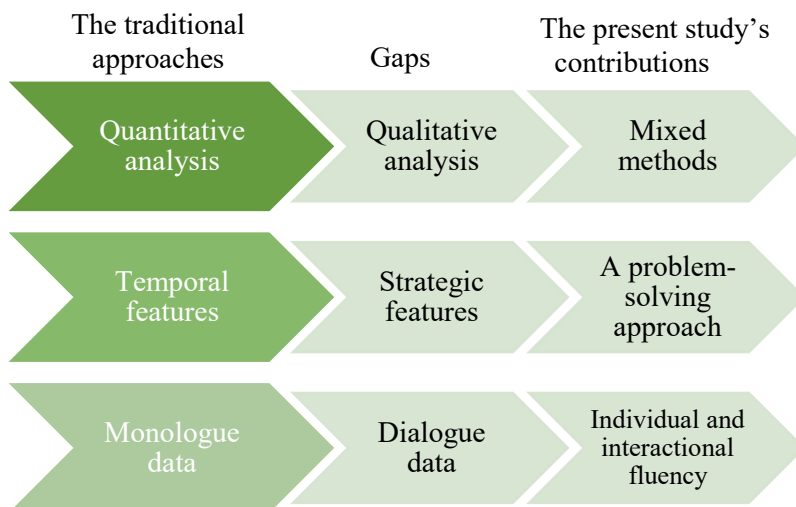


Figure 1. The traditional approaches, gaps, and the present study's contributions to L2 speech fluency research.

There is a need to broaden the methodological approaches to L2 speech fluency analysis: The mainstream quantitative approach, which is based on quantifying (dis)fluency features, is insufficient for comprehensively capturing as complex a phenomenon as L2 speech fluency. In particular, the quantitative analysis of relatively infrequently occurring, potentially multifunctional features, such as repetitions, filled pauses (FPs—e.g., *uh* and *um*), and fillers (e.g., *you know*), should be complemented with a qualitative analysis that focuses on the functions and contexts of these features and highlights individual differences in their use among speakers (for details, see Section 3.1). By triangulating these complementary methodological perspectives in a mixed-methods analysis, a more comprehensive picture of fluency and means for maintaining fluency can be achieved.

Related to the mixed-methods approach, the novel framework introduced in the dissertation acknowledges the multifunctional nature of certain fluency-related features and individual differences in their use (see also Degand, Gilquin, Meurant, & Simon, 2019; Dumont, 2018; Lennon, 2000). In the framework, the so-called disfluency features are viewed as potentially fluency-enhancing features that help in maintaining the flow of speech. Based on a problem-solving approach presented in Dörnyei and Kormos (1998) and adapted to the context of L2 fluency research, the framework combines fluency analysis with the analysis of strategic language use (*stalling mechanisms* and *communication strategies* [CSs]; see Section 3.2).

Finally, L2 fluency research has mostly focused on the analysis of monologic speech data, while few researchers have examined fluency in L2 interaction. To date, dialogue data have mainly been used to compare performance in monologue and dialogue contexts with established individual fluency measures, which have been adapted to dialogue contexts (e.g., Tavakoli, 2016, 2018; Witton-Davies, 2014). Building on and extending the focus of previous studies, in this dissertation, dialogue data are also analyzed from the perspective of *interaction*—the aspect that distinguishes monologic and dialogic performance. More specifically, collaboration between participants in maintaining the flow across turns is explored based on the notion of interactional fluency. In defining, measuring, and suggesting ways of assessing interactional fluency, the present study draws on research on *interactional competence* (IC) conducted in the CA-SLA tradition (conversation analysis [CA] as applied to L2 interaction; for a recent overview, see Salaberry & Kunitz, 2019) and in L2 assessment (Galaczi & Taylor, 2018; for details, see Section 3.3).

Based on the three gaps in research and the present study's contributions to filling these gaps, three main research aims were formulated. The research aims, including their connections to the original publications, are presented in Table 1. (For more specific foci of each publication, see Table 3 in Chapter 5.)

Table 1. The research aims and their connections to the original publications.

Research aims	Article I	Article II	Article III	Article IV
1) <i>To examine how a qualitative analysis contributes to illustrating differences in fluency across learners and complements a quantitative approach</i>	x	x	x	(x)
2) <i>To examine what a problem-solving framework can reveal about</i>				
a) <i>learners' speech fluency and resources for maintaining fluency at different school levels,</i>	(x)	x	x	
b) <i>the connections between learners' speech fluency and resources for maintaining fluency, and</i>		x		x
c) <i>the connections between L1 and L2 speech fluency</i>			x	
3) <i>To examine how fluency can be analyzed based on dialogue data from assessment and measurement perspectives</i>		x		x

As can be seen from Table 1, the first research aim relates to exploring the contribution of the qualitative approach. This aim is addressed in all four original publications. The second research aim relates to applying the problem-solving framework to fluency analysis and contains three aspects. More specifically, differences in speech fluency and resources for maintaining fluency between learners at different school levels (Article II in particular and, to some extent, Article III as well), the connections between learners' speech fluency and resources for maintaining fluency (Articles II and IV), and the connections between L1 and L2 speech fluency (Article III) are explored from a problem-solving perspective. Groups at different school levels are also compared in Article I, without explicit reference to the problem-solving framework. In Article IV, the problem-solving framework is applied to fluency assessment. Finally, the third research aim relates to the analysis of dialogue data and explores how fluency can be operationalized in an interactional context from the measurement and assessment perspectives. This aim is explored in the two articles using L2 dialogue data (Articles II and IV). The foci of the four original publications from the perspective of the main research aims will be presented in more detail below.

In **Article I**, the effects of the L1 (Swedish and Finnish) and school level (upper secondary school and university) on L2 speech fluency in English were examined. Comparisons across learners at different school levels were also included in Articles

II and III (aim 2a). Methodologically, Article I provides the basis for the other articles by introducing a mixed-methods approach to L2 speech fluency analysis as a solution for overcoming the limitations of a purely frequency-based analysis (aim 1). Thus, Article I also forms the foundation for the Fluency Resources Framework: Because some (dis)fluency-related features are multifunctional and can be used to enhance fluency, a qualitative analysis is essential for examining these features.

For Articles II–IV, a new data set was collected (see Section 4.2) to enable the further study of aspects identified as promising in Article I—notably the interactional aspect of fluency and the connections between fluency and strategic language use. **Article II** focused on the effect of school level on L2 fluency by comparing learners from the ninth grade (last year of compulsory education) and upper secondary school (aim 2a). That is, Article II complements the information about Finnish learners' fluency in L2 English in upper secondary school and university (Article I) by including a learner group that represents a lower proficiency level (ninth graders). From a theoretical–methodological perspective, Article II introduces a new framework to L2 speech fluency analysis by developing the argumentation in Article I further. In the Fluency Resources Framework, fluency is approached from a strategic, problem-solving perspective (see Dörnyei & Kormos, 1998), including an analysis of two types of so-called fluency resources: stalling mechanisms and CSs (aim 2b). In Article II, fluency was also studied in a new context (L2 dialogue) based on a paired problem-solving task. With the new framework and the interactional context, it was possible to examine how fluency can be maintained collaboratively (aim 3). Furthermore, the mixed-methods approach introduced in Article I was applied to dialogue and linked to the new problem-solving framework for the first time in Article II (aim 1).

In **Article III**, the problem-solving framework was applied to a current topic in L2 speech fluency research: How is the individual speaking style in the L1 connected to L2 speech fluency (aim 2c)? The connections between L1 and L2 fluency were examined from comparable monologue tasks in the L1 (Finnish) and L2 (English) produced by the same speakers. The participants' school levels were also taken into account in the analyses (aim 2a). In Article III, the Fluency Resources Framework was applied in a modified format: CSs, which were not used as frequently in the monologic data as in the dialogic data, were excluded. The framework was used to examine the extent to which the resources for maintaining fluency in the L2 relate to L1 fluency patterns, demonstrating that the framework can also be applied to L1 speech fluency analysis. Furthermore, the mixed-methods approach was applied in Article III to highlight individual differences and individual means for maintaining fluency in the L1 and L2 (aim 1).

In **Article IV**, a subset of the dialogue samples analyzed in Article II were used to examine how well fluency is maintained in L2 interaction from an assessment

perspective (listener's perspective, or *perceived fluency*; Segalowitz, 2010). Six dialogue samples were assessed by 15 raters from the perspectives of oral proficiency, individual fluency, interactional fluency, and strategic competence, mirroring the dimensions measured from an utterance fluency perspective with the Fluency Resources Framework in Article II, including temporal fluency, stalling mechanisms, and CSs (aim 2b). The main focus of the article was to examine the connections between the ratings and fluency measures (aim 3). Using a qualitative analysis, the study also examined which aspects of interaction were salient for the listener (individual, interactional, and/or strategic). Article IV thus also employed a mixed-methods approach (aim 1); yet, in contrast to Articles I–III, which examined utterance fluency using quantitative and qualitative approaches, the qualitative analysis in Article IV concerned the raters' comments and their justifications for their numeric scores. However, the aims of the mixed-methods approach in Article IV were similar to those of Articles I–III: to complement the information provided with a quantitative analysis and to provide insights into individual differences.

The compilation part of this dissertation starts with two theoretical chapters (2 and 3). In Chapter 2, fluency as an aspect of L2 oral proficiency is discussed from three main perspectives: definitions (Section 2.1), measurements (Section 2.2), and two central factors affecting L2 speech fluency, which are addressed in the present study (Section 2.3): general L2 proficiency (Section 2.3.1) and L1 speech fluency (Section 2.3.2). In Chapter 3, the theoretical frameworks related specifically to the contributions of the present study (cf. Figure 1) are discussed. The theoretical chapters are followed by a presentation of the participants, data and methods, and analyses in Chapter 4. The results of each article, as well as a summary of the results, are outlined in Chapter 5. The results are discussed in Chapter 6 (Section 6.1), along with the limitations and future directions (Section 6.2), as well as the implications for L2 assessment and teaching (Section 6.3). Concluding remarks are made in Section 6.4.

2 Fluency as an aspect of second language (L2) oral proficiency

In the present study, fluency is approached as one aspect of L2 oral proficiency. As a concept, fluency is used in various fields, including speech-language pathology (disfluencies in disordered speech, which are excluded from the present study; e.g., Marshall, 2000; Penttilä, 2019), psycholinguistics (notably Goldman-Eisler, 1968; Levelt, 1989; recently, e.g., Olkkonen, 2017), L2 teaching (e.g., Guillot, 1999), and L2 assessment (e.g., Fulcher, 2003), to name a few (for an overview, see De Jong, 2018). Thus, various definitions of fluency exist (e.g., Chambers, 1997; Koponen & Riggenbach, 2000); consequently, researchers from different backgrounds may emphasize different aspects of fluency and associate different features with fluency. Defining fluency is further complicated by its use in a nontechnical sense outside of academic contexts to broadly refer to general (oral) proficiency in the L2 (Lennon, 1990; see Section 2.1 for fluency definitions). The present study focuses on fluency in L2 speech, excluding fluency in writing and receptive skills (fluency in listening and reading; for an overview of fluency and the four skills, see Lintunen, Mutta, & Peltonen, 2020a). Furthermore, other components of L2 proficiency and performance, such as complexity or accuracy (for the complexity, accuracy, and fluency [CAF] framework, see Housen & Kuiken, 2009; Housen, Kuiken, & Vedder, 2012) are excluded from the present study to enable a detailed analysis of fluency, which is in itself multifaceted.

In the following section, fluency will be defined against this background, with a focus on characterizing fluency as one dimension of L2 oral proficiency. In Section 2.2, L2 fluency measurements will be discussed. Section 2.3 consists of two subsections discussing key factors affecting L2 speech fluency, which are examined in the present study: In Section 2.3.1, general L2 proficiency will be discussed, followed by a discussion of L1 speech fluency in Section 2.3.2.

2.1 L2 speech fluency definitions

L2 speech fluency is often defined based on Lennon's (1990) well-established distinction between the *broad* and *narrow* senses of fluency (later referred to as *higher-order* and *lower-order fluency*, respectively; Lennon, 2000). In general language use outside of the fluency research context, the use of fluency in a broad sense is common, with fluency being roughly comparable with overall (oral) language proficiency (Lennon, 1990). Recently, however, Tavakoli and Hunter (2018, p. 343) have suggested a further distinction between *broad* and *very broad* senses of fluency: The very broad sense of fluency refers to a certain level of language proficiency (encompassing multiple subskills in addition to speaking skills), while in the broad sense, fluency is specifically concerned with speaking skills. These broad and very broad senses have also been found to influence raters' assessments of fluency, especially when the raters have not been instructed to focus specifically on fluency in the narrow sense (as a specific component of oral proficiency; e.g., Cucchiari, Strik, & Boves, 2002; Freed, 1995; Kormos & Dénes, 2004; Riggenbach, 1991), and to reflect teachers' (Tavakoli & Hunter, 2018) and language learners' (Lintunen & Peltonen, 2020) understanding of fluency.

In research contexts, however, fluency is usually examined in the specialized, narrow sense, referring to one aspect of oral proficiency that relates to temporal features of speech, such as speed of talk and pausing (Lennon, 1990). Thus, in the narrow sense, fluency is viewed as an isolatable component of oral proficiency and performance, while other aspects, such as accuracy and complexity, are viewed as their own dimensions (e.g., the CAF framework mentioned above). While fluency in the narrow sense is essentially viewed in terms of the smoothness and effortlessness of speech (Chambers, 1997), disfluent speech can be described as uneven, disconnected, or hesitant (Fulcher, 2003, p. 30). In more concrete terms, speech that lacks fluency is characterized by the presence of disfluency features, such as (excessive) pauses or other hesitations (e.g., Lennon, 1990, p. 390; Segalowitz, 2000, p. 200; see also Section 2.2). In other words, L2 speech is considered disfluent if there are "too many" pauses or other hesitations, the implicit point of reference for "too many" generally being L1/native speaker (NS) speech, which presumably contains minimal pausing or hesitation.

The view of native speech as lacking disfluency and NSs as forming a homogenous group of speakers has been criticized by many researchers theoretically (e.g., Fillmore, 1979; Lennon, 1990; Riggenbach, 1991) and proven questionable empirically in studies demonstrating variation in L1 speech fluency and disfluency (Goldman-Eisler, 1968; Maclay & Osgood, 1959; Moniz, Batista, Mata, & Trancoso, 2014). In fact, some L2 fluency studies have used an NS control group to avoid implicit comparisons to (ideal) native speech and to provide a genuine point of comparison to learner speech (e.g., Dumont, 2018; Götz, 2013; Hilton, 2008; Kahng,

2014; Article I). Corroborating studies on the temporal aspects of L1 speech, these studies have also demonstrated variation in NS fluency. Even more recently, some studies have moved beyond using an “external” native group as a point of reference for L2 fluency³ and compared learners’ L2 speech against their own L1 production (De Jong, Groenhout, Schoonen, & Hulstijn, 2015; Duran-Karaoz & Tavakoli, 2020; Huensch & Tracy-Ventura, 2017a; Article III). Studies using L1 and L2 data from the same speakers can illustrate the influence of personal speaking style in the L1 on L2 fluency, along with differences between L1 and L2 fluency, and therefore provide new insights into L2 speech fluency. Both the between-subject (L1 control group) and within-subject (speaker’s own L1) approaches can be regarded as valid, albeit useful for examining slightly different research questions, and are generally preferred over implicit references to native speech in L2 speech fluency research.

In addition to comparisons to NS speech being common in definitions of L2 speech fluency, most fluency definitions are psycholinguistic in nature. For instance, fluency has been described as “purely a performance phenomenon” (Lennon, 1990, p. 391) and as an “automatic procedural skill” (Schmidt, 1992, p. 358), reflecting the idea that smoothly running cognitive processes underlie fluent speech. In a more recent, oft-cited definition by Lennon (2000), fluency is characterized as “the rapid, smooth, accurate, lucid, and efficient translation of thought or communicative intention into language under the temporal constraints of on-line processing” (p. 26). Similar to the earlier characterizations, Lennon’s (2000) definition reflects a psycholinguistic approach to fluency, which emphasizes the translation of underlying cognitive processing and its efficiency into fluent speech. A psycholinguistic approach is also reflected in Segalowitz’s (2010) influential threefold fluency framework (see also Segalowitz, 2016), due to the central role of cognitive fluency. In the framework, *cognitive fluency*, which refers to the speed and efficiency of underlying processing, forms the foundation of *utterance fluency* (the measurable features of speech associated with fluency). Then, in turn, the fluency-related features and their overall effect create an impression of fluency from the listener’s perspective (*perceived fluency*); the listener makes inferences about a person’s fluency based on his or her utterance fluency features. Thus, in the framework, cognitive fluency can be viewed as forming the foundation for other types of fluency (utterance and perceived fluency).

The present study is chiefly concerned with utterance fluency in Segalowitz’s (2010) terminology: Aspects of spoken productions that are associated with fluency and disfluency—especially temporal features, such as speed of talk and pausing—are analyzed (see Section 2.2 for details). Thus, the focus is on the narrow sense of

³ The problems with comparing learners to NSs have been highlighted, for instance, within the field of English as a lingua franca (see, e.g., Hynninen, 2020).

fluency following Lennon's (1990) terminology: Fluency is approached as a specific feature of L2 oral proficiency that is associated with the smoothness and effortlessness of speech. In addition to the utterance fluency analysis in Articles I–III, perceived fluency is considered in Article IV. Examining the connections between utterance and perceived fluency by analyzing both measured fluency and assessed fluency has become one of the main lines of inquiry in L2 speech fluency research, with studies generally finding moderate to strong correlations between (temporal) aspects of fluency and oral proficiency ratings (e.g., Brand & Götz, 2011; Ginther, Dimova, & Yang, 2010; Iwashita, Brown, McNamara, & O'Hagan, 2008; Kahng, 2014), as well as between measured temporal fluency and fluency ratings (Bosker, Pinget, Quené, Sanders, & De Jong, 2013; Cucchiari *et al.*, 2002; Derwing, Rossiter, Munro, & Thomson, 2004; Freed, 1995; Kormos & Dénes, 2004; Magne *et al.*, 2019; Préfontaine, Kormos, & Johnson, 2016; Rossiter, 2009; Suzuki & Kormos, 2020; on links between temporal fluency measures and self-assessed fluency, see Préfontaine, 2013). Cognitive fluency is left out of the scope of the present study, as examining it would require different methodological approaches and different types of data that were not collected for the present study (see Olkkonen & Peltonen, 2017).⁴ Overall, L2 fluency studies combining cognitive fluency and utterance fluency perspectives are relatively rare (but see Kahng, 2020).

A final point regarding definitions of fluency that are employed in the present study relates to the applicability of the psycholinguistic/cognitive foundation to dialogue data. While the present study acknowledges the psycholinguistic basis of L2 speech fluency research and considers (the lack of) speed and efficiency in cognitive processing as one potential explanation for speech (dis)fluency features (Segalowitz, 2010), it is argued that the social–interactional functions of disfluencies (related, for instance, to holding the floor and managing turn-taking; e.g., Clark, 1996, 2002; Clark & Wasow, 1998; for an overview, see De Jong, 2018) should not be overlooked when L2 fluency is studied in an interactional context (see also Carroll, 2004). In other words, the association between cognitive processing and its manifestations in speech (dis)fluency requires re-examination in dialogue data. As dialogue is fundamentally collaborative (cf. Clark's 1996 view of language use as joint action), the relationship between speech fluency features and an individual's cognitive processing may not be as straightforward as in monologue (see also

⁴ In Kahng's (2014) study, retrospective protocols are viewed as providing insights into cognitive fluency. While stimulated recall data were also collected for the present study, the data are regarded as providing relatively indirect information about cognitive fluency. Thus, the stimulated recall protocols are treated as secondary data in the present study, with the main purpose being to provide information about the use of strategies. (For details, see Section 4.2.)

Segalowitz, 2010). For instance, Gilabert, Barón, and Levkina (2011) found that there was a less straightforward relationship between proficiency and CAF measures in dialogue than in monologue. In the present study, the collaborative nature of dialogue is acknowledged by operationalizing interactional fluency as its own dimension of fluency (see further Section 3.3).

2.2 L2 speech fluency measurements

Temporal measures related to the speed of spoken production and the extent of pausing have been used to operationalize the notion “flowing speech” since the early exploratory L2 fluency studies in the 1980s and 1990s (e.g., Lennon, 1990; Raupach, 1980).⁵ Adding measures of repair to indicators of fluency, Skehan (2003, 2009, 2014; Tavakoli & Skehan, 2005) was among the first researchers to formulate a comprehensive framework of fluency dimensions. According to Skehan’s (e.g., 2003, 2009) framework, fluency measures can be grouped under three dimensions: *speed fluency* (notably articulation rate [AR]), *breakdown fluency* (referred to as *pausing* in Skehan, 2014), and *repair fluency* (e.g., false starts, repetitions, and reformulations; for an early L2 fluency study examining all three dimensions, see Riggenbach, 1991). In L2 speech fluency research conducted in the past 10–15 years, these three dimensions have become the standard for choosing representative fluency measures for a comprehensive fluency analysis (e.g., Bosker *et al.*, 2013). In addition to the three main dimensions, a separate *composite* dimension for measures combining aspects from the main dimensions (especially speed and breakdown) can be distinguished (e.g., Tavakoli, Nakatsuhara, & Hunter, 2020). The dimensions with representative measures are compiled in Figure 2.

⁵ The discussion in this section focuses on monologue-based fluency measures; challenges in measuring fluency in dialogue are discussed in Section 3.3.

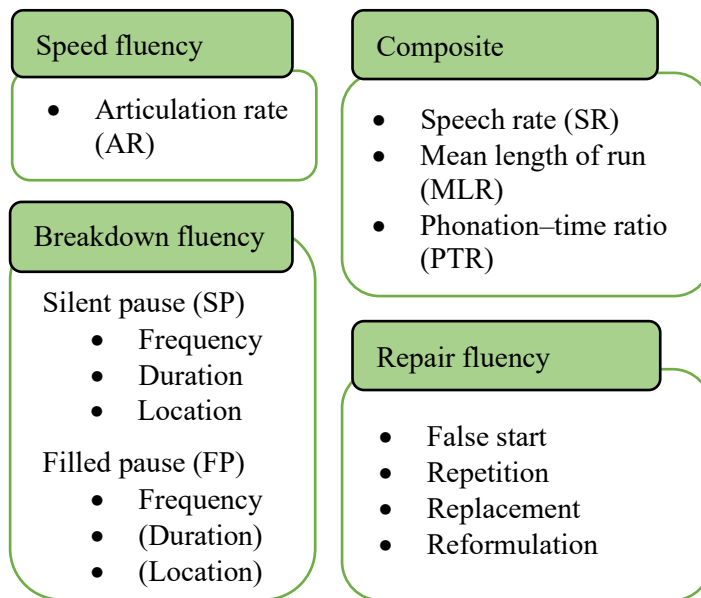


Figure 2. Skehan’s fluency dimensions and common L2 speech fluency measures.

Speed fluency, which refers to how fast speech is produced, includes a widely used and well-established fluency measure, articulation rate (AR). AR is similar to speech rate (SR), as both measure the number of speech units (syllables or words) per total time (in seconds or minutes), but AR excludes pause time from the total time, being a “purer” measure of speed (De Jong, 2016b, p. 212). While SR has traditionally been considered a measure of speed fluency along with AR, SR also contains information about pausing; thus, more recent studies regard it as a composite measure (see, e.g., Tavakoli *et al.*, 2020). SR has often been employed in studies that focus on all CAF dimensions—for instance, in task-based research (e.g., Gilabert *et al.*, 2011; Michel, 2011)—to gain an overall view of fluency. However, other researchers have argued for the use of more specialized measures, such as AR, to capture the different dimensions of fluency and to avoid overlap across dimensions (e.g., De Jong, 2016b). Both SR and AR are included in the present study: SR⁶ due

⁶ In some studies, a pruned variant of SR has been calculated; this entails removing different types of disfluencies (e.g., repetition) from the calculations (e.g., Lennon, 1990). However, this practice seems to be based on the idea that the features removed from the calculations are indicators of disfluency; as the phenomena are regarded as multifunctional in the present study and there was interest in examining these features individually (as stalling mechanisms, see Section 3.2), the SR measure was not pruned (see Section 4.3 for the measures used in the present study).

to its widespread use to ensure comparability across studies and AR to include a measure that captures only the speed of talk without the influence of pauses. Both measures have become well established in the field, demonstrating differences across learners at various proficiency levels (e.g., Iwashita *et al.*, 2008; Kormos & Dénes, 2004; Tavakoli *et al.*, 2020) or levels of fluency (e.g., Hilton, 2008; Riggenbach, 1991), differences between learner and native speech (e.g., Götz, 2013; Kahng, 2014), strong correlations with assessments of fluency (e.g., Baker-Smemoe, Dewey, Bown, & Martinsen, 2014; Cucchiariini *et al.*, 2002; Ginther *et al.*, 2010; Kahng, 2014; Magne *et al.*, 2019), and development during study abroad (SA; e.g., Huensch & Tracy-Ventura, 2017b; Mora & Valls-Ferrer, 2012; Tavakoli, 2018; Towell, Hawkins, & Bazergui, 1996).

Mean length of run (MLR) and phonation–time ratio (PTR) are also often included among speed fluency measures, but similar to SR, they incorporate information about pausing in addition to the amount of production. Therefore, they can be considered as combining the dimensions of speed and pausing, thereby forming “composite measures” (Skehan, 2014, p. 20; Tavakoli *et al.*, 2020). MLR can be considered a measure related to the degree of automaticity (Skehan, 2014; Towell *et al.*, 1996), and PTR provides information about the amount of pausing in relation to speech (Towell, 2002). Both are thus general measures of fluency (Skehan, 2014, p. 21; see also Huensch & Tracy-Ventura, 2017b, p. 287).

Breakdown fluency, or simply pausing, usually entails examining silent pauses (SPs) for their frequency and/or average duration (for a recent overview of pause measures in L2 fluency research, see Shea & Leonard, 2019). Pause frequency is a relatively well-established indicator of disfluency, as learners tend to pause more frequently than NSs (e.g., Kahng, 2014), and less advanced speakers tend to pause more often than more advanced speakers (e.g., Kormos & Dénes, 2004; Riggenbach, 1991; Tavakoli *et al.*, 2020). The number of pauses has also been shown to be linked to judgments of fluency (Cucchiariini *et al.*, 2002; Iwashita *et al.*, 2008; but cf. Préfontaine *et al.*, 2016).

The results concerning the average duration of pauses are more mixed: Some researchers have found pause duration to differ across fluent and non-fluent subjects (e.g., Kormos & Dénes, 2004) and to be linked to L2 proficiency (Shea & Leonard, 2019; Tavakoli *et al.*, 2020), while others have argued that pause length is not a reliable indicator of L2 fluency (e.g., Cucchiariini *et al.*, 2002; De Jong *et al.*, 2015). The mixed results could potentially be explained by cross-linguistic or individual differences in pause location: Some studies have suggested that pause length in the L1 may influence L2 speech and differ across languages (Campione & Véronis, 2002; Riazantseva, 2001), while others have not found support for this argument (De Jong *et al.*, 2015; De Jong & Mora, 2019; Derwing, Munro, Thomson, & Rossiter, 2009; Huensch & Tracy-Ventura, 2017a). There is also tentative evidence to suggest

that pause length in the L2 is influenced by individual speaking style in the L1 (De Jong *et al.*, 2015; Kolly, Leemann, Boula de Mareüil, & Dellwo, 2015; Shea & Leonard, 2019).

In addition to SP frequency and length, it has become increasingly common in recent years to incorporate measures of pause location into L2 speech fluency analyses, based on differences in mid-clause and clause boundary pausing in L1 and L2 speech. The observation that L1 and L2 pause location patterns differ is, however in itself not new: Based on studies of L1 speech suggesting that most SPs are located at clause boundaries (e.g., Hawkins, 1971), some temporal studies in the 1980s comparing L1 and L2 speech confirmed that in L1 speech, SPs mostly occur at clause boundaries, while L2 speech contains more mid-clause pausing (e.g., Raupach, 1980). The findings have been corroborated in more recent studies (De Jong, 2016a; Kahng, 2014; Tavakoli, 2011). Mid-clause pauses have also been shown to correlate strongly with fluency ratings (Kahng, 2014, 2018) and to be linked to L2 proficiency, with less proficient learners exhibiting more mid-clause pausing than more proficient learners (De Jong, 2016a; Shea & Leonard, 2019; Tavakoli *et al.*, 2020). In the present study, SPs are examined from all three perspectives: frequency, duration, and location.

Filled pauses (FPs; such as *uh* and *um*) are also included in Skehan's (2003, 2009, 2014) breakdown/pausing dimension along with SPs. FPs are most commonly examined for their frequency; they can also be examined for their length and location (e.g., Kahng, 2014; Tavakoli *et al.*, 2020), but this has been less common than for SPs. Overall, researchers have not found a similar connection between FPs and disfluency as for SPs and disfluency (e.g., Cucchiari *et al.*, 2002; Kormos & Dénes, 2004; Riggensbach, 1991). Furthermore, nonlinear progression from lower to higher proficiency levels has been found for FPs (Tavakoli *et al.*, 2020). In comparison to SPs, FPs have demonstrated weaker correlations with L2 proficiency (Shea & Leonard, 2019) and distinguished L1 and L2 speakers less clearly (Kahng, 2014). Furthermore, while differences across proficiency levels in FP use were found in Götz's (2019b) recent large-scale corpus study, the study also demonstrated substantial within-group variation in FP use and a more substantial role for other variables besides proficiency (e.g., country of origin) in predicting FP use in L2 speech (see also Götz, 2019a). Previous smaller-scale fluency studies have also shown that FPs are prone to individual variation, especially in L2 speech (e.g., Brand & Götz, 2011, Cucchiari *et al.*, 2002; Dumont, 2018; Götz, 2013; Kahng, 2014; for L1 speech, see, e.g., Clark & Fox Tree, 2002; Maclay & Osgood, 1959), and have multiple functions (e.g., Allwood, Nivre, & Ahlsén, 1990; Kosmala & Morgenstern, 2019; Segalowitz, 2010, p. 40).

These findings can contribute to explaining why it has been challenging to establish a link between FPs and (dis)fluency.⁷

As the view of FPs as straightforward indicators of disfluency receives little support based on previous studies, it is important to examine the potential strategic functions of FPs: They help to cope with processing time pressure and to avoid long pauses, thus contributing to fluency rather than to disfluency (Dörnyei & Kormos, 1998; Kahng, 2014; Olynyk, d’Anglejan, & Sankoff, 1990; see also, e.g., Allwood *et al.*, 1990; Maclay & Osgood, 1959 for a similar view adopted by L1 speech researchers; see also Section 3.1 for interactional functions of FPs). This alternative view of FPs is also adopted in Götz’s (2013) monograph: She views FPs, along with repetitions, for instance, as *fluency enhancement strategies*. Götz’s (2013) approach to FPs is comparable to the problem-solving approach adopted in the present study, which views FPs as stalling mechanisms that are used to buy time and avoid long SPs (see Section 3.2). Thus, there is a need to reconsider the practice of examining FPs and SPs under the same dimension as comparable indexes of speech “breakdown” (see also Dumont, 2018).

Finally, the repair fluency dimension, including measures for false starts, repetitions, replacements, and reformulations (Skehan, 2003, 2009, 2014), is a somewhat neglected dimension of fluency, as most studies have focused on the speed and breakdown dimensions. Furthermore, when the repair dimension has been studied, clear connections have not been found between repair phenomena frequencies and (dis)fluency (e.g., Kormos & Dénes, 2004; Lennon, 1990; Riggensbach, 1991) or between repairs and judges’ proficiency or fluency assessments (e.g., Baker-Smemoe *et al.*, 2014; Bosker *et al.*, 2013; Dumont, 2018), suggesting a complex link between repair and proficiency (e.g., Tavakoli *et al.*, 2020). Instead, repair phenomena seem to occur somewhat infrequently, especially in the relatively small data sets that are often used in L2 speech fluency studies (e.g., Freed, 1995; Kahng, 2014) but also in larger-scale learner corpora (Götz, 2013; Gráf & Huang, 2019; but see Dumont, 2018 for relatively high frequencies of repetition for French learners of English).

⁷ The nature of FPs as straightforward indicators of disfluency is not well established even in L1 speech research (see O’Connell & Kowal, 2004). In the field of psycholinguistics, two contrasting views have been presented: Traditionally, FPs have been considered essentially unplanned indicators of trouble during the speech production process (Levelt, 1989), whereas, more recently, some researchers have viewed them as planned units that are comparable to words (Clark & Fox Tree, 2002). Others have adopted stances that fall between the two extremes and suggested that the evidence for the view of FPs as words is inconclusive (Corley & Stewart, 2008). Approaching FPs from a different perspective in conversation analytical studies, FPs have been shown to be associated with a variety of conversational actions, depending on their positions in the conversations (e.g., Schegloff, 2010).

Similar to FPs, the use of different repair phenomena is associated with strong within-group variation and individual differences (e.g., Baker-Smemoe *et al.*, 2014; Cucchiarini *et al.*, 2002; Dumont, 2018; Freed, 1995; Gráf & Huang, 2019; Götz, 2013; Kahng, 2014; Tavakoli *et al.*, 2020). In addition to the mostly quantitative examinations of repair phenomena, some studies have used qualitative analyses to examine repair features. Generally, these studies have highlighted the role of repetitions in maintaining the flow of speech and avoiding breakdowns (Ejzenberg, 2000; Hilton, 2008; Olynyk *et al.*, 1990; for L1 speech, see Allwood *et al.*, 1990; Clark & Wasow, 1998; Maclay & Osgood, 1959; Tannen, 1989). That is, repetitions can be used strategically to cope with processing time pressure and to maintain fluent speech during planning (Dörnyei & Kormos, 1998) in a similar manner to FPs (see also Götz, 2013; Penttilä, Korpijaakko-Huuhka, & Kent, 2018).

While repetitions, along with FPs, have commonly been treated as indicators of disfluency, from the perspective of their potential fluency-maintaining function, FPs and repetitions can be viewed as more closely connected than FPs and SPs or repetitions and reformulations (cf. Dörnyei & Kormos's 1998 framework). Lennon's (1990) early suggestion of grouping fluency measures under two main dimensions (cf. Skehan's three dimensions, 2003, 2009, 2014) reflects a similar idea: *Temporal fluency* includes speed and SP measures and their combinations, and *vocal fluency* includes FPs and repair measures. Building on Lennon's (1990) suggestion, false starts, replacements, and reformulations (being roughly comparable to self-corrections) could be viewed as forming a more closely connected whole reflecting speech monitoring (see Kormos, 1999, 2000; Levelt, 1989; Olkkonen, 2017) and thus an orientation toward accuracy (Ellis & Barkhuizen, 2005, pp. 149–151; Gilabert, 2007) from a psycholinguistic perspective (see further Olkkonen & Peltonen, 2017). Especially in some task-based CAF studies (e.g., Gilabert *et al.*, 2011), repair-related measures have indeed been examined as indicators of accuracy, not fluency. Finally, in addition to the multifunctional nature of repairs, especially repetitions, a further challenge in the analysis of repairs is that in some studies, different types of repair have been grouped, for instance, into a single “repair” (Iwashita *et al.*, 2008), “disfluency” (Cucchiarini *et al.*, 2002; Kormos & Dénes, 2004), or “hesitation” measure (including FPs; Baker-Smemoe *et al.*, 2014). In these studies, it is perhaps not surprising that connections between repair and (dis)fluency have not always been found, as the different types and functions of repairs have not been considered.

To conclude, while several speed and pausing measures have proven to be reliable indicators of L2 speech fluency, more research is needed on the repair dimension (see also Bosker *et al.*, 2013; Kahng, 2014). In addition, the status of FPs as pauses that are equivalent to SPs requires reconsideration, especially from a problem-solving perspective. Similarly, whether repetitions should be included

under the repair dimension or perhaps as a part of vocal fluency (Lennon, 1990) is another important point for further consideration. These issues will be revisited in conjunction with the presentation of the problem-solving framework used to approach fluency in the present study (Section 3.2). Furthermore, to complement the discussion of monologue fluency measures in this section, measures of interactional fluency will be discussed in Section 3.3.

2.3 Factors affecting L2 speech fluency

Measures of fluency have been used to examine L2 speech from several different perspectives, including how different factors influence L2 speech fluency. Two factors—namely, general L2 proficiency (Section 2.3.1) and L1 speech fluency (Section 2.3.2)—are discussed in this section due to them being central to the present study. Other potential factors influencing L2 speech, such as situational factors (e.g., pre-task planning and task type) and affective factors (see Lintunen *et al.*, 2020b), are left out of the scope of the discussion, as they were not the focus of the present study. The effects of task mode (monologue vs. dialogue) on fluency are discussed in Section 3.3 in conjunction with interactional fluency.

2.3.1 General L2 proficiency

When studying the role of general L2 proficiency in L2 speech fluency, researchers have mostly used cross-sectional designs: Learner groups representing different proficiency levels or “fluent” and “non-fluent” speakers have been compared (e.g., Cucchiari *et al.*, 2002; Ejzenberg, 2000; Hilton, 2008, 2014; Kormos & Dénes, 2004; Riazantseva, 2001; Riggenbach, 1991). Generally speaking, these studies have shown that learners at higher proficiency levels also exhibit higher levels of L2 speech fluency; yet, some measures have indicated differences more clearly than others. In a relatively early influential study involving comparisons of two learner groups (eight Hungarian university students of English and eight lower-intermediate students of English), Kormos and Dénes (2004) demonstrated that out of 10 temporal fluency variables, university students outperformed the lower-intermediate students in five measures: SR, PTR, MLR, mean pause duration, and pace (the number of stressed words per minute). In a more recent study, Tavakoli *et al.* (2020) examined learners’ productions from four proficiency levels (levels A2–C1 in the Common European Framework of Reference for Languages [CEFR]; Council of Europe, 2001) using a comprehensive set of speed, composite, breakdown, and repair fluency measures. The findings indicated differences across fluency measures from different dimensions in regard to their ability to distinguish learners at different proficiency levels: While speed and composite measures distinguished levels A2, B1, and B2

(but not C1), comparable to Kormos and Dénes's (2004) earlier findings, breakdown fluency measures mostly differentiated only A2 from the higher levels. Finally, for repair and FPs, more complex links with proficiency level were detected, along with substantial within-group variation (cf. Section 2.2).

A similar approach to Kormos and Dénes's (2004), involving comparisons of learner groups from different school levels, was used in the present study: Speech samples from upper secondary school and university (Article I), as well as the ninth grade of compulsory education and upper secondary school (Articles II–IV), were examined. In contrast to most L2 fluency studies, which have often examined university students' and/or adult learners' fluency, the present study thus fills a gap by examining a group of relatively young, (lower)-intermediate-level students (ninth graders of about 15 years old), along with upper secondary school students and university students. The study also complements previous studies on L2 speaking skills and fluency conducted in the Finnish context, which have often focused on university-level students or adults (e.g., Pietilä, 1999; Toivola, Lennes, & Aho, 2009; Ullakonoja, 2011), although studies focusing on younger learners exist as well (e.g., Hildén, 2000; Olkkonen, 2017; Paananen-Porkka, 2007). In the present study, intact school-level groups were used to compare learners at different proficiency levels. To ensure that the school-level groups were representative of different proficiency levels on average, proficiency-level test scores (Lexical Test for Advanced Learners of English [LexTALE]; Lemhöfer & Broersma, 2012) obtained for Articles II–IV were consulted. The scores confirmed that sufficient within-group homogeneity existed in terms of the students' proficiency and that the more advanced student group was also on average more proficient. Yet, the use of intact school-level groups ensured sufficient within-group variation for correlational analyses (see Section 4.1 for details).

In addition to cross-sectional comparisons, some researchers have followed learners' L2 speech fluency development (along with their L2 proficiency development) over a certain period. Most of these longitudinal studies have been concerned with development during study abroad (SA⁸; e.g., Freed, 1995; Huensch & Tracy-Ventura, 2017b; Lennon, 1990; Mora & Valls-Ferrer, 2012; Segalowitz & Freed, 2004; Tavakoli, 2018; Towell, 2002; Towell *et al.*, 1996). Despite these studies generally pointing to the benefits of SA for fluency development, especially in SR and MLR (e.g., Tavakoli, 2018), researchers have also highlighted the importance of background factors in influencing developmental patterns—notably, the use of the target language during SA (e.g., Mora & Valls-Ferrer, 2012). Studies have also found that different dimensions of fluency may show different

⁸ For a recent overview of the benefits of SA for L2 development, including fluency as part of oral skills, see Borràs and Llanes (2019).

developmental patterns: For instance, Huensch and Tracy-Ventura's (2017b) study demonstrated gains in speed and breakdown fluency (yet faster development for speed than breakdown measures), but not for repair fluency (cf. Section 2.2). That is, while the benefits of SA for the development of speech fluency seem to be relatively well established, more research is needed on the potentially differing developmental patterns for different dimensions of fluency, as well as on the factors, such as engagement in target language use, that influence these patterns.

Some SA fluency studies have included a comparison group of students in formal instruction (e.g., Freed, 1995; Mora & Valls-Ferrer, 2012; Segalowitz & Freed, 2004) to enable comparisons of fluency development between students who stay at their home universities and those who spend time abroad. However, few studies have focused on fluency development solely in formal instruction settings (but see Derwing, Munro, & Thomson, 2008; Tavakoli, Campbell, & McCormack, 2016; Peltonen & Lintunen, 2019; Tonkyn, 2012). While SA studies typically examine development over several months, even years (Borràs & Llanes, 2019), the studies conducted in formal settings have focused mostly on the short-term development of fluency (ranging from four weeks in Tavakoli *et al.*, 2016 to 10 weeks in Tonkyn, 2012). A notable exception is Derwing and colleagues' (2008) longitudinal fluency development project in an immersion context. Two groups of L2 English learners (Slavic and Mandarin speakers) in Canada were followed over a two-year period in the project, and only the Slavic group demonstrated fluency improvement (Derwing *et al.*, 2008; see also Derwing *et al.*, 2009). Similar to some SA studies, Derwing *et al.* (2008) highlighted the importance of willingness to communicate and interaction outside the classroom in explaining their results.

In the studies examining short-term development in formal instruction, the interest has been in examining the effects of pedagogic interventions (Tavakoli *et al.*, 2016) or speaking skills instruction more generally (Peltonen & Lintunen, 2019; Tonkyn, 2012) on L2 fluency development. The studies have provided somewhat mixed results: Peltonen and Lintunen (2019) found overall improvement for three of the six examined fluency measures (SR, PTR, and mean length of SPs), but individual differences and differences across instructional groups were also observed. Similarly, the fluency gains observed in Tonkyn's study (2012) based on interview data were relatively limited, with improvement found for fluent runs and turn length. Instead, Tavakoli, Campbell, and McCormack's (2016) study showed that the experimental group receiving fluency instruction showed fluency development for four out of nine fluency measures (SR, AR, MLR, and PTR), while the control group did not develop in fluency. Based on these three studies, it seems that short-term oral skills practice can facilitate fluency development to some extent. However, targeted fluency instruction involving awareness-raising activities,

strategy instruction, and fluency practice seems to be even more beneficial for fluency development.

2.3.2 L1 speech fluency

As discussed in Section 2.1, learners' fluency has often been compared to native speakers' (NSs') fluency both implicitly and explicitly. While generally not the main focus in studies comparing learner and native fluency, NSs have, nevertheless, been shown to vary in their fluency (e.g., Götz, 2013; Kahng, 2014; see also Moniz *et al.*, 2014). As early as in the 1960s, Goldman-Eisler (1968) demonstrated that NSs vary in their temporal aspects of speech, with some speaking more slowly than others and some exhibiting more pausing than others, for instance (on individual differences in hesitation phenomena, see also Maclay & Osgood, 1959). Thus, it is important to examine the extent to which L1 speaking style, as demonstrated by the variation in the temporal aspects of L1 speech, is reflected in L2 speech fluency.

Despite some early examples of studies comparing speech in the L1 and L2 by the same speakers (Lehtonen, 1979; Möhle, 1984; Raupach, 1980), most L2 speech fluency studies have generally not obtained L1 speech samples from the subjects and have focused only on L2 samples (or used an external group of NSs as controls for the learners, as explained earlier). Yet, a growing body of research incorporating L1 and L2 speech samples from the same speakers suggests that L2 speech fluency can, at least to some extent, be explained by L1 speaking style (De Jong & Mora, 2019; De Jong *et al.*, 2015; Derwing *et al.*, 2009; Duran-Karaoz & Tavakoli, 2020; Huensch & Tracy-Ventura, 2017a; Kahng, 2020; Shea & Leonard, 2019; see also Bradlow, Kim, & Blasingame, 2017). Derwing *et al.* (2009) found initial correlations in L1 and L2 speech fluency for two different L1 groups (Slavic and Mandarin), but only for the Slavic group on later data collection times, suggesting that the relationship between L1 and L2 fluency can change over time. De Jong *et al.*'s (2015) study examining L1 English and L1 Turkish speakers of L2 Dutch showed correlations across the L1 and L2 on several fluency measures (replicated for three measures with L1 Spanish and L2 English participants in De Jong & Mora, 2019), with SP duration between Analysis of Speech Units (AS-units; Foster, Tonkyn, & Wigglesworth, 2000) in the L1 explaining the most variation in the L2 fluency measures. Finally, examining the effects of SA, L1 background (Spanish, English, and French), and L1 speaking style on L2 speech fluency, Huensch and Tracy-Ventura (2017a) demonstrated correlations across L1 and L2 fluency with several measures but also highlighted the importance of cross-linguistic differences together with the L1 profile in explaining shifts in L1–L2 correlations before and after SA.

The studies conducted so far on the effects of L1 speaking style on L2 speech fluency are relatively difficult to compare due to their different designs and L1s, but

the overall picture emerging from the studies is that regardless of the proficiency level or the participants' L1 (or the L1–L2 pair), L1 fluency is at least to some extent linked to the L2. In addition to studies focusing specifically on fluency, the links between L1 and L2 temporal aspects are supported by findings from other fields examining L1 speech and/or L1 influence. For instance, studies have suggested that SR in particular (e.g., Bradlow *et al.*, 2017; see also Guz, 2015; Towell & Dewaele, 2005), including information about pausing, and pausing separately (Kolly *et al.*, 2015) in the L2 are prone to L1 speaking-style influence. In addition to measures of speed and pausing, results from Fehringer and Fry's (2007) study comparing FPs, filler words, repetitions, and reformulations in the L1 and L2 suggest that preferences for certain hesitation types in the L1 are also visible in L2 production and are thus linked to L1 speaking style and individual differences (see also Sections 2.2 and 3.2 on individual differences in stalling mechanisms).

Various factors have been suggested to explain the differences among L1 speakers in regard to their fluency; these range from cross-linguistic factors to individual differences in cognitive processing (Bradlow *et al.*, 2017, p. 887), personality (e.g., Ramsay, 1968), and social/demographic factors (such as age, gender, and geographic region; e.g., Bortfeld, Leon, Bloom, Schober, & Brennan, 2001; Jacewicz, Fox, & Wei, 2010; Quené, 2008; Yuan, Liberman, & Cieri, 2006; for further examples, see De Jong, 2018). In addition to the need for further studies examining the reasons for L1-related differences reflected in the L2, the role of mediating factors in L1–L2 connections, such as cross-linguistic differences and proficiency levels, also requires further study. The present study (Article III) examines the influence of the latter factor in L1–L2-fluency connections by taking into account the learners' school levels in the analyses.

3 Broadening the scope of L2 speech fluency analysis

In this chapter, broadening the scope of L2 speech fluency analysis will be considered from three main perspectives (cf. the three main contributions of the present study outlined in Figure 1). First, in Section 3.1, a mixed-methods approach to L2 fluency analysis is presented. Then, in Section 3.2, the Fluency Resources Framework introducing a problem-solving perspective to fluency is discussed. Finally, in Section 3.3, expanding the scope of fluency analysis from individual to interactional fluency is presented.

3.1 A mixed-methods approach

As discussed in Section 1.2, one of the central contributions of the present study that fills a gap in research is that it combines qualitative analysis with quantitative analysis. Overall, a mixed-methods approach has the potential to improve the validity of a study through the triangulation of complementary methodological approaches (Dörnyei, 2007). Specifically, incorporating a complementary qualitative approach provides a more comprehensive picture of fluency, enabling detailed analyses of fluency-related features in their immediate contexts in the present study. As discussed in Section 2.1, L2 speech fluency research has mostly been quantitative in nature, involving frequency-based analyses of fluency and disfluency features, while little attention has been paid to the functions or contexts of these features. Yet, the few studies that have included mixed-methods analyses (e.g., Brand & Götz, 2011; Ejzenberg, 2000; Hilton, 2008, 2014; Towell *et al.*, 1996), along with purely quantitatively oriented L2 speech fluency studies (e.g., Cucchiaroni *et al.*, 2002; Kahng, 2014) and L1 speech studies (Allwood *et al.*, 1990; Maclay & Osgood, 1959), have shown that certain features associated with disfluency, such as filled pauses (FPs) and repetitions, are prone to individual variation and are not straightforward indicators of disfluency (see also Section 2.2). Instead, they can be used as resources for maintaining fluent speech (see Section 3.2).

Whereas a quantitative approach is well suited for speed and breakdown (SP) fluency measures (as all speech necessarily includes both syllables and silences;

Grosjean, 1980), the relatively infrequently occurring and potentially multifunctional phenomena, including FPs and repetitions, benefit from a complementary qualitative approach that illustrates the functions, contexts, and individual differences associated with these phenomena (see also Degand *et al.*, 2019; Dumont, 2018; Lennon, 2000). While the aforementioned studies have mostly been based on monologic production (with the exception of Ejzenberg, 2000), examining the functions of disfluency markers in dialogue data is even more crucial due to the added element of turn-taking. As studies on L1 speech have shown, repair phenomena, especially repetitions, as well as FPs, have important interactional functions in addition to their fluency-enhancing functions from an individual's perspective, including holding the floor (Allwood *et al.*, 1990; Bortfeld *et al.*, 2001; Clark, 1996, 2002; Clark & Wasow, 1998; Maclay & Osgood, 1959; Tannen, 1989; see further Section 3.3).

In practice, the mixed-methods approach involved complementing a quantitative, frequency-based analysis of utterance fluency features with a qualitative examination of the contexts and functions of selected features in the present study. The purpose of triangulating two data analysis methods was to gain insights into individual differences across learners and to examine how the particular features are used in speech to maintain its flow in addition to quantitative, group-level tendencies. As the qualitative analyses focused on only a small number of individuals and their use of certain features, the generalizability of the observations should be regarded as relatively limited. Nevertheless, the observations can be used as a basis for studying the findings further and to inform theory building and hypothesis testing in future studies. (For details of the mixed methodology and qualitative analyses, see Section 4.4.2.)

3.2 Combining fluency and problem-solving: The Fluency Resources Framework

The mixed methodology forms the basis for the novel problem-solving framework (the Fluency Resources Framework) that is applied in the present study. The problem-solving framework is based on Dörnyei and Kormos's (1998) taxonomy of mechanisms for managing problems related to L2 speech production. Two problem-solving mechanisms—namely, stalling mechanisms and communication strategies (CSs)—are central to the present study due to their potential to contribute to enhanced utterance fluency. These problem-solving mechanisms (*fluency resources*) are linked to L2 (temporal) fluency in the Fluency Resources Framework.

First, stalling mechanisms (including, e.g., FPs and repetitions, among other similar features) can be viewed as problem-solving mechanisms that can contribute to maintaining utterance fluency by providing solutions for coping with the

processing time pressure that is involved in speaking in an L2 (Dörnyei & Kormos, 1998). Compared to highly automatized L1 speech production processes, speaking in an L2 involves more attentional processes and requires the use of problem-solving mechanisms, among them stalling mechanisms, at different stages of speech production (Kormos, 2006). Based on Levelt's (1989) model of L1 speech production and Kormos's (2006) L2 adaptation of the model, the speech production process consists of three main stages: 1) *conceptualization* (involving message generation through the processes of macroplanning and microplanning, resulting in a preverbal message), 2) *formulation* (involving the development of linguistic representation through the processes of lexico-grammatical, morpho-phonological, and phonetic encoding, resulting in a phonetic plan), and 3) *articulation* (involving the execution of the phonetic plan by articulating the message).⁹ In addition to these three stages, speech is monitored for accuracy (*monitoring*; see also Kormos, 1999, 2000). From the perspective of speech production, stalling mechanisms can be used to provide more processing time both during message macro- and microplanning and during the encoding of the preverbal message (Dörnyei & Kormos, 1998); from the perspective of utterance fluency, stalling mechanisms help in reducing time spent in silence and keeping the flow of talk going.

In addition to stalling mechanisms, another type of problem-solving mechanism that can potentially contribute to enhanced utterance fluency is CSs (e.g., paraphrases), which are used to solve difficulties related to lexical gaps or lexical retrieval (Dörnyei & Kormos, 1998). These lexical problem-solving mechanisms can thus also be used during the planning and encoding of the preverbal message, with the main purpose of coping with trouble related to retrieving lemmas that correspond to the concepts outlined in the preverbal message (Dörnyei & Kormos, 1998). CSs, together with stalling mechanisms, can function as fluency resources, as they help in overcoming troubles during the L2 speech production process and thus help in maintaining the flow of speech. The relationship between these two types of fluency resources and temporal fluency is illustrated in Figure 3 on the following page.

⁹ As stated in Section 2.1, as the present study is not focused explicitly on the processes underlying speech production (cognitive fluency), Levelt's (1989) model is presented here in a simplified form to contextualize the problem-solving framework. (For a detailed discussion of speech production from an L2 perspective, see Kormos, 2006.)

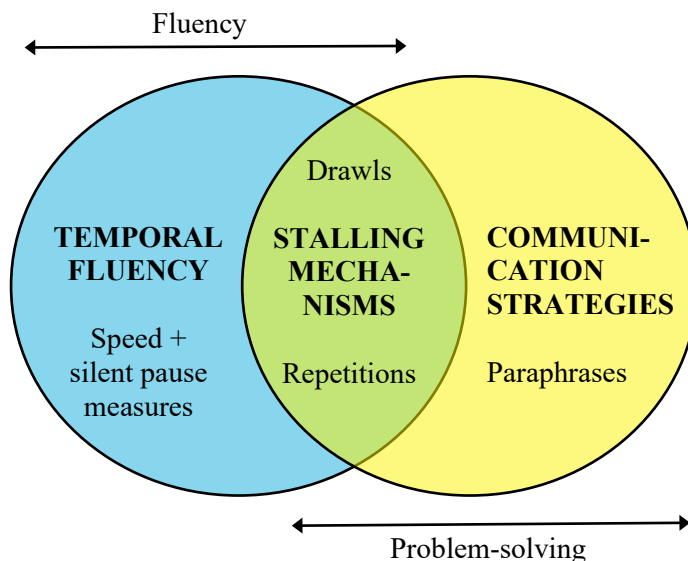


Figure 3. The Fluency Resources Framework (originally published in Article II, p. 3).

In Figure 3, the left side represents temporal fluency measures that have been found to be reliable indicators of L2 speech fluency, notably different speed and silent pause measures (cf. Section 2.2). The middle part, which overlaps with temporal fluency, represents stalling mechanisms (such as drawls and repetitions): measures that are not as straightforwardly linked to disfluency and that can be used as fluency-maintaining devices. Stalling mechanisms include the previously mentioned FPs and repetitions along with fillers¹⁰ (e.g., Hasselgren, 2002; Wolk, Götz, & Jäschke, 2020) and drawls (Möhle, 1984; Raupach, 1980; see also Dumont, 2018; Riggenbach, 1991). These features help to maintain speech fluency by reducing time spent in silence and by providing additional planning time (Dörnyei & Kormos, 1998).¹¹ Learners have been shown to exhibit strong individual variation related to stalling mechanism use (cf. Section 2.2) and may prefer certain stalling mechanisms over others (e.g., Dumont, 2018; Fehringer & Fry, 2007; Götz, 2013; Hasselgren, 2002;

¹⁰ Various terms have been used for these phenomena (for a recent overview, see Wolk *et al.*, 2020). The term adopted here (based on Dörnyei & Kormos, 1998) includes two main categories: *discourse markers* (e.g., *you know*, *like*, and *well*) and *smallwords* (*sort of* and *kind of*; based on Götz, 2013). In the present study, fillers thus include “lexicalized” pauses, while “non-lexicalized” FPs (*uh* and *um*) are considered separately (cf., e.g., Corley & Stewart, 2008, in which the term *fillers* refers to FPs).

¹¹ Stalling mechanisms have also been referred to as *time-gaining strategies* (e.g., Celce-Murcia, Dörnyei, & Thurrell, 1995) and *speech management strategies* (Rühlemann, 2006; see also Götz, 2013, pp. 34, 41–42).

Riggenbach, 1991; Wolk *et al.*, 2020; cf. Seliger's 1980 low and high input generators; for L1 speech, see Maclay & Osgood, 1959).

As stated above, in addition to stalling mechanisms, CSs can also contribute to maintaining fluency (the right side in Figure 3). Because CSs comprise problem-solving mechanisms that are used for coping with vocabulary-related L2 resource deficits (Dörnyei & Kormos, 1998), using these resources efficiently to overcome trouble related to lexical retrieval can reflect positively on utterance fluency.¹² Following Færch and Kasper's (1983, p. 36) oft-cited definition, CSs are defined as "*potentially conscious*" problem-solving behaviors, which include different types of achievement strategies (notably paraphrases, transfer, and appeals for assistance) and reduction strategies (e.g., message abandonment), with only the former being examined in the present study due to the focus on potentially fluency-enhancing strategies.¹³ While there is a long tradition of studying learners' CSs in their L2s (for reviews, see Dörnyei & Scott, 1997; Kasper & Kellerman, 1997; Nakatani & Goh, 2007) and several L2 fluency researchers have discussed the potential links between strategies and fluency in theory (e.g., Lennon, 1990; Koponen & Riggenbach, 2000; Segalowitz, 2010), CSs have generally not been examined in L2 speech fluency studies along with temporal indicators of fluency (but see Dörnyei, 1995; Hilton, 2008). Furthermore, CSs have not usually been studied along with other problem-solving mechanisms, such as stalling mechanisms, either (Dörnyei & Scott, 1997, pp. 178–179); the Fluency Resources Framework thus links these three areas of study (fluency, stalling mechanisms, and CSs) and provides the basis for examining them empirically.

Previous CS studies have demonstrated overall differences in CS frequencies across proficiency levels, with less proficient learners using more strategies than more proficient learners (Chen, 1990; Dobao, 2001; Poulisse, 1990), whereas the results regarding more specific CS categories are more mixed (Paribakht, 1985; Poulisse, 1990; for the CS types examined in the present study, see Section 4.3). It should also be noted that several models of CS categorizations have been presented and employed in empirical studies, making it challenging to compare results across studies. Nevertheless, researchers have pointed to the importance of various factors,

¹² Dörnyei and Kormos (1998) also mention grammatical and phonological/articulatory problem-solving mechanisms, but lexical mechanisms are more common and have also been studied more extensively.

¹³ Due to the focus on problem-solving mechanisms related to maintaining fluency either via buying more time (stalling mechanisms) or overcoming lexical problems (CSs), strategies related to the *negotiation of meaning* (e.g., Long, 1996; also known as "problem-solving mechanisms related to other-performance problems" in Dörnyei and Kormos' 1998 framework, pp. 375–376) were not included in the scope of the present study.

such as task type (e.g., Dobao, 2001; Poulisse, 1990) and the efficiency of the strategies (Chen, 1990), in explaining potentially conflicting results from different studies. There is also some evidence to suggest that the development of CSs may follow a U-shaped pattern: Dobao (2001) found that Spanish and Galician elementary- and advanced-level students of English used more strategies than intermediate-level learners in their L2. Due to the mixed results and different theoretical backgrounds, the usefulness of teaching CSs in the classroom has also been debated (Dörnyei, 1995).

While research on CSs was most popular during the 1980s and 1990s, combining the study of CSs with the current CAF framework of L2 proficiency, especially its fluency dimension, provides a new angle to this traditional topic and addresses the general aim of the dissertation: What kinds of resources, including CSs, do learners use to maintain their speech fluency? Because the connections between problem-solving mechanisms and fluency have not previously been examined systematically, the inclusion of temporal fluency, stalling mechanisms, and CSs in a single study has the potential to provide novel insights into understanding the different mechanisms for maintaining fluent speech and into individual differences related to the use of these mechanisms. As Segalowitz (2010) has noted, “strategy use reveals *a need* to cope with a potential fluency vulnerability and, at the same time, indicates *an ability* to compensate for this vulnerability” (p. 42, added emphasis), making CSs a particularly interesting phenomenon to be studied in conjunction with fluency. For instance, despite a learner’s speech containing disfluencies, such as SPs caused by word searches, the speech may also include FPs, repetitions, and/or CSs that help in overcoming the disfluencies.

So far, in this section, the problem-solving framework and the two types of fluency resources (stalling mechanisms and CSs) have been considered mainly from an utterance fluency perspective: that is, how these aspects can be measured from speech productions (see Article II). However, the framework was also adapted for the purpose of examining the concepts from an assessment perspective (perceived L2 fluency) in Article IV. Along with individual and interactional fluency, the raters in Article IV were asked to assess the participants’ strategic competence. The instructions guided the raters toward a relatively narrow operationalization of strategic competence, including only CSs to correspond with the Fluency Resources framework and the measurement perspective introduced in Article II (see Section 4.2). In terms of models of communicative competence, the approach is thus broadly comparable to the understanding of strategic competence in Canale and Swain’s

(1980) and Celce-Murcia, Dörnyei, and Thurrell's (1995) models, which conceptualize strategic competence chiefly in terms of CSs.¹⁴

3.3 From individual to interactional fluency

While stalling mechanisms and CSs can be used to maintain the flow of speech from an individual speaker's perspective, these resources can also be employed in interaction, along with additional dialogue-specific resources. As established in Chapter 1, the main interest in L2 speech fluency research has nevertheless been on examining fluency from an individual's perspective, echoing the mainstream view of L2 competence as an individual's property in SLA (for criticism of this view, see, e.g., Atkinson, 2019; Firth & Wagner, 1997). Similarly, fluency definitions have tended to emphasize the cognitive basis of speech fluency (e.g., Lennon, 2000; see also Section 2.1). However, recent developments in the field of L2 speech fluency research suggest a shift toward exploring fluency in the dialogic context and acknowledging the social dimension of fluency (see, e.g., Segalowitz, 2016; Tavakoli, 2016; Wright & Tavakoli, 2016). Yet, so far, relatively few empirical studies applying the concept of fluency to dialogue data have been conducted (but see Sato, 2014; Tavakoli, 2016; Witton-Davies, 2014 for analyses of L2 peer interaction; Gilabert *et al.*, 2011; Michel, 2011; Tavakoli, 2018 for task-based studies on CAF; Van Os, De Jong, & Bosker, 2020 for an experimental study). Most studies use monologic data, and while some studies have used interactional data—for instance, interviews between NSs and non-native speakers (NNSs; e.g., Derwing *et al.*, 2004; Dumont, 2018; Götz, 2013; Tonkyn, 2012)—the focus has still been on individual (NNS) performance. Theoretical accounts have not typically acknowledged interactional perspectives on fluency either (but see De Jong, 2018; Guillot, 1999; Götz, 2013; Kirk & Carter, 2010; Lauranto, 2005), and there have been few attempts to define fluency in an interactional context (but see McCarthy, 2010). The present study thus addresses this gap in research by presenting a theoretical–methodological approach to fluency in an interactional context.

A notable exception to the mainstream, monologue-based approach to L2 fluency is presented by McCarthy (2010). Applying the notion of interaction as co-constructed (Jacoby & Ochs, 1995) to the context of fluency and highlighting the

¹⁴ Bachman's (1990; Bachman & Palmer, 1996) models of communicative language ability, which are influential in language testing, reflect a broader understanding of strategic competence. In Bachman and Palmer (1996), strategic competence is viewed in terms of metacognitive strategies, which relate to the management of language use in terms of goal setting (e.g., choosing tasks), assessment (e.g., assessing one's own knowledge and task characteristics), and planning (e.g., selecting the knowledge needed for the task).

participants' collaboration in maintaining the flow of the interaction, McCarthy (2010) suggests that fluency in interaction can be captured with the concept of *confluence*, which emphasizes the jointly constructed nature of fluency: In interaction, the participants are not only responsible for their own contributions but also share the responsibility of maintaining fluency across turn boundaries. For instance, by jointly reducing pause time between turns (Stivers *et al.*, 2009), the participants maintain the flow of talk collaboratively (see also Clark, 1996). The view of fluency as co-constructed forms the core of the approach to fluency in interaction in the present study: *Interactional fluency* refers to the collaborative, between-turn aspects of fluency (comparable to confluence), while *individual fluency* in an interactional context, in turn, refers to a participant's within-turn fluency. Although other researchers have also suggested that individual and interactional fluency may form separate constructs (Sato, 2014; Tavakoli, 2016), the present study is among the first to operationalize these two aspects systematically with distinct measures.

As discussed previously, while the analysis of interactional fluency has been relatively rare in L2 speech fluency research to date, the co-constructed nature of L2 interaction has been studied from other viewpoints. Notably, in L2 assessment, learners' interactional practices as indicators of *interactional competence* (IC; Kramsch, 1986) have been examined both from raters' (Borger, 2019; Ducasse & Brown, 2009; May, 2009) and learners' performance (Galaczi, 2014) perspectives (see also May, Nakatsuhara, Lam, & Galaczi, 2020 on classroom assessment of and feedback on IC).¹⁵ L2 interaction and its development have also been studied within CA-SLA by applying conversation analysis (CA; Sacks, Schegloff, & Jefferson, 1974) to L2 interactional data (e.g., Pekarek Doehler & Pochon-Berger, 2015; Pekarek Doehler & Berger, 2018; for overviews, see Hall & Pekarek Doehler, 2011; Kasper & Wagner, 2011; Pekarek Doehler, 2019; Salaberry & Kunitz, 2019). Furthermore, studies on English as a lingua franca examining collaboration in interaction have explored participants' mutual efforts in maintaining the flow of interaction (e.g., Hüttner, 2009; for an overview, see Hynninen, 2020).

¹⁵ In addition to L2 assessment research on interactional competence, "fluency" and "interaction" are also commonly included in the rubrics of L2 speaking tests (for an overview of the role of fluency in four tests of English, see De Jong, 2018). For instance, in the CEFR (Council of Europe, 2001), both dimensions have their own descriptors in the "qualitative aspects of spoken language use" assessment scale (pp. 28–29), with "fluency" mainly including aspects of (individual) speed, pausing, and repair fluency, and "interaction" drawing on dimensions of IC (e.g., turn-taking management), thus also containing elements that could be considered indicators of interactional fluency.

As much of the research on L2 IC has been conducted from a CA-SLA perspective (Pekarek Doehler, 2019), it is useful to consider the notion of interactional fluency, understood as the participants' mutual efforts to maintain flow across turns and minimize silences between turns, through a CA lens. From the perspective of the CA tradition, interactional fluency is linked to the *progressivity* of turn-taking or to the smooth continuation from one element to the next, wherein a turn indicates understanding of a previous turn (e.g., Schegloff, 2007). Successful L2 interaction thus entails monitoring the ongoing interaction and adapting one's own turns accordingly (including, for instance, their appropriate timing; e.g., Pekarek Doehler & Pochon-Berger, 2015; see also Clark 1996, 2002 regarding coordinating actions).¹⁶ These skills in turn-taking practices, including the increased approximation to the “no gap no overlap” pattern that is typical of L1 interactions (Sacks *et al.*, 1974), develop with increasing L2 IC (Galaczi, 2014; Pekarek Doehler & Pochon-Berger, 2015) and are closely linked to the notion of interactional fluency.

As the participants' joint efforts to ensure smooth turn-taking are essential for establishing interactional flow, interactional fluency can be viewed as roughly comparable to a dimension of L2 IC that is referred to as *turn-taking management* (Galaczi, 2014, Borger, 2019; *turn management* in Galaczi & Taylor, 2018). For instance, in Galaczi's (2014) study on interaction in paired L2 speaking tests, the smoothness and rapidity of turn-taking was identified as one of the main aspects of distinguishing learners' interactions at different proficiency levels. Similarly, in Borger's (2019) study on IC from an assessment perspective, turn-taking management formed one of the main categories in the raters' comments, including aspects such as the smoothness of transitions between turns, which is comparable to interactional fluency. However, IC is a multidimensional concept, much like fluency itself, and encompasses dimensions in addition to turn-taking management. While there are some differences between IC taxonomies and in the operationalizations of IC across studies (see also Pekarek Doehler, 2019, p. 50), generally managing the topics of the interaction (*topic development/management*), demonstrating active listening (*listener support*), using *nonverbal resources* to support interaction, and using *repair* to address interactional trouble are viewed as aspects of IC, especially in the L2 testing tradition, in addition to turn-taking management (see He & Young, 1998; recently Galaczi & Taylor, 2018). Thus, within the broader notion of IC, turn-taking management can be viewed as the aspect of IC that most resembles the notion

¹⁶ From a psycholinguistic perspective, lower-level learners might spend more resources on planning their individual turns and thus focusing on individual fluency, with more proficient learners being able to allocate more resources to discourse-level monitoring, including turn-taking smoothness, and thus interactional fluency (see also Kormos, 2006; Levelt, 1989, pp. 110–123).

of interactional fluency, as it relates to how closely turns are linked across speakers (e.g., Galaczi, 2014).

In addition to the concept of interactional fluency coming close to the notion of turn-taking management, the notions of *alignment* (Pickering & Garrod, 2004; see also Costa, Pickering, & Sorace, 2008) and *accommodation* (Giles, Coupland, & Coupland, 1991), which have been studied from both the psycholinguistic and sociolinguistic perspectives, can also be viewed as being related to the notion of L2 interactional fluency. Both phenomena entail the participants acknowledging the interlocutor's contributions and adapting their own behavior accordingly (e.g., Dings, 2014), which are also essential elements for establishing interactional fluency. Participants have been shown to align their SR (Street, 1983), pause length (Jaffe & Feldstein, 1970), and interactional rhythm (Auer, Couper-Kuhlen, & Müller, 1999), for instance, and the alignment of these features can be viewed as contributing to interactional fluency.

While phenomena related to interactional fluency have been examined from other perspectives and within other frameworks, L2 fluency studies have mostly focused on individual fluency and only rarely explored the interactional aspects of fluency. A notable, pioneering exception is Riggensbach's (1991) microanalytical study of fluency in six conversations between NSs and NNSs (with a focus on NNS contributions). In the study, Riggensbach (1991) examined aspects that can be considered indicators of interactional fluency (referred to as *conversational fluency*; e.g., backchannels, other-repetitions, and collaborative completions) along with aspects of individual fluency. In a more recent study, Sato (2014) approached fluency in L2 interaction from a perceived fluency perspective and found that raters paid attention to collaborative elements, including turn-taking and scaffolding, in addition to individual elements, when rating fluency based on peer interaction samples. Similar results have been obtained from experimental settings involving manipulated speech. In a study focusing on turn-taking behavior and perceived fluency, Van Os, De Jong, and Bosker (2020) examined the effects of delay steps (different gap and overlap conditions in question–answer sequences) and SR (fast vs. slow) on the fluency assessments of native and non-native speech. The results demonstrated an interaction between the two aspects, suggesting that listeners attend to both when making their fluency judgments. Van Os *et al.*'s (2020) experiments thus also suggest that raters pay attention to both individual fluency (SR) and interactional fluency (gaps and overlaps between turns) when rating fluency based on dialogic data.

Other notable studies examining fluency in L2 dialogue (Tavakoli, 2016; Witton-Davies, 2014) have mainly focused on comparing monologue and dialogue performance (for task-based studies, see Gilabert *et al.*, 2011; Michel, 2011; Tavakoli, 2018). As monologue-based measures have been adapted to interactional

contexts in these studies, individuals' contributions have received more prominence than the specifically interactional aspects of fluency. Tavakoli's (2016) and Witton-Davies' (2014) studies can nevertheless be considered pioneering, having explored fluency in a rarely studied L2 dialogue context and having discussed the treatment of features that are characteristic of interactional data, especially (between-)turn pauses, in fluency analysis.

Building on previous studies' analyses of fluency based on L2 dialogue data, the present study aimed to develop ways to capture learners' collaboration in maintaining fluency (interactional fluency) in interactional contexts. Two related but distinct aspects (cf. Skehan's 2003, 2009, 2014 monologue fluency dimensions) were viewed as providing insights into how fluency is maintained across turn boundaries collaboratively: *temporal fluency* and *cohesive devices* (see also Section 4.3). Temporal fluency includes measures related to (between-)turn pauses and thus captures the previously mentioned mutual responsibility of minimizing pause time between turns. The joint management of pause time, in turn, is based on the notion of time as a shared resource in interaction (Auer *et al.*, 1999): In addition to maintaining fluency within their own turns, the participants share the responsibility of ensuring smooth turn transitions by reducing turn pause times collaboratively. Between-turn pauses have been found to distinguish learners at different proficiency levels (Galaczi, 2014) and to influence perceptions of fluency in L2 peer interaction (Sato, 2014) and in experimental settings involving native and non-native speech (Van Os *et al.*, 2020).

The second aspect, cohesive devices, involves different means of creating links across turns. Thus, it captures how the participants acknowledge the interlocutor's contributions and co-construct fluency across turns with lexical means (McCarthy, 2010). Two potential indicators of this dimension, other-repetitions and collaborative completions (see Peltonen, 2017), demonstrate acknowledgment of the interlocutor's contribution and extend beyond a single speaker's contribution while also creating cohesive flow in the interaction (e.g., Lerner, 1996; Tannen, 1989; see also Costa *et al.*, 2008). Both turn pausing and linking across turns are inherently collaborative phenomena because they entail attuning to the other speaker's contributions. Thus, in contrast to within-turn, individual fluency, they can be viewed as aspects of interactional fluency.¹⁷ In addition to examining linking across

¹⁷ The dimensions and measures presented here should be regarded as forming a starting point for exploring fluency in dialogue data from the perspective of the participants' collaboration: They are thus exploratory in nature and should be confirmed by future studies. Furthermore, while it is useful to treat individual and interactional fluency as separate concepts (Sato, 2014; Tavakoli, 2016), they are likely to be intertwined to some extent in practice. For instance, collaborative completions contribute to interactional fluency by maintaining flow across turns, but they also help in maintaining

turns with other-repetitions and collaborative completions, as is done in the present study, indicators of “relexicalization” (McCarthy, 1998), which refers to the use of near-synonyms to link turns (pp. 112–116) and is thus related to (direct) repetition, could also be explored as features of interactional fluency. In addition, McCarthy (2010) suggests that turn openings, creating links to what has been said previously, and turn closings, including items triggering speaker change, could be examined as potential indicators of confluence or interactional fluency. Both relexicalization and turn-opening/-closing items, similar to other-repetitions and collaborative completions, entail acknowledgment of the interlocutor’s contributions.

In future studies on interactional fluency, the role of *listening fluency*¹⁸ (see Anckar & Veivo, 2020) or *interactive listening* as a domain of IC (Galaczi & Taylor, 2018; see also Lam 2018, 2019) should also be considered, because participants in interaction are engaged in both listening and speaking, alternating between productive and receptive skills (see also Guillot, 1999). While the aspects suggested here (especially related to linking across turns) as candidates for interactional fluency indicators focus on productive skills, they also involve receptive skills: To be able to react in a manner that links one’s own contribution to the interlocutor’s contribution, the participant has to have processed what the interlocutor has said. Therefore, phenomena such as other-repetitions and collaborative completions can be viewed as parallel to some of the aspects that have been examined in the IC literature from a listening perspective. For instance, the contingent responses produced by the participants in Lam’s (2018) study (see also Lam, 2019), including formulating the previous speaker’s contributions, demonstrate attentiveness to previous talk and the ability to link one’s own contributions to previous turns, much like other-repetitions and collaborative completions, and could thus be viewed as reflecting listening fluency in an interactional setting.

To conclude, as L2 speech fluency studies are increasingly being conducted using interactional data along with the “mainstream” studies using monologue data, there is a need to develop approaches and measures to capture the interactional aspect of fluency in dialogic contexts. In the present study, it is suggested that both

individual, within-turn flow (Peltonen, 2017). Furthermore, from an individual, psycholinguistic perspective, other-repetitions in particular can help in coping with processing time pressure by freeing processing resources, thus facilitating individual fluency (Dörnyei & Kormos, 1998; Pickering & Garrod, 2004; Tannen, 1989).

¹⁸ Compared to speaking fluency, listening fluency as a term is not equally well established. Listening fluency can be viewed from a processing perspective (including speed and accuracy of processing) or in terms of the result of the process (accuracy of comprehension; Anckar & Veivo, 2020); from an interactional perspective, the main interest is in participants’ indicators of understanding in interaction (cf. the analysis of utterance fluency features from spoken data).

individual contributions and individual performance (within-turn aspects, i.e., individual fluency), as well as interactional, collaborative phenomena (between-turn aspects, i.e., interactional fluency), can and should be examined from L2 dialogue data to form a comprehensive picture of fluency. As has recently been argued by Roever and Kasper (2018) in the context of L2 assessment and IC, psycholinguistic–individualist and sociolinguistic–interactional perspectives of L2 performance “are not necessarily mutually exclusive” (p. 332; see also Galaczi & Taylor, 2018; McNamara, 1997). Similarly, interactional and individual approaches to L2 speech fluency should be regarded as providing complementary perspectives to the same communicative situation.

4 Data and methods

In this chapter, the data and methods of the present study are discussed. In Section 4.1, the participants are presented, followed by the details of the data and data collection procedures in Section 4.2. Measures and analyses are discussed in Sections 4.3 and 4.4, respectively. Two different data sets are presented, because the CAF corpus, compiled in the Department of English at the University of Turku (2019), was used in Article I, and a new data set was collected for Articles II–IV.

All students participated in the data collections voluntarily, and signed informed consent was obtained from all participants. The participants who had turned 18 signed the consent forms themselves, and in the case of underage participants, written permission to participate in the data collection was obtained from the subjects' guardians. The participants were informed of the purpose of the data collection and that their anonymity would be secured in all research outputs. The participants' anonymity was protected by replacing their names with subject codes in the background information data and using these codes in the research reports. The participants were also informed that their performance would not influence their English grades and that there were no right or wrong answers in the tasks. The participants were also given the opportunity to cancel their participation by contacting the researchers. In addition to the permissions obtained from the participants, the schools' principals were contacted for permission to organize the data collection at the schools with the help of English teachers. The participating teachers gave their permission for data collection in their classes and assisted with the practical arrangements.

4.1 Participants

The participants in **Article I** were chosen from the CAF corpus, compiled in the Department of English at the University of Turku (2019). I participated in the data collection for the corpus in 2012 and 2013. Altogether, 50 participants were chosen to be included in the study, representing two L1s (Finnish and Swedish) and two school levels (upper secondary school and university). In addition to the 40 learners of English, 10 native speakers were included as a control group (5 females, 5 males; average age 25.7 years). In the subject selection, participants with impairments

related to language production or relatively long stays in English-speaking countries (more than one month) were not included in the sample to maximize the within-group homogeneity of the learner groups regarding their L2 English proficiency.

The 40 selected learners were divided into four groups, each consisting of 10 subjects, as follows: Finnish-speaking first-year upper secondary school students (6 females, 4 males), Swedish-speaking first-year upper secondary school students (7 females, 3 males), university students of English from the University of Turku (L1 Finnish; 9 females, 1 male), and university students of English from Åbo Akademi University (L1 Swedish; 5 females, 5 males). The upper secondary schools and the universities were located in Southwest Finland. The upper secondary school subjects were on average 16.1 years old, while the university subjects were on average 20.9 years old.

The subjects' backgrounds regarding their duration of studies in English at their respective school levels were relatively similar: Finnish L1 upper secondary school subjects had studied English for an average of 7.0 years in basic education and the Swedish L1 subjects for approximately 6.67 years. Furthermore, the distribution of English grades in both upper secondary school groups was similar (L1 Finnish mean 7.9; L1 Swedish mean 8.1). While the CAF corpus does not contain separate proficiency-level test information for the subjects (cf. data collection for Articles II–IV), the goal in English for an average student is to reach the lower range of the B2 level (B2.1; level B indicating independent user) in the CEFR (Council of Europe, 2001) by the end of upper secondary school (Finnish National Agency for Education 2019, p. 177). At the university level, the Swedish L1 subjects had studied Finnish for 10–11 years and English for an average of 8.6 years, while the Finnish L1 subjects had studied English for slightly longer ($M = 10.0$ years). While there is no official CEFR goal for university studies in English in Finland, university students are expected to represent levels C1/C2 in the CEFR (level C indicating proficient user), having passed a competitive entrance exam to study English. Furthermore, they form a select group of upper secondary school graduates. The selection of subjects from different school levels, with differing amounts of instruction in English and the estimated CEFR level, was regarded as sufficient for ensuring that the school-level groups on average represented different proficiency levels.

As noted previously, for Articles II–IV, a new data set was collected (for details, see Section 4.2). In **Article II**, in which the subjects' performance during a pair task was analyzed, 42 subjects from two school levels (final, ninth grade of compulsory education and the second year of upper secondary school) were included from the total of 50 participants (20 ninth graders, 30 upper secondary school students). To ensure the relative homogeneity of the school-level groups, four subjects and their interlocutors were excluded from the sample due to background factors that could have influenced their oral proficiency level in English (one subject due to an

unspecified language impairment; three subjects due to stays abroad that lasted over eight weeks).¹⁹ The ninth-grade group consisted of 16 subjects (8 pairs; 10 males, 6 females; *M* age = 15.00 years), and the upper secondary school group had 26 subjects (13 pairs; 17 males, 9 females; *M* age = 17.35 years). All subjects spoke Finnish as their L1, with the exception of one subject who reported Swedish as his L1 and Finnish as his home language. One subject reported additional languages spoken at home, and one identified as bilingual (Spanish as the other L1). Ninth graders represented the whole age group, whereas upper secondary school students formed a selected group of students, having completed their compulsory education. (Note that some students attend vocational education instead of upper secondary school after compulsory education.²⁰)

As with subject selection in Article I, the groups in Article II differed in the number of years they had studied English: The ninth graders had studied English for an average of 6.06 years and upper secondary school students for approximately two years longer (*M* = 8.58 years). Furthermore, the results of a vocabulary test providing estimates of the subjects' overall proficiency in English (LexTALE) confirmed that the upper secondary school students represented a higher proficiency level on average (*M* score = 72.89%; approximately CEFR level B2) than the ninth graders (*M* score = 59.38%; approximately CEFR level B1; for the CEFR level estimates, see Lemhöfer & Broersma, 2012, p. 341). The difference in the mean scores was statistically significant ($t(40) = 3.98, p < .001$), but it should be noted that within-group variation was quite large (ninth grade *SD* = 10.83; upper secondary school *SD* = 10.58). Nevertheless, the fact that the subjects came from different school levels, with a difference in the number of years they had studied English, as well as in their overall proficiency levels, was regarded as sufficient for considering the groups as representing different proficiency levels on average.²¹

The **Article III** subjects ($n = 42$) came from the same pool of 50 participants as the subjects in Article II, but the focus was on the analysis of the participants' L1 Finnish and L2 English monologue productions. Altogether, eight subjects were

¹⁹ Participation in content and language integrated learning (CLIL) during compulsory education was also considered a potential influencing factor. However, four upper secondary school participants who had participated in CLIL did not differ from the overall upper secondary group on any fluency or CS measure according to Mann-Whitney U-tests, so they were not excluded from the study.

²⁰ For a graph about the Finnish school system, see Ministry of Education and Culture (2018, p. 3).

²¹ The students also provided self-evaluations of their English skills as a part of their background information, and in an unpublished analysis, the upper secondary school subjects were found to rate their productive skills (discussion, speaking, and writing) in L2 English statistically significantly higher than ninth graders, further corroborating the differences between the two groups.

excluded from the study: In addition to the four subjects excluded from Article II, one subject was excluded due to having a different L1 (Swedish) and three subjects due to their speech samples being very short in the L1 or L2 or both and thus preventing reliable analyses of their speech. Similar to Article II, the 42 subjects included in the study represented two school levels: 17 ninth-grade subjects (11 males, 6 females; M age 15.00 years) and 25 second-year upper secondary school subjects (14 males, 11 females; M age 17.24 years). All subjects reported Finnish as their L1. Similar to the subjects in Article II, the subjects in Article III differed in the number of years they had studied English (ninth graders $M = 6.06$ years; upper secondary school group $M = 8.52$) and in their estimated overall proficiency in English; the ninth graders represented CEFR level B1 on average (LexTALE M score = 59.63%, $SD = 10.53$) and the upper secondary school students level B2 on average (M score = 73.40%, $SD = 10.23$), with the difference in the mean LexTALE scores being statistically significant ($t(40) = 4.23, p < .001, d = 1.33$).

For **Article IV**, which focused on examining the connections across rated and measured fluency, six samples of peer interaction (from 12 participants) from the pool of 25 samples were chosen to be rated by a group of judges. The use of six samples was regarded as sufficient for ensuring variation across the samples, and limiting the number of samples was considered important to avoid rater fatigue. According to the subjects' LexTALE scores, CEFR levels B1, B2, and C1/C2 were represented in the sample: Two pairs represented the upper range of B1/lower range of B2 (scores 56.25%–67.50%), three pairs represented the B2 level (scores 63.75%–75.00%), and one pair represented levels B2 and C1–C2 (scores 72.50% and 96.50%). In addition to ensuring sufficient variation among the chosen samples based on the LexTALE scores²² (see also, e.g., Cucchiari *et al.*, 2002), the following criteria were applied when choosing the samples to be rated: sufficient talk from both participants (more than 100 syllables during the interaction), relatively symmetrical interaction (based on my estimate), and the use of CSs during the interaction to ensure that the raters could reliably assess the “strategic competence” dimension.

In Article IV, the ratings were completed as part of an MA-level course on language assessment.²³ The group of raters consisted of 15 advanced students of English at a Finnish university (13 females, 2 males; M age = 26 years; L1 Finnish).

²² Variation across the pairs regarding their proficiency was considered in sample selection to ensure that the raters would detect differences across the samples based on the 9-point scales and thus assess the samples reliably.

²³ The course included a theoretically oriented component (lectures) and practical assessment exercises. During the course, the participants familiarized themselves, for instance, with the CEFR (Council of Europe, 2001).

The average duration of the subjects' studies in English at the university was 4.40 years ($SD = 1.30$) and 10.03 years ($SD = 1.32$) at school before university studies. For 12 of the subjects, English was their major. All subjects reported normal hearing. The raters reported having relatively little experience in spoken language assessment prior to the language assessment course; yet, 12 of the subjects had participated in teacher training during their studies and had thus gained some practice in language assessment. Including students of English (many of them preservice teachers) as raters in the study was in line with the practice adopted in some other L2 fluency studies that have used less experienced raters and/or L2 speakers as raters (e.g., Bosker *et al.*, 2013; Derwing *et al.*, 2004; Magne *et al.*, 2019; Rossiter, 2009), as opposed to the more commonly used rater populations, such as native speakers or practicing teachers. The choice of the rater population was motivated by the exploratory nature of the study: Especially for the qualitative analysis, a broad range of interpretations of the assessed dimensions was preferred, and the range was expected to be broader among inexperienced than experienced raters (for details, see Article IV).²⁴ Yet, the subjects' similar backgrounds as advanced university students of English and their interests in language assessment yielded sufficiently homogenous assessments of the samples and ensured acceptable inter-rater reliability (cf. Rossiter's 2009 findings of minor differences across different rater groups; see Article IV for inter-rater reliability statistics).

4.2 Data and procedure

For **Article I**, the data were selected from the CAF corpus (see Section 4.1). While the corpus contains different types of L2 production data, including written samples, monologue speech samples were used in the present study. The spoken monologue productions consisted of the participant describing a cartoon strip consisting of six frames. The cartoon strip depicted a boy and a girl growing up, reflected in a growing tree. Similar picture narration tasks have been commonly used in L2 speech fluency studies (e.g., Kormos & Dénes, 2004; Lennon, 1990). Furthermore, standardizing the prompt with a set of pictures ensures comparability across the speech samples (see also Segalowitz, 2010). The research assistants, including myself, followed the same standardized data collection procedure in different schools. During the collection of the monologue speech samples, the participants were first given two minutes of individual planning time. After the planning time, the participants described the story depicted in the cartoon strip in their own words in English. The participants were allowed to look at the prompt during the task to exclude the

²⁴ For a similar approach involving untrained raters and minimal instructions, see Jarvis (2017), who explores raters' intuitions regarding lexical diversity.

influence of a potential intervening factor (retrieval from memory) on task performance. The tasks were audio-recorded in a quiet room during regular school days with no additional people present.

The selected CAF corpus speech samples were transcribed and double-checked by a group of research assistants, including myself, for Article I. Silent pauses of 0.40 seconds²⁵ or longer were manually identified with the help of a waveform in audio editing programs (Amadeus Lite for Mac and Transcriber for Windows). Compared to automatized pause identification, the manual process enabled us to distinguish acoustic silences from silent pauses, as well as to identify SPs with paralinguistic noises (e.g., lip smacks). It was also suitable for the data set, as some of the samples contained background noise.²⁶

For **Articles II–IV**, a new data set was collected (see Figure 4 on the following page) to further explore promising aspects identified in Article I—notably the different ways to maintain fluency or strategic aspects of fluency. The new data set also enabled the study of interactional fluency, which was not possible with the monologue speech samples included in the CAF corpus. Finally, the new data set focused on intermediate-level learners (ninth graders and upper secondary school students), excluding the more proficient subjects (university students) examined in Article I. By focusing on intermediate-level learners, more variation in fluency and the used fluency resources was expected compared to university students.

²⁵ A slightly lower pause cut-off was applied in Articles II–IV based on De Jong and Bosker's (2013) recommendation.

²⁶ In subsequent articles (II–IV), a combination of automatized and manual pause identification was applied.

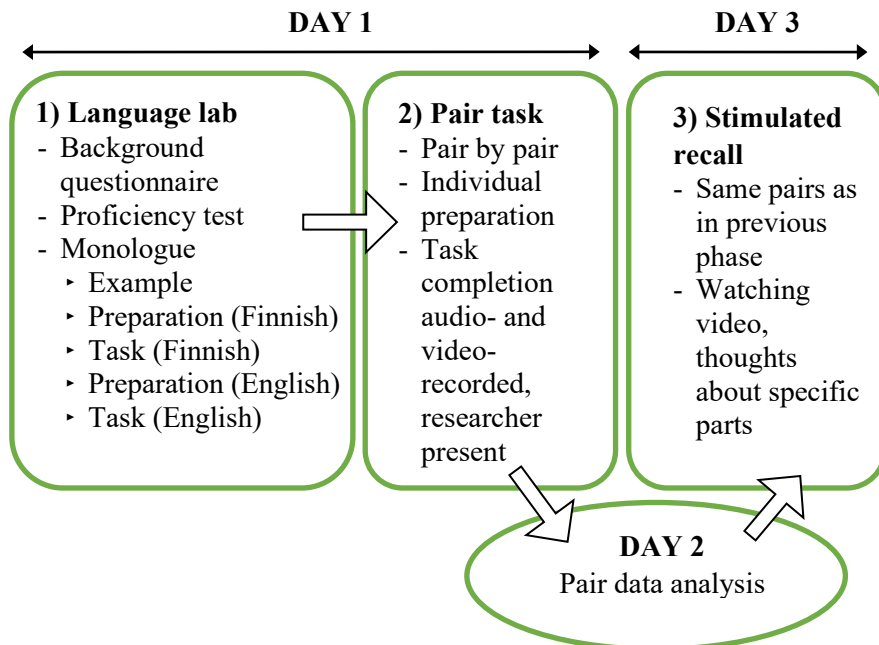


Figure 4. Phases of data collection (Articles II–IV).

I designed the data collection and collected the new data set with the help of research assistants in December 2014 and from March to April 2015. Before the data collection, the materials were piloted in November 2014 with six ninth graders who were not included among the actual study participants. The pilot was successful, and only minor modifications were made to the data collection materials. The data collection was organized at two schools (participants from several English classes) in the Uusimaa region and in Southwest Finland. The data were collected during regular school days in the schools' language laboratories (monologue data collection) and classrooms (dialogue data collection and stimulated recall protocols) according to the procedure described in Figure 4.

The data collection started in a language lab, with the participants completing a background questionnaire form and the LexTALE proficiency test (in pen-and-paper format). The background questionnaire included questions about the participants' previous language studies, visits abroad, and language impairments, among other relevant background factors, as well as a short self-reported English skills component (see Section 4.1). Following a brief introduction to the monologue tasks, the participants spent two minutes preparing for their first monologue picture description task (pictures in Appendix 1). The participants described a picture in their L1, Finnish, while looking at it. The same procedure, with two minutes of planning

time followed by a description of a (different) picture, was repeated with the participants describing the second picture in their L2, English.

As described in Section 4.2, monologue picture description tasks have often been used in fluency studies because the standardized input ensures similar content and comparability across samples (see also Segalowitz, 2010); yet, the production is still less constrained than, for instance, in reading tasks. For the data collection, two different yet maximally similar pictures were used. The task type and instructions were the same for both pictures: The participants were asked to describe the people, surroundings, and events in the pictures. Both pictures involved narrative elements, as ensured by the interactions between different elements in the pictures. The use of similar tasks in different languages (instead of repeating the description of one picture in different languages) was preferred to ensure the participants' motivation across the tasks and separate content planning for the two tasks to avoid transfer effects (see also De Jong *et al.*, 2015; Huensch & Tracy-Ventura, 2017a).²⁷

In the second phase of the data collection (on the same day), the participants completed a dialogue task. The dialogue task consisted of the participants completing a problem-solving activity in pairs.²⁸ The activity can be characterized as a communicative problem-solving task or a consensus task, as the participants were instructed to discuss 16 objects and to rank the objects in the order of usefulness for survival on a desert island (see Appendix 2). The instructions were created based on Klippel (1984, pp. 63–64) and Ur (1990, pp. 70–72). The objects included two types of items: those that the participants were likely to be able to name in English (e.g., umbrella) and those that were less familiar and more difficult to name in English (e.g., flares) to ensure the use of CSs and an appropriate level of difficulty in the task. Ninth-grade and upper secondary school textbooks and one of the teachers of the ninth-grade participants were also consulted to ensure that the level of the vocabulary was suitable for the participants.

The participants had two minutes of individual preparation time for the task, followed by six minutes for completing it in pairs. The tasks were completed in a quiet space. In addition to the pair, only the researcher (myself) was present in the room to audio- and video-record the task. I did not intervene in the interaction, except

²⁷ Despite these measures taken to ensure comparability across the monologue tasks, it should be acknowledged that a potential limitation of this design is the unknown role of picture content in influencing the productions due to the lack of counterbalancing.

²⁸ To ensure that there would be no major L2 proficiency-level differences within the pairs that could have influenced the performance (see, e.g., Csépes, 2009; Davis, 2009; Iwashita, 1996), they were formed in collaboration with the learners' teachers based on their estimations of the subjects' skills in English. In addition to acknowledging L2 proficiency level as a potential interlocutor effect (see, e.g., Taylor & Wigglesworth, 2009), age, L1, and acquaintanceship were also controlled in the present study.

to announce when the participants had one minute left to complete the task. Learner–learner pairs (as opposed to, e.g., researcher–learner pairs with standardized prompts) were used to ensure relatively symmetrical interaction and to encourage genuine two-way interaction and collaboration (cf. information gap tasks that can generate one-sided interaction and less varied and elaborate turns; e.g., Foster, 1998; Nakahama, Tyler, & van Lier, 2001). Overall, the task type was chosen as the most suitable for the purposes of the present study due to its relatively structured nature (ensuring comparability across samples; cf. e.g., free conversation), genuine purpose for talk (solving a problem collaboratively), inclusion of difficult elements (enabling the study of CSs), and opportunity for collaboration.

After the monologue and dialogue data collection, I reviewed the dialogue data and prepared for the third phase of data collection (a stimulated recall session), which was undertaken one day after the first two phases.²⁹ The participants were not informed about the stimulated recall session beforehand—only that data collection would continue after the first two phases. The purpose of the stimulated recall session was to gain further insights into the participants' CS use; the stimulated recall reports were consulted when identifying and classifying CSs based on the speech samples (Article II; see Kahng, 2014; Kormos, 1998). In other words, the data collection methods (production data and stimulated recall protocols) were triangulated to enhance the reliability of CS analysis.

During the stimulated recall session, the pairs watched the video recording of their task performance with me. Following Kormos's (1998) and Gass and Mackey's (2000) recommendation, the session was conducted in the participants' L1 (Finnish) to ensure that their proficiency level would not influence their reports (see also Kahng, 2014; Poulisse, Bongaerts, & Kellerman, 1987). Gass and Mackey's (2000) guidelines for stimulated recall were followed: Importantly, the participants were instructed to report any thoughts they recalled having *when performing the task* (not opinions or retrospective interpretations). If the participants recalled something, they could stop the video themselves; occasionally, I also stopped the video at potentially

²⁹ While it is generally recommended that the stimulated recall sessions be organized as soon as possible after task completion (e.g., Ericsson & Simon, 1993; Gass & Mackey, 2000), preferably immediately following the task (see, e.g., Kahng, 2014; Poulisse *et al.*, 1987), for the purpose of the present study, the opportunity to familiarize myself with the data and to preliminarily identify some problem-solving sequences in the interaction to prepare for the sessions was prioritized over the immediate review of the data with the participants (see also Dobao, 2001). The sessions were thus organized on the third day of data collection, with the second day being dedicated to my preliminary analysis of the data (see Figure 4). The purpose of the analysis was to ensure that all parts of the videos related to problem-solving would be covered during the stimulated recall sessions, especially because the pilot had demonstrated that the subjects often did not stop the video themselves (yet did provide comments when I stopped the video).

interesting places, identified beforehand, and asked whether the participants wanted to comment on anything. The questions were formulated in as general and neutral manner as possible (e.g., “What did you think about here?”) so as not to guide the participants’ thoughts toward certain answers (see, e.g., Ericsson & Simon, 1993).

After data collection, all monologue and dialogue samples were transcribed and double-checked with the help of research assistants. The assistants were given detailed instructions, but I was also available for any questions or consultations regarding potentially difficult cases. The assistants followed a slightly modified version of the transcription guidelines developed during the CAF corpus compilation (e.g., with additional instructions related to dialogue transcription). The set of conventions was influenced by conversation analytical transcription conventions. After the transcription phase, I identified silent pauses of 0.25 seconds or longer (De Jong & Bosker, 2013) in the speech analysis software Praat (Boersma & Weenink, 2019) using a script (De Jong & Wempe, 2009), followed by manual checks and adjustments. Shorter pauses were regarded as micropauses and were not taken into account in subsequent analyses. An upper limit of three seconds for SPs was applied (Witton-Davies, 2014), and pause times exceeding three seconds were excluded from the total sample duration calculations. A script calculating the total durations and frequencies of labeled intervals (Lennes, 2002) based on Praat annotations was used to extract information (e.g., speaking time and pause time at different locations) for temporal fluency measures (see Section 4.3).

In addition to the utterance fluency data (monologue and dialogue speech samples), I collected a complementary data set for examining perceived fluency in Article IV in the form of fluency ratings in November 2016. The ratings of six pair tasks were completed as part of an MA-level course on language assessment, as mentioned in Section 4.1. The raters had two weeks to complete their ratings on their own time. To secure the participants’ anonymity, anonymized versions of pair task videos were used, and the background information given to the raters was limited to the participants’ school level (two pairs from ninth grade and four pairs from upper secondary school). While the ratings were mainly completed based on the videos, the raters were given transcriptions of the interactions to, for instance, enable them to check any unclear parts of the videos. The raters were instructed to assess all six samples on 9-point scales (based on, e.g., Bosker *et al.*, 2013; Cucchiarini *et al.*, 2002; Derwing *et al.*, 2004; Götz, 2013; Rossiter 2009) for four dimensions. In addition to a holistic rating of oral proficiency, the raters evaluated the samples according to three analytic dimensions—(individual) fluency, interactional fluency, and strategic competence—which reflected the Fluency Resources Framework (see Section 3.2). For the analytic ratings, the rating sheet (see Appendix 3) contained a separate space for comments about the aspects to which the raters paid attention during the assessment process (see also Suzuki & Kormos, 2020).

Instead of providing descriptors to the raters (e.g., based on the CEFR), only brief instructions for assessing each dimension were given to the raters to capture their understanding and interpretation of each concept (see also Suzuki & Kormos, 2020, for a similar approach). For (individual) fluency, the raters were instructed to focus on the dimensions of speed, pausing, and repair in their assessments (following Skehan, 2009) to capture fluency in the narrow sense (as opposed to fluency in the broad sense; following, e.g., Bosker *et al.*, 2013; Derwing *et al.*, 2004, Rossiter, 2009). For interactional fluency, no additional instructions were given. In contrast to individual fluency and strategic competence, a joint score was given to the pair to reflect the co-constructed and collaborative nature of interactional fluency (see also May, 2009; Taylor & Wigglesworth, 2009). For strategic competence, the instructions guided the raters toward a relatively narrow understanding of the construct, with a focus on CSs, to reflect the Fluency Resources Framework.

4.3 Measures

A set of commonly used, reliable fluency measures was compiled for fluency analyses in Articles I–IV based on earlier influential fluency studies (see Table 2). The measures have been found to demonstrate fluency development (e.g., Towell *et al.*, 1996) and to distinguish subjects according to their fluency (e.g., Cucchiaroni *et al.*, 2002; De Jong *et al.*, 2015; Götz, 2013; Kormos & Dénes, 2004; Riggenbach, 1991; see also Section 2.1). As has been common in L2 speech fluency studies, the measures were chosen to represent different dimensions: In Article I, Skehan's (2003, 2009, 2014) division into speed, breakdown, and repair was followed, whereas in Articles II–IV, the measures reflected the Fluency Resources Framework. In Article I, all frequency measures were standardized by expressing them per minute of total time (following, e.g., Kormos, 2006, p. 163). In Articles II–IV, following a more recent recommendation (De Jong, 2016b), the frequency measures were standardized per minute of speaking time, thus excluding the influence of SPs on the measures. The frequency of a feature was thus divided by speaking time (time spent talking, excluding SPs) and multiplied by 60 to obtain a measure per minute instead of per second (De Jong, 2016b).

Table 2. Fluency measures, their operationalizations, their use in different articles, and links to Skehan's dimensions and the Fluency Resources Framework categories.

Measure	Operationalization	Article			Skehan	TFRF
		I	II + IV	III		
Speech rate	Total number of syllables per minute of total time (in Articles II and IV, excluding turn pauses [TPs]).	✓	✓	✓	C	TF
Articulation rate	Total number of syllables per minute of speaking time (excluding SPs).	✓	✓	✓	S	TF
Mean length of run	Syllables per number of runs (a run = a stretch of speech occurring between SPs).	✓		✓	C	TF
Mean length of turn	Syllables per number of turns (following Nitta & Nakatsuhara, 2014, p. 155).		✓		C	TF
Phonation–time ratio	Total speaking time as percentage proportion of total length of sample (total speaking time divided by sample length).	✓			C	TF
Number of silent pauses (SPs)	Number of SPs (minimum duration in Article I = 0.40 seconds, in Articles II–IV = 0.25 seconds).	✓	✓	✓	B	TF
Number of mid-clause SPs	Number of SPs located mid-clause.	✓	✓	✓	B	TF
Number of clause boundary/AS-unit SPs	Number of SPs located at independent clause boundaries or at AS-unit boundaries (see the definition for AS-unit below). Note that this measure was not reported in Article IV.	✓	✓	✓	B	TF
Number of turn pauses (TPs)	All pauses between turns were regarded as shared TPs. Adjacency pairs (e.g., questions) were considered an exception, and a pause after the first pair part was analyzed as belonging to the interlocutor (see Sacks <i>et al.</i> , 1974, pp. 714–715). The number of TPs was divided equally between the participants (Witton-Davies, 2014).		✓		B	IF

		Article			Skehan	TFRF
Measure	Operationalization	I	II + IV	III		
Mean length of SPs	Total length of SPs divided by total number of SPs.	✓	✓	✓	B	TF
Mean length of mid-clause SPs	Total length of mid-clause SPs divided by total number of mid-clause SPs.			✓	B	TF
Mean length of clause boundary SPs	Total length of clause boundary SPs divided by total number of clause boundary SPs.			✓	B	TF
Mean length of turn pauses	Total length of TPs divided by total number of TPs. In addition to TPs, all clause boundary pauses lasting over 2 seconds were considered TPs and the responsibility of both speakers (Witton-Davies, 2014).		✓		B	IF
Number of other-repetitions	Number of words or longer stretches of interlocutor's speech that were repeated without modification.		✓		-	IF
Number of collaborative completions	Number of completions of an utterance that previous speaker has started (e.g., Lerner, 1996; Dings, 2014).		✓		-	IF
Number of hesitation groups	Number of clusters of at least two co-occurring hesitations: SPs, FPs, and/or repairs. At least one feature (SP or FP) lasted 0.4 seconds or longer.	✓			-	-
Number of filled pauses (FPs)	Number of non-lexicalized FPs (e.g., <i>uh</i> and <i>um</i>).	✓	✓	✓	B	FR/SM
Number of drawls	Number of vowel elongations of 0.30 seconds or longer (e.g., Witton-Davies, 2014).		✓	✓	-	FR/SM

Measure	Operationalization	Article			Skehan	TFRF
		I	II + IV	III		
Number of fillers	Number of the following lexicalized pauses (Dörnyei & Kormos, 1998): - discourse markers (<i>like, you know, well, and I mean</i>) - smallwords (<i>sort of/sorta, kind of/kinda, and quite</i>) (Götz, 2013).	✓	✓	✓	-	FR/SM
Number of repetitions	Number of words or longer utterances repeated without modification.	✓	✓	✓	R	FR/SM
Number of false starts	Number of rejected, cut-off sounds, words, or longer utterances.	✓			R	-
Number of replacements	Number of words replaced with another word without additional modifications.	✓			R	-
Number of reformulations	Number of utterances longer than one word repeated with some modification.	✓			R	-
Number of paraphrases	See subcategories below.		✓		-	FR/CS
- Number of circumlocutions	Number of illustrations or descriptions based on the characteristics of the target object or action.		✓		-	FR/CS
- Number of approximations	Number of alternative lexical items (e.g., superordinate/related term) that share semantic features with the target word or structure.		✓		-	FR/CS
- Number of all-purpose words	Number of general, “empty” lexical items in contexts in which specific words are lacking.		✓		-	FR/CS
- Number of other strategies	Number of strategies in the following subcategories: word coinage, use of similar-sounding words, restructuring, and tip-of-the-tongue phenomenon.		✓		-	FR/CS

Measure	Operationalization	Article			Skehan	TFRF
		I	II + IV	III		
Number of transfer strategies	Number of strategies in the following subcategories: <i>literal translation</i> (of a word or structure from L1 to L2) and <i>code-switching</i> (including L1 words with L1 pronunciation in L2 speech).		✓		-	FR/CS
Number of appeals for assistance	Number of instances in which help was elicited from the interlocutor directly (asking a question) or indirectly (expressing the lack of a vocabulary item and gazing toward interlocutor; gazing away from interlocutor was regarded as a solitary word search—not a request of help; Goodwin & Goodwin, 1986).		✓		-	FR/CS

Note: Skehan’s dimensions: S = speed, B = breakdown, R = repair, C = composite; the Fluency Resources Framework (TFRF) categories: TF = temporal fluency, IF = interactional fluency, FR/SM = fluency resources 1: stalling mechanisms, FR/CS = fluency resources 2: communication strategies.

The Analysis of Speech Unit (AS-unit; Foster *et al.*, 2000) was used as the basis for coding silent pause location in Articles I–IV (see also, e.g., De Jong *et al.*, 2015; Huensch & Tracy-Ventura 2017a, 2017b; Kahng, 2014, 2018). Foster *et al.* (2000, p. 365) defined the AS-unit as “consisting of *an independent clause, or sub-clausal unit, together with any subordinate clause(s) associated with either*” (emphasis in the original). SPs were thus coded as occurring either mid-clause or at clause/AS-unit boundaries. In Articles II and IV, in which dialogue data were analyzed, (between-)turn³⁰ pauses were also examined separately. The frequencies of SPs overall and in different locations were examined in all articles. The duration of SPs at specific locations was examined in Article III to enable comparisons with other recent articles examining the connections between L1 and L2 (monologue) speech

³⁰ The speaker’s total contribution before the other speaker’s contribution (before a speaker change) was regarded as a turn. Backchannels (e.g., laughter, *mm*, and *yeah*) in the middle of the other speaker’s turn were not considered to cut it off nor to form a turn of their own. In other locations, these were counted as short constructional units and considered as forming a turn of their own. In the case of overlapping and parallel turns, intonation and syntax were used to disambiguate where a turn ended (see also Witton-Davies, 2014).

fluency (De Jong *et al.*, 2015; Huensch & Tracy-Ventura, 2017a).³¹ Thus, overall, SPs were examined comprehensively from the perspectives of frequency, duration, and location in the present study.

Some temporal fluency measures that were used only in Article I were excluded from subsequent articles. To avoid overlap between measures, phonation–time ratio was excluded, as it provides information about the relationship between speech and pausing (and only rather global information about pausing; see also De Jong, 2016b, p. 211), similar to a more widely used measure—mean length of run. Only MLR, not PTR, was thus used in Article III and in a modified form in Articles II and IV as mean length of turn (MLT; referred to as *average turn length* in Nitta & Nakatsuhara, 2014) to capture how long the learners could maintain their speech flow in a dialogue context. Furthermore, the exploratory measure of hesitation groups, which was developed to capture co-occurring hesitations in Article I (based on Hilton, 2009; see also Freed, 1995; Riggenbach, 1991), was not used in subsequent articles. While based on previous studies, the analysis in Article I, which highlighted the importance of analyzing the different functions and contexts of hesitations, provided little support for the use of a measure that groups different hesitations in subsequent studies. Instead, it is more useful to examine hesitations separately by type (see also qualitative analysis in Section 4.4.2). Finally, while in Article I, four measures were used to examine the repair dimension (based on Foster & Skehan, 1999), in Articles II–IV, only repetitions of the original four repair measures were included (as stalling mechanisms in the Fluency Resources Framework). The other measures were excluded from Articles II–IV because self-correction was viewed as more closely connected to accuracy and the process of monitoring (e.g., Dörnyei & Kormos, 1998; see also Section 2.2).

In addition to the FPs and repetitions that were already included in Article I, two new measures for other potential stalling mechanisms were included in Articles II–IV as part of the Fluency Resources Framework. Based on Dörnyei and Kormos (1998), *drawls* and *fillers* (see also Götz, 2013) were viewed as fulfilling a similar function in obtaining more planning time while maintaining the flow of speech as FPs and repetitions. In addition to stalling mechanisms, CSs were examined as another type of problem-solving mechanism in Articles II and IV using dialogue data.³² The CS types and their definitions were based on previously established and widely used categorizations (notably Dörnyei & Scott, 1997; see also Dobao, 2001; Færch & Kasper, 1983; Montero, Serrano, & Llanes, 2017). CSs were identified

³¹ The mean durations of SPs and TPs were nevertheless calculated separately in Articles II and IV.

³² As the monologues did not contain as frequent CS use as the dialogues, CSs were excluded from Article III to ensure reliable analyses.

based on strategy markers in the dialogue data and the information provided in the stimulated recall protocols (following, e.g., Dobao, 2001; Kasper & Kellerman, 1997; Nakatani & Goh, 2007). Three main categories of achievement strategies (Færch & Kasper, 1983) were examined: paraphrases, transfer, and appeals for assistance (based on, e.g., Dobao, 2001; see Table 2 for definitions). The categories cover the main achievement strategies studied in the CS literature (for an overview, see Dörnyei & Scott, 1997), and thus provided a well-established basis for choosing representative CS measures for the present study. Paraphrases contained four subcategories: In addition to the main categories—*circumlocution*, *approximation*, and *all-purpose words*—the *other strategies* category contained paraphrase strategy types with only a few occurrences of each type in the data.

In addition to examining stalling mechanisms and CSs along with temporal fluency measures, in Articles II and IV, four measures were used to capture dialogue or interactional fluency: the number of turn pauses (TPs), the mean length of TPs, the number of other-repetitions, and the number of collaborative completions. The first two measures relate to temporal interactional fluency and the latter two to cohesive devices (see Section 3.3). The first two pausing measures complement within-turn pause measures by providing information about “joint” pause time in dialogue. The two novel measures for examining cohesion across turns, other-repetitions and collaborative completions (for details, see Peltonen, 2017), were chosen based on previous studies having examined them as indicators of collaboration (e.g., Dings, 2014; Hüttner, 2009; Lerner, 1996; Riggenbach, 1991; Taguchi, 2014), as well as their prominence in the interactional data set of the present study. Other-repetitions refer simply to words or longer utterances of repeated interlocutor’s speech (without additional modifications), while collaborative completions (also known as, e.g., *anticipatory completions*; Lerner, 1996) complete an utterance started by the interlocutor (e.g., Dings, 2014).

A final detail related to the measures concerns Article III, in which cross-linguistic comparisons across Finnish (L1 monologues) and English (L2 monologues) were conducted. As Finnish is characterized by its rich inflectional morphology, many words being longer in Finnish than in English, words were considered a less suitable unit for cross-linguistic comparisons than syllables (see also Bradlow *et al.*, 2017; on temporal measures for L1 and L2 Finnish, see also Lehtonen, 1979; Penttilä *et al.*, 2018; Toivola *et al.*, 2009). Syllables were thus used, for instance, in speech rate and articulation rate calculations to ensure comparability across the samples in Article III. Syllables were also calculated in the other substudies, enabling comparisons of the results across articles.

4.4 Analyses

The articles of the present study incorporated both quantitative and qualitative elements. This section contains a presentation of the quantitative analyses in the articles (Section 4.4.1), followed by an outline of the mixed methodology and complementary qualitative analyses (Section 4.4.2).

4.4.1 Quantitative analyses

The quantitative analyses in the articles included group comparisons with analysis of variance (ANOVA) tests (Article I) and Mann-Whitney U-tests (Articles II and III) to examine whether the groups differed in terms of their fluency according to the fluency measures. Connections across fluency measures were examined with rank-order correlation coefficients (Spearman's rho) in Articles II–IV. Nonparametric correlations were preferred over parametric Pearson's correlations due to non-normal distributions in the data sets. The correlational analyses formed the basis for further regression analyses in Articles III. All statistical analyses were performed using SPSS.

In **Article I**, the quantitative research questions addressed the differences between L1 groups (Finnish and Swedish) and school-level groups (upper secondary school and university), as well as the differences between learner groups and the native control group. A one-way ANOVA (followed by Tukey's HSD post hoc tests) was used as the main method of comparing the differences across the five groups in terms of fluency measures. (A complementary two-way ANOVA was used to confirm the findings; for details, see Article I.) Before the ANOVA, the data were examined for normality using Shapiro-Wilk's tests of normality. For non-normally distributed variables, a log transformation was used. (For four measures that did not reach normal distribution after the transformation, the nonparametric independent samples Kruskal-Wallis test was used.)

In **Article II**, the two quantitative research questions addressed group-level differences in fluency between ninth graders and upper secondary school students, as well as correlations across temporal fluency measures and measures for fluency resources. The differences across the two school-level groups were examined using nonparametric Mann-Whitney U-tests (*r* effect sizes calculated based on Larson-Hall, 2010, pp. 377–378) because non-normal distributions for some fluency measures were discovered with Shapiro-Wilk's tests of normality. Due to the use of multiple U-tests, the false discovery rate (FDR) method was applied to control for potential false positives (type I error; see Larson-Hall, 2010, pp. 251–252). To examine the connections across fluency measures, rank-order correlation coefficients (Spearman's rho) were calculated.

In **Article III**, the quantitative research questions addressed the connections across L1 and L2 fluency measures (examined using correlations) and predicting L2 fluency measures based on L1 fluency measures (examined using regression analyses). In a preliminary analysis, potential group-level differences between the ninth-grade and upper secondary school groups in terms of their L1 and L2 fluency were examined with Mann-Whitney U-tests and r effect sizes. The nonparametric tests were used due to the Shapiro-Wilk tests of normality indicating non-normal distributions for several L1 and L2 fluency variables. As in Article II, eventual false positives were controlled using the FDR method. After the preliminary analysis of group-level differences, rank-order correlation coefficients were computed for the sample as a whole and for the two groups separately. Effect sizes were interpreted as follows: r_{ss} close to .25 = small, .40 = medium, and .60 = large (Plonsky & Oswald, 2014, p. 889). Following the correlational analyses, simple linear regressions (L2 fluency measures as dependent variables and L1 fluency measures as independent variables) were conducted for variables that had demonstrated statistically significant correlations. Hierarchical regressions were also performed to examine whether adding school level as an explanatory (dummy) variable in the analyses would increase the explained variance in the L2 fluency measures (L1 fluency measure added first and school level second). The assumptions of regression (linearity, normal distribution of errors, homogeneity of variances, and multicollinearity for hierarchical regression) were mostly met, but for variables for which clear nonlinearity, non-normal distribution of errors, or both nonlinearity and heterogeneity of variances were observed, x^2 was applied to the independent variables in the models.

In **Article IV**, the connections across rated and measured fluency were examined. The research questions addressed correlations across ratings of different aspects of fluency and correlations between fluency measures and fluency ratings. In contrast to Article III, which included both correlational and regression analyses, only correlations were performed in Article IV due to the limitations of the small sample size. Before the analyses related to the research questions, the reliability of the ratings was examined with intraclass correlation coefficients (two-way random model). Thereafter, rank-order correlation coefficients (Spearman's rho) were calculated to examine the connections across the rated aspects and across fluency measures and mean ratings. For the purpose of the analyses, the joint interactional fluency scores were treated as individual scores (both participants in a pair were coded as receiving the same score). Effect sizes were interpreted based on Plonsky and Oswald's (2014) previously mentioned guidelines.

4.4.2 The mixed methodology and qualitative analyses

As stated previously (see Section 3.1), the mixed methodology³³ applied in the articles was based on the idea that purely quantitative analyses of utterance fluency are insufficient for capturing all facets of fluency: In particular, multifunctional fluency-related features should also be examined from a qualitative perspective. A qualitative analysis is also needed for understanding variation in learners' fluency and individual differences across learners. Thus, in Articles I–III, the mixed methodology involved complementing quantitative utterance fluency analysis with qualitative utterance fluency analysis. Based on Dörnyei's (2007) typology of the main types of mixed-methods designs in applied linguistics, the approach can be characterized as a "QUAN + qual" design. The capitalized "QUAN" indicates that the main methodology in the articles was based on the quantitative analyses of fluency, providing insights into the general, group-level tendencies, and was complemented by a qualitative approach (represented by the lowercase letters). The plus sign indicates that both methods of analysis were employed in parallel (a concurrent design), as opposed to sequential designs (including, e.g., separate data collection for quantitative and qualitative data; Dörnyei, 2007). Furthermore, the quantitative and qualitative analyses were applied to the same set of data (i.e., monologue and dialogue L2 speech samples).

In Article IV, a mixed-methods approach was also used, but the qualitative analysis differed from the other articles in its focus and in terms of the applied qualitative methodology, as the analysis involved grouping the raters' comments thematically based on qualitative content analysis (e.g., Dörnyei, 2007) to illustrate the aspects to which the raters paid attention when assessing the pair dialogue samples. In other words, the qualitative analysis focused on perceived fluency in Article IV. Thus, as opposed to Articles I–III, Article IV included separate quantitative (fluency ratings and fluency measurements based on speech samples) and qualitative (raters' written comments about their numeric ratings) data sets. The design can, nevertheless, also be viewed as a "QUAN + qual" design, with the quantitative, correlational analyses forming the main component of the study and the qualitative content analysis providing complementary insights into the raters' views on L2 fluency. Furthermore, the quantitative (numeric ratings) and qualitative

³³ While mixed-methods research can be criticized for attempting to combine paradigms or worldviews that are incompatible, from the perspective of the present study, I would argue that the two analytical approaches provide two different and complementary perspectives of the same phenomenon—L2 fluency (see also Dörnyei, 2007). By integrating quantitative and qualitative components within a single study, a more comprehensive picture of L2 fluency and the resources for maintaining it can be obtained than with either approach alone.

(comments about the ratings) data were collected concurrently; the mixed methodology was thus applied at both the data collection and data analysis stages in Article IV.

The qualitative research questions in Articles I–III related to examining individual learners’ use of stalling mechanisms and other fluency resources that have been found to demonstrate within-group variation and individual differences. In Articles I–II, subjects from different groups with contrasting profiles were selected for the qualitative analysis, following studies that have focused on comparing the most and least fluent subjects qualitatively (e.g., Brand & Götz, 2011; Eijzenberg, 2000).³⁴ By focusing on subjects representing extremes (“extreme case sampling”; Dörnyei, 2007, p. 128), rather than analyzing average cases, the whole range of potential fluency profiles could be explored. The sampling procedure was thus in line with the broader aims of the qualitative analyses: to explore within-group variation and individual differences. With the focus on maximally different cases, “the limits of the experience” could be identified (Dörnyei, 2007, p. 128), while, at the same time, the central aspects of fluency could be elucidated: Identifying particular phenomena across extreme cases suggests that they are likely to occur in less extreme forms in average cases as well.

In **Article I**, three subjects with zero or few repair phenomena were contrasted with three subjects with the most repair phenomena ($n = 6$). In **Article II**, 10 subjects were included in the analysis: Six subjects (selected from both groups) with the most stalling mechanisms and CSs (“high fluency resource users”) were chosen for the qualitative analysis and contrasted with those interlocutors ($n = 4$) who used the least fluency resources in their respective groups (“low fluency resource users”). As opposed to Articles I–II, in **Article III**, contrasting profiles were not targeted, and the sampling procedure represented “critical case sampling” (Dörnyei, 2007, p. 128) instead: Six participants with the most stalling mechanisms in their L2 from the two school-level groups were chosen for the analysis. The purpose of this approach was to gain insights into the potential idiosyncratic use of stalling mechanisms in their L2 and determine whether the use could be connected to corresponding L1 behavior. The rationale for the sampling in Article III was therefore similar to the approach to sampling adopted in Articles I–II: By gaining insights into “extreme” profiles, the

³⁴ This approach is similar to the “extreme case analysis” discussed by Dörnyei (2007, p. 272) as one potential approach to data analysis in mixed-methods designs. In this type of analysis, “unusual cases” are selected based on one method (in the present study, based on the quantities of certain fluency-related features) and analyzed based on another (qualitative analysis).

limits of stalling mechanism use could be identified and related to the average tendencies in the sample.³⁵

The qualitative approach used in Articles I–III was informed by other studies that have incorporated a qualitative L2 fluency analysis (Brand & Götz, 2011; Ejzenberg, 2000; Hilton, 2008, 2014; Towell *et al.*, 1996). While the specific foci of the qualitative analyses varied slightly across Articles I–III depending on the research questions, generally speaking, the qualitative analyses were conducted to examine fluency-related features for aspects other than their frequency to illustrate the individual variation in their use. In particular, the features' potential functions, the reasons for their use, and the contexts of use were examined. For instance, in Article II, the features' contexts were taken into account by examining the use of different types of resources together. The use of different types of resources was also compared in Articles II and III. The subjects' overall fluency profiles were taken into account in the analyses in Articles I and II; in Article II, the subjects' vocabulary test scores and stimulated recall protocols were also consulted. In Article III, the focus was on comparing the use of stalling mechanisms across L1 and L2 productions. The use of stalling mechanisms by the individual learners included in the qualitative analysis was also related to group-level patterns.

³⁵ With this type of sampling, particular care should be taken so as not to overgeneralize the findings, which applies to the findings of qualitative analyses in general (see also Dörnyei, 2007, pp. 41, 128). However, the mixed-methods approach adopted in the present study, including the quantitative analyses of average tendencies, helps in mitigating this potential weakness by providing a point of comparison for the qualitative findings.

5 Results

In this chapter, the main results of each original publication will be presented in Sections 5.1–5.4. The three main research aims (see Section 1.2) were to examine

- 1) how a qualitative analysis contributes to illustrating differences in fluency across learners and complements a quantitative approach;
- 2) what a problem-solving framework can reveal about
 - a) learners' speech fluency and resources for maintaining fluency at different school levels,
 - b) the connections between learners' speech fluency and resources for maintaining fluency, and
 - c) the connections between L1 and L2 speech fluency; and
- 3) how fluency can be analyzed based on dialogue data from the assessment and measurement perspectives.

Each article also included more specific foci; these are presented in Table 3 along with connections to the research aims.

Table 3. Focus of the articles and connections to the research aims.

Focus of the articles	Aim 1	Aim 2			Aim 3
		a	b	c	
Article I					
- Differences in fluency and ability to maintain fluency across school levels			x		
- The role of qualitative analysis in illustrating differences in fluency across learners and complementing a quantitative approach	x				
Article II					
- Differences in fluency and ability to maintain fluency across school levels			x		
- Connections between fluency and fluency resources				x	
- The role of qualitative analysis in illustrating differences in fluency across learners and complementing a quantitative approach	x				
- Measurement of fluency in dialogue data					x
Article III					
- Connections across L1 and L2 speech				x	
- Differences in fluency and ability to maintain fluency across school levels			x		
- The role of qualitative analysis in illustrating differences in fluency across learners and complementing a quantitative approach	x				
Article IV					
- Connections across assessment and measurement perspectives on fluency in dialogue data					x
- Connections between fluency and fluency resources				x	
- The role of qualitative analysis in complementing a quantitative approach	x				

The discussion in Sections 5.1–5.4 follows the themes and research aims compiled in Table 3. In Section 5.5, the results of each article will be summarized from the perspective of their specific foci.

5.1 Results of Article I

In Article I, differences in monologic speech fluency across Finnish-speaking and Swedish-speaking learners of English at two school levels (upper secondary school and university) were examined using the CAF corpus data (aim 2a). Methodologically, Article I forms the basis for the other articles by introducing the mixed-methods approach, thus contributing to the first research aim.

The quantitative results of Article I demonstrated differences across the two school-level groups (upper secondary school and university), especially for the Finnish-speaking participants. The measures that demonstrated statistically significant differences across the groups based on a one-way ANOVA (7 of the total 14 measures), along with means, standard deviations, and effect sizes, are compiled in Table 4.

Table 4. Statistically significant differences across the groups based on a one-way ANOVA in Article I (for all comparisons, see Table 2 in Article I, p. 223).

Measures	Upper secondary school		University		Natives
	L1 Finnish Mean (SD)	L1 Swedish Mean (SD)	L1 Finnish Mean (SD)	L1 Swedish Mean (SD)	Mean (SD)
Speech rate** (SR)	110.49 (18.81)	148.06 (27.59)	158.21 (30.26)	177.68 (28.15)	197.76 (36.44)
Articulation rate** (AR)	193.75 (18.33)	196.76 (25.15)	211.59 (24.31)	243.36 (26.89)	257.18 (32.71)
Mean length of run** (MLR)	4.20 (0.84)	7.21 (2.42)	8.73 (3.38)	8.83 (2.75)	11.50 (3.67)
Phonation–time ratio** (PTR)	57.05 (8.82)	74.94 (7.39)	74.64 (10.00)	73.09 (8.04)	76.59 (8.03)
Silent pauses (SPs)/min.**	25.50 (4.41)	20.70 (4.78)	18.29 (4.50)	20.28 (3.47)	16.83 (3.28)
Mid-clause SPs/min.**	11.81 (3.37)	7.33 (4.04)	5.35 (2.60)	7.46 (2.49)	3.91 (1.93)
Mean length of SPs in seconds*	1.03 (0.28)	0.73 (0.10)	0.81 (0.14)	0.79 (0.13)	0.82 (0.25)

Note: * $p < .05$, ** $p < .001$

Effect sizes (partial eta squared): SR = 0.54, AR = 0.52, MLR = 0.55, PTR = 0.44, SPs/min. = 0.36, mid-clause SPs/min. = 0.47, mean length of SPs = 0.25.

The pairwise comparisons following a one-way ANOVA demonstrated that Finnish-speaking university (Fin-Uni) subjects were statistically significantly more fluent than Finnish-speaking upper secondary school students (Fin-USS) based on the following five fluency measures: SR ($p = .005$), MLR ($p < .001$), PTR ($p < .001$), SPs per minute ($p = .003$), and mid-clause SPs per minute ($p = .001$). That is, Fin-Uni subjects spoke faster, produced more syllables between pauses, spent less time pausing overall, and paused less frequently especially in mid-clause positions than Fin-USS subjects.

Compared to the Finnish-speaking subjects, the differences across two school levels were found to be less prominent for the Swedish-speaking subjects. Swedish-speaking upper secondary school (Swe-USS) and university subjects (Swe-Uni) were found to differ statistically significantly only in AR ($p = .002$). In fact, on average, the Swe-USS subjects came close to the Swe-Uni group (and the Fin-Uni group; the university groups did not differ statistically significantly) in their fluency. The Swe-USS students' high level of fluency was also demonstrated in the L1 comparisons by school level. The Swedish-speaking students outperformed their Finnish-speaking peers at upper secondary school based on four temporal fluency measures—SR ($p = .042$), MLR ($p = .007$), PTR ($p < .001$), and mean length of SPs ($p = .006$)—indicating a beneficial L1 Swedish influence on L2 English fluency, especially at the upper secondary school level (L1 effects being minor at university level).

Overall, while statistically significant differences across the five groups were found for seven of the fluency measures with the one-way ANOVA, statistically significant differences were not found for clause boundary pauses, FPs, hesitation groups, or any of the four repair measures. The findings regarding FPs and repetitions are particularly interesting, considering the dissertation as a whole, as well as the methodological implications: For both FPs and repetitions, the within-group variation was substantial. For FPs/min., the *SDs* were between 2.86 and 3.65 for the four learner groups (*Ms* between 5.25 and 5.90); for repetitions/min., the *SDs* ranged from 0.51 to 2.40 for the four learner groups (*M* less than one repetition per minute in the learner groups, with the exception of the Swe-Uni group, for which $M = 2.68$ repetitions per minute).

To elucidate the individual differences in the use of repair features and FPs, a qualitative analysis was conducted. The analysis demonstrated that the subject who used the most repair at the upper secondary school level used repetitions to maintain the flow of speech and to reduce time spent pausing. (His average pause duration was only 0.61 seconds.) Yet, due to his high input generator (Seliger, 1980) type of profile, involving planning while talking with only brief pauses and frequent repair, he paused more often mid-clause for planning than the subjects in his group did on average. In contrast, the university-level subject with the highest repair frequency

mostly paused at clause boundaries; his mid-clause SP frequency was the lowest in the Swedish university group. The subject's FPs and repetitions occurred at clause boundaries (in other words, at native-like planning locations) and therefore did not disturb the flow of speech. These devices were used by the subject to buy time, keep the flow of talk going, and reduce mid-clause pausing.

Overall, the qualitative analysis complemented the quantitative analysis by demonstrating how selected learners used repetitions and FPs in particular in fluency-enhancing ways. Some differences were also found between two participants who represented different proficiency levels: While the upper secondary school participant with the most repair used fluency-enhancing devices to reduce time spent in silence in the middle of clauses and clause boundaries, the university student used similar resources mostly at clause boundaries. The qualitative analysis also highlighted the importance of analyzing different types of repairs separately, as they can have different functions: Notably, repetitions differ from other types of repairs (false starts, replacements, and reformulations) in their stalling function. The analysis further showed that repetitions and FPs are not straightforward disfluency markers: They were used to a different extent by individual learners. Thus, the within-group variation can also at least partly explain the lack of statistically significant group-level differences for these resources in the quantitative analysis. These findings, which underscore individual variation and the fluency-enhancing aspects of the use of FPs and repetitions and draw attention to the importance of analyzing the contexts and functions of these phenomena in addition to their frequency, form the basis for the Fluency Resources Framework in Article II and the mixed-methods approach used in Articles II and III.

5.2 Results of Article II

By comparing ninth graders' (G1) and upper secondary school students' (G2) fluency according to the Fluency Resources Framework, Article II contributes to research aim 2a. Article II also contributes to examining connections across fluency and different aspects of problem-solving (fluency resources) with correlational analyses (aim 2b). Complementing quantitative analysis with qualitative analysis, the article also contributes to the first research aim. Finally, as dialogue data were used in the article, it also contributed to the third research aim.

In Article II, seven measures of individual fluency (SR, AR, MLT, SPs/min., mean length of SPs, mid-clause SPs/min., and clause boundary SPs/min.) and four measures of interactional fluency (TPs/min., mean length of TPs, other-repetitions, and collaborative completions) were examined, along with four measures of stalling mechanisms (FPs, drawls, fillers, and repetitions) and eight measures of CSs (all strategies, paraphrases, circumlocutions, approximations, all-purpose words, other

paraphrase strategies, transfer, and appeals for assistance). All frequency measures were standardized per minute of speaking time. The statistically significant differences between the groups according to Mann-Whitney U-tests, along with means, standard deviations, and effect sizes, are compiled in Table 5.

Table 5. Statistically significant differences between the groups based on Mann-Whitney U-tests in Article II (for all comparisons, see Table 1 in Article II, p. 6).

Measure	Ninth grade (G1)	Upper secondary school (G2)	z	p
	Mean (SD)	Mean (SD)		
Speech rate (SR)	132.38 (29.63)	168.44 (34.38)	3.212	.001
Silent pauses (SPs)/min.	37.07 (11.38)	27.63 (10.48)	2.720	.007
Mean length of SPs (sec.)	0.90 (0.20)	0.67 (0.13)	3.807	< .001
Clause boundary SPs/min.	20.51 (5.70)	15.75 (5.15)	2.331	.020
Turn pauses (TPs)/min.	29.64 (19.89)	13.19 (10.09)	3.574	< .001
Mean length of TPs (sec.)	1.95 (0.36)	0.97 (0.39)	5.081	< .001
Other-repetitions/min.	0.14 (0.38)	1.12 (1.82)	2.799	.005
Fillers/min.	1.74 (5.24)	3.00 (3.82)	2.307	.021
Repetitions/min.	0.77 (1.56)	1.54 (1.21)	2.783	.005

Effect sizes (*r*): SR = 0.50, SPs/min. = 0.42, mean length of SPs = 0.59, TPs/min. = 0.55, mean length of TPs = 0.78, other-repetitions/min. = 0.43, repetitions/min. = 0.43.

As can be seen from Table 5, considering the differences between G1 and G2, the individual temporal fluency measures demonstrated that upper secondary school students were statistically significantly more fluent than ninth graders based on three measures: SR, SPs/minute, and average (within-turn) pause duration. The final measure demonstrated the largest effect size. The differences in clause boundary SPs were not significant after the FDR correction. The tendencies for the other temporal measures were similar, indicating upper secondary school students' higher fluency on average, but the differences did not reach statistical significance.

Regarding interactional fluency (referred to as dialogue fluency in the article), differences between the groups were found in the number of (between-)turn pauses per minute and turn pause duration. Considering both individual temporal fluency measures and interactional fluency measures, based on the effect sizes, the most prominent difference between the groups was found for the average length of TPs: G2 subjects spent less time in silence between turns than G1 students, demonstrating

higher interactional fluency. In addition to the temporal interactional fluency measures, the measures for cohesive devices demonstrated a statistically significant difference regarding other-repetitions per minute: While G1 seldom used other-repetitions, G2 used them once per minute on average. The differences in collaborative completions per minute were not statistically significant between the two groups.

Regarding the four types of stalling mechanisms (FPs, drawls, fillers, and repetitions), a statistically significant difference between the groups was found only for repetitions: G2 used repetitions more than G1 subjects on average. G2 also used more fillers on average than G1, but the difference did not reach statistical significance after the FDR correction. It should also be noted that the within-group variation was high for fillers. For FPs and drawls, the group means did not differ to a significant extent, and within-group variation was substantial for both measures (for FPs, especially in G1; FP G1 $M = 6.87$, $SD = 6.33$, G2 $M = 6.21$, $SD = 3.91$; drawls G1 $M = 4.10$, $SD = 4.30$, G2 $M = 3.55$, $SD = 4.25$).

Finally, for communication strategies, no statistically significant differences were found between the groups in the overall use of strategies or in the use of specific CS types. G2 students used fewer CSs on average per minute ($M = 3.99$) than G1 students ($M = 6.55$), but the difference between the groups was not statistically significant, and within-group variation was substantial, as indicated by the standard deviations (G1 $SD = 4.39$, G2 $SD = 2.95$). Paraphrases were used more often by both groups (G1 $M = 4.74$, G2 $M = 3.58$) than transfer strategies (G1 $M = 1.29$, G2 $M = 0.28$) or appeals of assistance (G1 $M = 0.52$, G2 $M = 0.12$), with the most commonly used paraphrase strategy in both groups being approximations (G1 $M = 2.39$; G2 $M = 2.15$). All-purpose words were somewhat more common among G1 subjects ($M = 1.32$) than among G2 subjects ($M = 0.49$), but the difference did not reach statistical significance.

In addition to differences between the two groups, correlations across temporal fluency/interactional fluency and different types of problem-solving mechanisms (stalling mechanisms and CSs) were examined. The statistically significant correlations are compiled in Table 6.

Table 6. Statistically significant correlations across problem-solving mechanisms and temporal/interactional fluency in Article II.

Problem-solving mechanism	Temporal fluency/interactional fluency	Correlation
Drawls/min.	Articulation rate (AR)	-.336*
Fillers/min.	Speech rate (SR)	.479**
Fillers/min.	Articulation rate (AR)	.356*
Fillers/min.	Mean length of turn (MLT)	.547**
Fillers/min.	Silent pauses (SPs)/min.	-.398**
Fillers/min.	Mean length of SPs	-.451**
Fillers/min.	Clause boundary SPs/min.	-.404**
Fillers/min.	Turn pauses (TPs)/min.	-.645**
Fillers/min.	Mean length of TPs	-.492**
Fillers/min.	Other-repetitions/min.	.314*
Repetitions/min.	Turn pauses (TPs)/min.	-.392*
Repetitions/min.	Mean length of TPs	-.444**
Repetitions/min.	Other-repetitions/min.	.491**
Circumlocutions/min.	Mid-clause SPs/min.	.425**
Circumlocutions/min.	Other-repetitions/min.	-.313*
Approximations/min.	Silent pauses (SPs)/min.	.342*
Approximations/min.	Mean length of SPs	.330*
Approximations/min.	Clause boundary SPs/min.	.322*
All-purpose words/min.	Mid-clause SPs/min.	.376*
Other/min.	Other-repetitions/min.	.306*

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

As can be seen from Table 6, regarding temporal fluency and the four types of stalling mechanisms, most statistically significant correlations were found between fillers and temporal fluency/interactional fluency measures. The strongest correlations were found between the frequency of TPs and fillers, as well as between MLT and fillers. That is, the use of fillers was associated with fewer TPs (interactional fluency) and longer turns on average (individual fluency). Repetitions were also found to be statistically significantly and moderately correlated with three interactional fluency measures: other-repetitions per minute, mean length of TPs, and frequency of TPs. In other words, repeating one's own words was connected to repeating the interlocutor's words. Repetitions were also associated with shorter and fewer TPs, being associated in particular with interactional fluency rather than individual fluency.

For the other type of problem-solving mechanisms—that is, CSs—only seven statistically significant correlations with temporal fluency/interactional fluency measures were found, with small to medium effect sizes ranging from $r_s = .306$ to $r_s = .425$. Notably, paraphrase strategies were correlated with different SP measures: For instance, circumlocutions and all-purpose words were moderately positively correlated with mid-clause SPs. In addition, approximations were moderately positively correlated with three temporal measures related to pausing more generally. That is, the use of approximations was associated with SP frequency, SP duration, and specifically with clause boundary SP frequency. None of the correlations reached statistical significance for the other two main strategy categories (transfer and appeal for assistance) and were weak overall.

While the quantitative analysis pointed to some connections between disfluency (pausing) and CSs, the analysis was complemented with a qualitative analysis focusing on individual learners' profiles (high fluency resource users contrasted with low fluency resource users). The qualitative analysis demonstrated that the four subjects who frequently used CSs often used combinations of strategy types: In particular, all-purpose words were often followed by a circumlocution. Stalling mechanisms also often co-occurred with CSs. These resources, such as repetitions and fillers, compensated for the effects of local disfluencies (especially mid-clause pauses) that resulted from word searches. In addition to the subjects who used many fluency resources (stalling mechanisms and/or CSs), selected subjects' productions with few fluency resources were analyzed. The analysis revealed two different potential reasons for the lack of strategy use: While some subjects seemed not to use strategies due to poor vocabulary knowledge combined with inadequate strategic skills, other subjects demonstrated good vocabulary knowledge and thus a lack of need for strategy use.

Overall, the qualitative analysis complemented the quantitative findings and provided potential explanations for the quantitative findings. The quantitative

results, which demonstrated few group-level differences in CS use and weak correlations with the temporal aspects of fluency (with the exception of some disfluencies—especially mid-clause SPs), could be explained, at least in part, by the qualitative analysis, which showed that no single fluency profile could be associated with strategy use; strategies were employed by subjects with differing levels of temporal fluency. Regarding stalling mechanisms, for which strong within-group variation and few statistically significant correlations were found in the quantitative analysis (with the exception of fillers being connected to higher temporal fluency), the qualitative analysis highlighted their local fluency-enhancing functions in conjunction with word searches. Thus, for stalling mechanisms, a qualitative analysis is useful to complement the picture provided by the quantitative, group-level analysis.

5.3 Results of Article III

In Article III, the problem-solving framework was applied to study the connections between individual speaking style in L1 and L2 speech fluency (aim 2c): Speech samples from ninth-grade and upper secondary school participants in their L1 (Finnish) and L2 (English) were examined. As a mixed-methods approach was used in the article, it also contributed to answering the first research aim. Finally, the article also addressed to a lesser extent the differences between learners at two school levels (aim 2a) and took school level into account in the correlational and regression analyses. The individual temporal fluency measures were the same as in Article II (with the difference being that MLR being used instead of MLT), with two additional measures (totaling nine measures): mean length of mid-clause SPs and mean length of end-clause SPs. Regarding measures of fluency resources, only the four types of stalling mechanisms were included.

Starting with the school-level differences, Mann-Whitney U-tests were used to examine potential differences between the two groups (ninth grade, G1; upper secondary school, G2) in their L1 and L2 monologue fluency in a preliminary analysis. The results demonstrated that while G2 subjects were somewhat more fluent than G1 subjects on average based on most of the nine temporal L2 fluency measures that were examined, only one of the differences reached statistical significance (mean length of end-clause SPs), and the effect sizes ranged from small to medium. The differences were clearer in the use of stalling mechanisms: G2 subjects were found to use drawls, fillers, and repetitions statistically significantly more than G1 subjects in their L2 (medium effect sizes; no statistically significant differences found for FPs). Regarding differences in L1 fluency, the differences were more prominent, as six statistically significant differences were found (with medium to large effect sizes), demonstrating G2's higher level of fluency: SR, AR, MLR,

clause boundary SPs/min., mean length of SPs, and mean length of end-clause SPs. For L1 stalling mechanisms, the differences were not statistically significant.

The correlational analyses (Spearman's rank-order correlation coefficients) used to examine the connections between L1 and L2 speech fluency were conducted for the two groups combined (whole sample) and separately. The first analysis for the whole sample ($N = 42$) revealed statistically significant positive correlations for the majority of the fluency measures (10 of the total 13), with medium to large effect sizes ranging from $r_s = .411$ for FPs to $r_s = .682$ for the mean length of end-clause SPs/minute. In the subsequent analysis, the correlations were calculated separately for the two groups to reveal potential school-level effects. The analysis demonstrated stronger and more statistically significant correlations for G2 ($n = 25$) compared to G1 ($n = 17$). Ten of the 13 measures in L1 and L2 correlated statistically significantly for G2 (all temporal fluency measures and FPs), while for G1, six measures were statistically significantly correlated (SR, MLR, mid-clause SPs, mean length of end-clause SPs, FPs, and drawls). The correlations ranged from moderate to strong for both groups. As a whole, the results of the correlational analyses suggest that L1 and L2 fluency are connected irrespective of the school level for measures that were statistically significantly correlated for both groups—that is, SR, MLR, mid-clause SPs/min., the mean length of end-clause SPs, and FPs/min. For five temporal measures (AR, SPs/min., end-clause SPs/min., mean length of SPs, and mean length of mid-clause SPs), the connection seemed to vary according to school level: The connections were stronger for G2 and statistically significant solely for them—not for G1. For drawls, the opposite pattern was detected: The correlations were stronger and statistically significant for G1 only. For fillers or repetitions, statistically significant whole-sample or group-level correlations across L1 and L2 were not found.

In addition to the correlational analyses, the connections between L1 and L2 fluency were examined using regression analyses. Despite the differences across groups revealed by the correlational analyses, adding school level as a predictor improved the model only for drawls per minute (hierarchical regression); for the other measures, simple regression models with L1 measures as predictors are reported in Table 7.

Table 7. Simple regression models predicting L2 fluency measures based on L1 fluency measures in Article III (Table 5 in Article III, p. 685).

Measure	<i>B</i>	Std. error	Slope (β)	<i>F</i> (1,40)	<i>p</i>	Adj. <i>R</i> ²
<i>Temporal fluency</i>						
Speech rate (SR)	.384	.084	.587	21.04	< .001	.328
Articulation rate (AR)	.001	.000	.591	21.52	< .001	.334
Mean length of run (MLR)	.312	.069	.580	20.29	< .001	.320
<i>Pause frequency</i>						
Silent pauses (SPs)/min.	.823	.152	.650	29.22	< .001	.408
Mid-clause SPs/min.	.794	.149	.646	28.58	< .001	.402
Clause boundary SPs/min.	.532	.131	.540	16.48	< .001	.274
<i>Pause duration</i>						
Mean length of SPs	.199	.046	.561	18.37	< .001	.298
Mean length of mid-clause SPs	.074	.179	.066	0.17	.680	-.021
Mean length of end-clause SPs	.307	.035	.815	79.12	< .001	.656
<i>Stalling mechanisms</i>						
Filled pauses (FPs)/min.	.778	.282	.400	7.62	.009	.139

As can be seen from Table 7, the results of the simple regression models predicting L2 fluency based on L1 fluency measures were statistically significant for 9 of the 10 examined measures. Fillers and repetitions were not included in the analyses, as they did not correlate statistically significantly across L1 and L2 performance. For temporal measures, the highest percentage for explained variance regarding the temporal measures was found for the mean length of end-clause SPs (66%); for the other variables, the explained variance ranged from 27% for end-clause pause frequency to 41% for silent pauses per minute. For FPs, the amount of explained variance was 14%. Only the model for the mean length of mid-clause SPs was nonsignificant. For drawls, the hierarchical regression model showed that after the addition of the school level as a dummy variable, the explained variance increased from 8% to 26% (Adj. $R^2 = .257$; $F(2,39) = 8.07$, $p = .001$), but L1 drawls were not a significant predictor in the model after school level had been added (L1 drawls $B = 0.034$, Std. error = 0.018, $\beta = 0.259$, $p = .063$; Group $B = 5.031$, Std. error = 1.532, $\beta = 0.445$, $p = .002$). Overall, the results of the regression analyses corroborate the

whole-group correlations: Most temporal fluency variables in the L2 can be predicted to a certain extent from L1 measures, but the predictive power varies across the measures. Yet, compared to the correlations, school level emerged as a significant predictor only for the measure drawls and thus had a more minor role than L1 fluency measures in explaining variation in L2 fluency measures.

In the complementary qualitative analysis, connections across L1 and L2 fluency were examined among individual speakers, with a focus on stalling mechanisms. Generally speaking, for the six participants included in the analysis, G1 participants used FPs more in the L2 than in the L1, while the results were more mixed for G2 participants. Based on an example analysis of one G1 participant, the use of FPs was not only more common in the L2, but FPs also differed in their distribution across the L1 and L2: In the L1, FPs occurred at clause boundaries, while the L2 also contained mid-clause FPs. For drawls, the analysis demonstrated that the pattern of using drawls more in the L2 was clearer for the three G2 subjects than for G1 subjects: Thus, overall, G2 subjects used this resource more in their L2. For fillers and repetitions, the patterns were more mixed for both groups (cf. lack of group-level correlations for these measures across the L1 and L2 mentioned above). Overall, the three G1 participants rarely used fillers and repetitions; in G2, two participants used more fillers in English, while repetitions were produced by only one participant in Finnish (all produced them in English). Based on a detailed analysis of one G2 participant's use of different types of stalling mechanisms across her L1 and L2 productions, the preference for the types of stalling mechanisms was shown to vary across languages: While the participant tended to use a particular filler word in her L1, she mostly used drawls to gain more planning time and to fill silences in her L2. Another G2 participant used repetitions solely in the L2 as his main stalling mechanism. This suggests that learners can have different ways of coping with processing time pressure and might rely on some stalling mechanisms more than others. Furthermore, these tendencies do not necessarily transfer across languages; some stalling mechanisms may be L2-specific.

Together, the quantitative and qualitative analyses revealed different aspects of the connections between L1 and L2 fluency. Regarding temporal fluency, the quantitative analyses demonstrated that most measures were correlated across the L1 and L2. Furthermore, most L2 measures could even be predicted from the L1 measures, with the explanatory power ranging from 27% to 66%. For stalling mechanisms, correlations between L1 and L2 measures were found only for FPs and drawls, and the percentages of explained variances were relatively low. The qualitative analysis complemented the quantitative findings by illustrating the substantial individual variation in the use of stalling mechanisms; this variation can also partly explain the lack of group-level correlations for repetitions and fillers. The qualitative analysis, which included examinations of different stalling mechanism

types, also complemented the quantitative findings of differences across groups in L2 productions. The analysis showed that G2 subjects used draws, fillers, and repetitions more extensively than G1 subjects, whereas FPs were used quite often by the subjects in both groups. Overall, based on the qualitative analysis, G2 subjects thus used a broader repertoire of resources to keep the flow of talk going in their L2 than the G1 subjects, with some differences in preferences across languages and subjects.

5.4 Results of Article IV

In Article IV, the focus was on examining the fluency assessments of dialogue data. Fifteen raters assessed six samples of peer interaction on 9-point scales and commented on their scores. By examining the connections between assessed and measured fluency in an interactional context, Article IV thus contributed to the third research aim in particular. Furthermore, as the Fluency Resources Framework was employed in the article, connections across ratings of strategic competence and fluency were also examined (aim 2b). Finally, Article IV also contributed to the first research aim, despite not focusing on a qualitative assessment of utterance fluency but on perceived fluency (judges' written comments about their numeric assessments).

First, the connections between measured and assessed fluency for the six interactional samples were examined with Spearman's rank-order correlation coefficients. The measures were the same as in Article II, with the difference being that one individual temporal fluency measure (clause boundary SPs) was excluded. The results for (individual) temporal fluency and mean ratings of individual fluency demonstrated strong and statistically significant correlations for three temporal variables: the strongest correlation was found for SR ($p < .01$, $r_s = .828$), followed by AR ($p < .05$, $r_s = .681$) and MLT ($p < .05$, $r_s = .646$). The silent pause measures did not correlate statistically significantly with individual fluency ratings and demonstrated small (mid-clause SPs $r_s = -.298$) to medium (mean length of SPs $r_s = -.428$, SPs/min. $r_s = -.495$) negative correlations. For stalling mechanisms, the correlations between the measures and mean ratings of individual fluency were weak (ranging from $r_s = .042$ for FPs to $r_s = -.232$ for draws) and not statistically significant.

Regarding correlations between the four interactional fluency measures and the mean interactional fluency ratings, the negative correlation between the mean length of TPs and interactional fluency ratings was found to be very strong ($r_s = -.943$). It was also the only statistically significant correlation ($p < .01$). For the other temporal interactional fluency measure, TPs per minute, the negative correlation was medium in effect ($r_s = -.495$) but not statistically significant. The positive correlations for the

two interactional fluency measures related to cohesive devices were medium (other-repetitions $r_s = .381$) or weak (collaborative completions $r_s = .237$) in effect and not statistically significant.

Finally, regarding measures for CSs and the mean strategic competence ratings, no statistically significant correlations were found. Most of the correlations were negative and weak (ranging from $r_s = -.033$ for appeals for assistance to $r_s = -.232$ for all-purpose words), suggesting that the more frequent use of strategies was associated with lower mean ratings. Two correlations were positive—for circumlocutions ($r_s = .218$) and other paraphrase strategies ($r_s = .431$)—with small to medium effects.

In addition to the analysis of connections across measured and assessed fluency, Spearman's rank-order correlation coefficients across the ratings of individual fluency, interactional fluency, and strategic competence were examined. The analysis demonstrated that all the ratings were strongly and statistically significantly correlated. For individual and interactional fluency mean ratings, the correlations were the strongest ($r_s = .773, p < .05$). The correlation between individual fluency and strategic competence mean ratings was also strong ($r_s = .712, p < .05$). The correlation between interactional fluency and strategic competence was slightly weaker ($r_s = .622, p < .05$), but it can still be considered strong.

The connections across the ratings and the links between the measurement and assessment perspectives were further examined using a qualitative content analysis of the raters' comments, which included details about the aspects to which the raters' paid attention in the interactions. The themes emerging from the data were categorized according to the assessed dimensions: individual fluency, interactional fluency, and strategic competence, as well as the category "other" for comments that did not fit into the previously mentioned categories.

First, for *individual fluency*, five themes emerged from the data (30.6% of all comments): speech tempo and flow (9.1%), pausing (8.0%), hesitations and corrections (6.4%), amount of speech (3.7%), and stalling mechanisms (3.4%). That is, the raters seemed to pay attention to the features that they were instructed to take into account in their assessments (SR, pausing, and repair). When it comes to rating individual fluency, the raters' perspective thus seemed to match the quantitative, measurement-based perspective relatively well.

For *interactional fluency*, seven themes were identified (30.9% of all comments): acknowledging or encouraging interlocutor's contributions (9.0%), the flow of discussion (6.0%), balance in turn-taking (5.7%), pausing during the discussion (3.1%), mutual understanding (3.1%), task accomplishment (2.3%), and other aspects of collaboration (1.8%). Of these themes, the flow of discussion and pausing during the discussion were most closely linked to the temporal interactional fluency measures (frequency and duration of TPs). Furthermore, the function of

“acknowledging or encouraging interlocutor’s contributions” was reflected in the two measures for cohesive devices examined in the present study—namely, other-repetitions and collaborative completions. However, the raters’ comments overall revealed that they viewed interactional fluency as a broader concept than what was captured by the interactional fluency measures. For instance, they also paid attention to aspects such as balanced turn-taking and indicators of mutual understanding. Furthermore, the raters’ comments revealed some challenges in distinguishing the different dimensions: For instance, comments related to the category “collaborative problem-solving” could be viewed either as part of interactional fluency or strategic competence (see below). Individual and interactional fluency were also viewed to some extent as interrelated in the raters’ comments: Of the comments categorized under interactional fluency, 9.9% were mentioned among the answers for individual fluency. Some of the raters also expressed difficulties in assigning strategy scores to those participants who were perceived as possessing the needed vocabulary for the task and consequently did not need to use strategies.

Three main themes were identified for *strategic competence* (23.7% of all comments): verbal communication strategies (14.8%), collaborative problem-solving (4.6%), and nonverbal communication strategies (4.3%). As can be seen from the percentages, different types of compensations for gaps in vocabulary with CSs were most often mentioned by the raters, reflecting the instructions for the ratings, which guided the raters to focus on CSs. Yet, a relatively large proportion (4.6%) of the comments related to other collaborative means for problem-solving, such as the negotiation of meaning. That is, for strategic competence, the measurement perspective, with a focus on individual CSs and their frequencies, did not fully match the raters’ assessment perspective. Consequently, this can, at least to some extent, explain the lack of correlations between the mean strategic competence ratings and strategy measures. Finally, *other comments* (14.8% of all comments) related to the following categories: vocabulary (4.6%), content (4.1%), other (1.4%), proficiency level (1.3%), accuracy (1.2%), pronunciation (0.8%), the use of the L1 (0.7%), and personality (0.7%).

5.5 Summary of the results

Article I addressed the second research aim (2a) by examining differences in L2 monologue fluency across learners from two school levels (upper secondary school and university) and two L1 backgrounds (Finnish and Swedish). The results demonstrated that university subjects were more fluent than upper secondary school subjects on average. However, the differences in fluency were clearer among the Finnish-speaking learners at different school levels than among the Swedish-speaking learners; the Swedish-speaking upper secondary school students were close

to university students in their L2 English fluency. The results of Article I also demonstrated that temporal fluency measures were better suited to group-level examinations than, for instance, repetitions and FPs, which were associated with substantial within-group variation. Article I thus also contributed to the first research aim by demonstrating that repetitions and FPs could be used in a fluency-enhancing manner, questioning the status of these features as straightforward disfluency markers.

Article II also addressed the second research aim (2a): In addition to examining differences in fluency and the ability to maintain fluency across school levels (ninth grade and upper secondary school) based on dialogue data using a problem-solving framework, connections across different aspects of problem-solving were examined. The results demonstrated that upper secondary school students were, in general, more fluent than ninth graders. However, in contrast to the monologue data examined in Article I, the most prominent differences between the groups in Article II were not found in temporal individual fluency measures but, rather, in a measure that can be considered an indicator of interactional fluency: the mean length of (between-)turn pauses (cf. the third research aim). For stalling mechanisms, the results were comparable to the Article I results: these fluency resources were associated with substantial within-group variation. For the other type of problem-solving mechanisms (CSs), differences in their frequencies were not found between the two groups.

The correlational analyses in Article II (aim 2b) showed that fillers, in particular, were strongly correlated with temporal (individual) fluency and interactional fluency measures; this finding underscores the role of fillers as resources for maintaining fluent speech in dialogue. Conversely, CSs were found to be correlated with pause measures, pointing to an association with disfluency. In particular, circumlocutions and all-purpose words were moderately positively correlated with mid-clause SPs. Yet, the qualitative analysis, contributing to the first research aim, showed that stalling mechanisms, such as fillers, were often used in conjunction with CSs and helped to mitigate the negative effects of disfluencies (especially mid-clause SPs) caused by word searches. In addition to stalling mechanisms often being used in conjunction with CSs, combinations of CSs were also commonly used to ensure successful communication. The qualitative analysis also provided insights into the lack of differences between school-level groups in the quantitative analysis. Two polarized reasons for the lack of strategy use were identified: poor (lack of competence to use strategies) vs. very good vocabulary skills (no need to use strategies). Overall, Article II also contributed to the third research aim by demonstrating the importance of complementing individual fluency measures with interaction-specific measures in a dialogue context.

In **Article III**, monologue data in the L1 (Finnish) and L2 (English) from the same learners were used to examine how L1 and L2 fluency were connected from the perspective of the problem-solving framework (excluding CSs; the second research aim). It also contributed to aim 2c by examining school-level differences in a preliminary analysis and taking the school level into account in subsequent correlational and regression analyses. The Article III results demonstrated clearer differences between the two school-level groups in their L1 Finnish monologue productions than in their L2 English monologue productions, especially regarding temporal fluency. Interestingly, however, upper secondary school subjects were found to use more stalling mechanisms than ninth graders in their L2. These school-level differences were also, to some extent, reflected in the groupwise correlational analyses: stronger and more statistically significant correlations were found for the upper secondary school group compared to the ninth-grade group. However, including school level as a predictor rarely improved the regression models (the only exception being draws).

Simple regression models predicting L2 fluency measures from L1 measures were statistically significant for 9 of the 10 examined measures. The explained variance differed to some extent across individual measures; the strongest predictive power was found for end-clause SP duration (66%). Thus, overall, most L2 fluency measures could be predicted from L1 measures, at least to some extent. The qualitative analysis (contributing to the first research aim) provided further insights into variation in the use of stalling mechanisms and especially into the idiosyncratic connections between L1 and L2 fluency. Within-individual comparisons across L1 and L2 productions demonstrated that some of the preferences related to stalling mechanisms were language specific: Certain stalling mechanisms were used in the L1, but a similar fluency-enhancing function was filled with another resource in L2.

The main focus in **Article IV** was on the connections between assessment and measurement perspectives on fluency in dialogue data (the third research aim). The results demonstrated differing connections across the different dimensions of fluency: For temporal (individual) fluency, SR, AR, and MLT were strongly correlated with individual fluency ratings. No statistically significant correlations were found between stalling mechanisms and individual fluency ratings. For interactional fluency, the mean length of TPs was found to be strongly correlated with the interactional fluency ratings. Finally, regarding CSs, no statistically significant correlations were found between the frequency-based measures and the judges' strategic competence ratings. Overall, the correlations were weak. In addition to the correlations between the fluency measures and assessments, the results demonstrated strong correlations between the ratings of different aspects of fluency and problem-solving (aim 2b).

The qualitative analysis, contributing to the first research aim, provided further insights into the raters' interpretations of the rated dimensions. The raters' interpretations were shown to match the measurement perspective for individual fluency more closely than for interactional fluency or strategic competence. While individual fluency ratings were mostly based on temporal aspects, in line with the instructions provided for the raters, the interactional fluency comments reflected a rather broad understanding of interactional fluency (with 9.0% of all comments belonging to the "acknowledging or encouraging interlocutor's contributions" category). Yet, the raters also commented on aspects that corresponded with the measurement perspective (e.g., the "pausing during the discussion" category). Strategic competence comments, in particular, included references to other, collaborative problem-solving mechanisms in addition to individual CSs. The analysis also provided insights into the fluid boundaries between the different categories: Individual and interactional fluency were viewed to some extent as interrelated, and collaborative problem-solving was linked to both strategic competence and interactional fluency.

6 Discussion

In this chapter, the results of the present study will be discussed and related to previous studies (Section 6.1), followed by a discussion of the limitations and generalizability of the present study along with suggestions for further study (Section 6.2). Implications for L2 assessment and teaching are discussed in Section 6.3. The final section, 6.4, concludes this dissertation.

6.1 Theoretical and methodological considerations for L2 fluency research

In this section, the results will be related to the three research aims and the general aim of the present study (to examine how learners maintain speech fluency in L2 English monologue and dialogue, using a mixed-methods approach). The results concerning each research aim are discussed in Sections 6.1.1–6.1.3, starting with the qualitative analysis in Section 6.1.1, followed by the problem-solving perspective in Section 6.1.2 and the examination of fluency in dialogue data in Section 6.1.3. These three perspectives and the main research aims associated with them were based on three theoretical and methodological gaps in previous research (see Section 1.2); by exploring these research aims, the present study thus seeks to fill these gaps. In the final section, 6.1.4, the discussions in the previous subsections are drawn together in the form of a theoretical–methodological classification for analyzing L2 speech fluency in monologue and dialogue from a problem-solving perspective.

Figure 5 on the following page forms the basis for the discussion in Sections 6.1.1–6.1.3 by providing an overview of the key findings of the present study. It presents the general aim and condensed main results related to the three specific research aims.

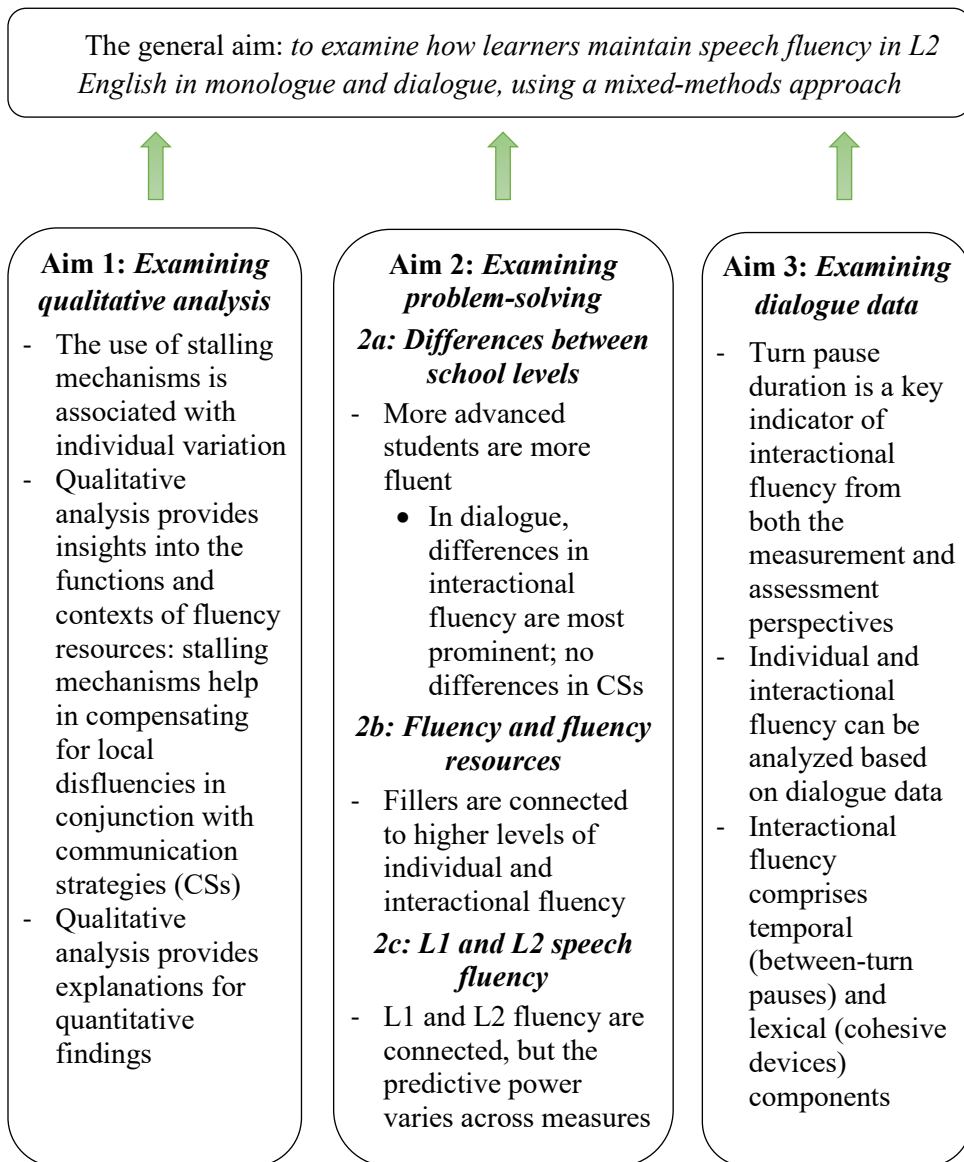


Figure 5. Main results based on the three research aims and their contributions to the general aim.

6.1.1 Qualitative analysis

The first research aim (to examine how a qualitative analysis contributes to illustrating differences in fluency across learners and complements a quantitative approach) was approached in Articles I–III from the perspective of a qualitative

utterance fluency analysis that complemented the quantitative analysis; Article IV also included a qualitative component but addressed perceived fluency (raters' comments). A qualitative component was included in each substudy to complement previous L2 speech fluency research, which has mostly been quantitative in nature and has focused less on the functions and contexts of fluency-related features than on their quantities (see Sections 1.2, 2.1, and 3.1). As illustrated in Figure 5, three main contributions of the qualitative analysis could be identified based on the analyses in the different substudies. First, the qualitative analyses illustrated the variation across learners in the use of stalling mechanisms both in individual monologic production (Articles I and III) and dialogic production (Article II). Second, the analyses provided insights into the functions and contexts of fluency resources: Notably, stalling mechanisms were shown to have important functions, especially in compensating for local disfluencies in conjunction with CSs (Article II; note that this finding is closely linked to aim 2b). Third, the qualitative analyses provided explanations for the findings of the quantitative analyses—for the lack of group-level differences or statistically significant correlations, in particular. Triangulating quantitative and qualitative analyses thus contributed to providing a comprehensive view of fluency and resources for maintaining it. Importantly, the qualitative analysis complemented the frequency-based quantitative analysis by focusing on aspects that could not have been analyzed using a purely quantitative approach.

In Articles I–III, notable individual variation was demonstrated regarding the use of stalling mechanisms in both monologue and dialogue contexts across learner groups representing different school levels. In particular, FPs and repetitions demonstrated substantial within-group variation, which can explain, at least in part, the lack of group-level differences in the quantitative analyses. In Articles II and III, individual variation was also found in the use of fillers and drawls. These findings are in line with previous studies demonstrating individual variation in the use of FPs and repair phenomena, including repetitions (e.g., Cucchiari *et al.*, 2002; Götz, 2013, 2019a, 2019b; Kahng, 2014). While drawls and fillers have not typically been included among indicators of L2 fluency in recent studies (for an early study including both, see Riggensbach, 1991), some studies have shown that fillers are also prone to individual variation (Fehringer & Fry, 2007; Götz, 2013; Wolk *et al.*, 2020). In a recent learner corpus study, Dumont (2018) also found variation related to the use of fillers (discourse markers) among French L2 English speakers and native English speakers; for drawls, within-group variation was substantial, especially in the learner group. Compared to previous mainstream L2 fluency research, which has rarely highlighted individual variation among L2 speakers, the present study focused specifically on these within-group differences to illustrate individual variation in fluency.

In the present study, the individual differences associated with the use of stalling mechanisms, as well as their role in maintaining fluency, were explored in the qualitative analyses. Notably, in Article II, the lack of differences across groups in the use of CSs, as well as the correlation between mid-clause SPs and certain CS types, was further clarified with the qualitative analysis: While some learners did not use strategies due to their high levels of vocabulary knowledge and competence, others did not use them due to their poor vocabulary knowledge and lack of strategic skills (see also Hilton, 2008). These findings echo Dobao's (2001) results regarding the U-shaped behavior associated with strategy use. Paralleling these utterance fluency findings from a perceived fluency perspective, some of the raters in Article IV also commented on the challenges of scoring strategic competence for those participants who were viewed as having sufficient vocabulary knowledge for the task and not having to rely on strategies. The analysis in Article II also revealed that stalling mechanisms had an important role in compensating for local disfluencies, such as mid-clause SPs, occurring during CSs. Thus, the qualitative analysis complemented the quantitative analysis by showing that stalling mechanisms contribute to maintaining fluency during word searches and that other aspects of strategy use are related to fluency in a complex way. That is, despite the link between disfluency (mid-clause pausing) and CSs demonstrated in the correlational analyses, the resources used by the individual learners, including the use of stalling mechanisms and multiple strategies together, helped to overcome the local disfluencies resulting from word searches and contributed to overall fluency.

These observations, based on triangulating quantitative and qualitative analyses, point to the challenges in analyzing CSs in an interactional context both from the assessment and measurement perspectives: Quantifying CSs does not seem to provide a complete picture of the connections between fluency and strategic language use. Thus, in addition to the frequencies of CSs, other aspects should be analyzed, including the collaborative use of strategies, the contexts of strategy use, and the use of multiple strategies together (see also Bialystok, 1990; Chen, 1990). Methodologically, the qualitative analysis thus highlights the usefulness of a mixed-methods approach to CS analysis, including an analysis of relatively infrequently occurring and idiosyncratic fluency-related features in their contexts to complement a frequency-based analysis of average tendencies. The frequencies may not always reveal differences in fluency across learner groups nor demonstrate statistically significant correlations, but how these features are used provides information about the learners' skills in maintaining fluency.

In addition to variation in the use of problem-solving mechanisms in L2 speech, the qualitative analysis also provided important insights into personal preferences regarding the use of stalling mechanisms across L1 and L2 monologue production. Interestingly, some of the learners seemed to display preferences for particular

stalling mechanisms, but these preferences did not necessarily transfer from the L1 to the L2. While certain stalling mechanisms were preferred in the L1, other stalling mechanisms could be used for a similar, fluency-maintaining function in the L2. Previous studies have also suggested that speakers (e.g., Fehrer & Fry, 2007; Hasselgren, 2002; Maclay & Osgood, 1959; Riggensbach, 1991; Wolk *et al.*, 2020) may prefer particular stalling mechanisms and can be grouped according to different profiles (Dumont, 2018; Götz, 2013; Seliger, 1980). Yet, the findings of the present study provide a new angle to these findings and complement them by suggesting that how the stalling function is realized (i.e., the preferred mechanism) may vary not only across learners but also within particular learners across the languages in their repertoire. Therefore, it is important to examine the different types of stalling mechanisms holistically and to compare learners' production in their L1 and L2. However, it is also important to note that because these observations were based on a limited number of learners and on only one L1–L2 pair (Finnish–English), these profiles related to stalling mechanism use should be studied further with other languages, with other learner populations, and preferably also with larger samples.

6.1.2 A problem-solving perspective on L2 speech fluency

Based on the observations in Article I, a problem-solving framework incorporating the analysis of stalling mechanisms and CSs as a means to maintain fluency (i.e., as fluency resources) along with temporal fluency measures was applied in Articles II–IV. The Fluency Resources Framework was applied to examine the second research aim from three perspectives: differences in fluency and ability to maintain fluency across learners at different school levels (aim 2a), connections between learners' speech fluency and resources for maintaining fluency (aim 2b), and connections across L1 and L2 speech fluency (aim 2c). Based on Figure 5, the main findings related to the second research aim can be summarized as follows: More advanced students are more fluent (in dialogue contexts, differences between groups are most prominent in interactional fluency, while there are no differences in CS frequencies); the use of fillers is linked to higher levels of individual and interactional fluency; and L1 and L2 fluency are connected, but the predictive power of L1 fluency measures varies.

Starting with the differences across learner groups, as demonstrated in Articles I and II in particular from the perspectives of monologue and dialogue data, more advanced students also seem to be more fluent in their L2 in general. In particular, the clearest differences were found between upper secondary school and university students (among Finnish-speaking subjects) regarding commonly used, well-established temporal fluency measures related to both overall speech rate and pause location (SR, MLR, PTR, SPs/min., and mid-clause SPs/min.). In a dialogue context,

two of these measures (SR and SPs/min.) also showed differences between ninth graders and upper secondary school students, along with the mean length of SPs. These results are in line with those of previous studies that have shown fluency development regarding temporal measures in formal education (Peltonen & Lintunen, 2019; Tavakoli *et al.*, 2016, Tonkyn, 2012) and study abroad (e.g., Huensch & Tracy-Ventura, 2017b; Mora & Valls-Ferrer, 2012), as well as demonstrated differences in learners' fluency at different proficiency levels (e.g., Ejzenberg, 2000; Hilton, 2008, 2014; Kormos & Dénes, 2004; Riggenbach, 1991; Tavakoli *et al.*, 2020).

Complementing previous findings, the present study provided novel insights into differences in fluency in a dialogue context and in the use of problem-solving mechanisms. Notably, the clearest differences between the groups were found in the temporal fluency measures in Article I, which used monologue data, while the clearest differences between the groups were not found for individual, temporal fluency measures but for interactional fluency (mean length of TPs) in Article II, which used dialogue data. From a processing perspective, less advanced students potentially spent more attentional resources on individual within-turn planning, while more advanced students had more resources available for monitoring the overall progression of the interaction across turns (i.e., higher, discourse-level aspects; see Kormos, 2006; Levelt, 1989, pp. 110–123). These findings highlight the importance of incorporating measures for interactional fluency along with individual fluency measures when fluency is examined in dialogue contexts (see further Section 6.1.3). In the future, when further studies on L2 interactional fluency and the underlying processing are conducted, the findings could also be used to incorporate a dialogic perspective into the (largely monologue-based) L2 speech production models.

Regarding differences related to problem-solving mechanisms, despite some previous studies finding differences across proficiency-level groups regarding CSs (Chen, 1990; Dobao, 2001; Poulisse, 1990), CS frequencies did not demonstrate clear group-level patterns in the present study; the differences in their use were more idiosyncratic (cf. Section 6.1.1). Similarly, in both monologue and dialogue performance, the use of stalling mechanisms to maintain fluency was associated with individual differences (within-group variation), with few prominent group-level differences, as noted in Section 6.1.1. Yet, in Article III, in addition to within-group variation, between-group variation was found for drawls, fillers, and repetitions: Upper secondary school students used more of these resources in their L2 monologue production than ninth graders did. In the dialogue data, of the four types of stalling mechanisms, only repetitions were found to be used more often by upper secondary school students than by ninth graders. These findings suggest that while it is possible to detect some broader, group-level patterns related to the use of stalling

mechanisms, individual differences related to their use are particularly prominent (especially regarding FPs; see also, e.g., Cucchiarini *et al.*, 2002; Götz, 2013, 2019a, 2019b; Kahng, 2014).

While the frequencies of stalling mechanisms or CSs rarely distinguished learners at different school levels in the present study, some correlations between temporal fluency and problem-solving mechanisms were found. Of the stalling mechanisms examined in Article II, fillers in particular were connected to temporal fluency measures from both an individual and interactional perspective: In particular, they helped in maintaining within-turn flow (strong positive correlation between MLT and fillers) and reducing TPs (strong negative correlation between the frequency of TPs and fillers).

The role of fillers in enhancing fluency has not been highlighted extensively in previous L2 fluency studies (but see, e.g., Hasselgren, 2002); perhaps this is due to the lack of studies on interactional fluency and the relatively narrow operationalization of fluency focused on temporal measures. However, Götz (2013) and Wolk *et al.* (2020) emphasized the role of different types of fillers—notably discourse markers and smallwords—as fluency indicators (or “fluencemes”; Götz, 2013), along with, for instance, FPs, from an individual fluency perspective. Dumont (2018) also underscored the importance of examining discourse markers (“lexicalized” pauses) along with FPs (“non-lexicalized” pauses) as indicators of fluency. Thus, the results of the present study complement these previous examinations of fillers as fluency indicators from a peer interactional viewpoint and further underscore the central role of fillers in maintaining fluency. Based on these results, fillers could be included as potential indicators of (individual) fluency, especially in examinations of dialogue data.

Regarding strategies, the quantitative analysis suggested some links between CSs and disfluency: More specifically, mid-clause SP frequencies were positively correlated with circumlocutions and all-purpose words. From a processing perspective, this association can be viewed as reflecting problems associated with the formulation stage during the speech production process (Kormos, 2006; Levelt, 1989), because mid-clause SPs have been regarded as indicating trouble in relation to lexical retrieval and encoding in particular (Lambert, Kormos, & Minn, 2017). Thus, difficulties in lexical retrieval and the associated need for alternative formulations, resulting in an increased processing load, are likely to be reflected in mid-clause pausing (see also Dörnyei & Kormos, 1998; Hilton, 2008). Nevertheless, the qualitative analysis in Article II showed that the disfluency effects could be mitigated, at least to some extent, by the use of stalling mechanisms to reduce mid-clause pausing and other hesitations (as discussed in Section 6.1.1).

Article IV provided further insights into the connections between different aspects of problem-solving: In a correlational analysis, all rated dimensions were

found to be strongly connected (see also Van Batenburg, Oostdam, Van Gelderen, & De Jong, 2018). The raters' comments also revealed connections across dimensions, especially across individual and interactional fluency. Collaborative problem-solving was associated with both strategic competence and interactional fluency in the raters' comments, suggesting links between the strategic and interactional aspects of L2 dialogue. In future studies, to complement the findings of the present exploratory study, these connections across rated dimensions could be studied further, for instance, by designing novel rating scales (see also Sato, 2014) or by applying existing ones (e.g., in the CEFR, 2001) specifically targeting collaborative and strategic aspects of interaction.

The final aspect of the second research aim related to the connections between L1 and L2 fluency. The results demonstrated that there were clear connections between L1 and L2 fluency; many L2 temporal fluency measures could be predicted from their equivalent L1 measures. These results corroborate previous findings of L1 fluency measures being connected to L2 fluency (notably De Jong *et al.*, 2015; Huensch & Tracy-Ventura, 2017a). More specifically, the strongest predictive power was associated with end-clause SP duration (66%), corroborating De Jong and colleagues' (2015) previous findings. As end-clause SPs are associated with content planning (e.g., De Jong, 2016a) and the conceptualization stage of the speech production process (Kormos, 2006; Levelt, 1989), the connection could have reflected overall preparedness for the task. This measure also showed a difference between the groups: The ninth graders spent more time at clause boundaries in both languages, indicating that they spent more time planning during the task than the upper secondary school group. As Lambert *et al.* (2017) noted, more advanced learners may be able to allocate more attentional resources to content planning (conceptualization) due to the overall higher automaticity of the speech production processes, while lower-level learners might need more attentional resources at other stages of speech production, including encoding processes, due to a lack of automaticity (see also Kormos, 2006).

Overall, it is also important to note that participants may differ in their L1 fluency across school levels, as demonstrated in Article III. Compared to most other studies examining connections between L1 and L2 fluency among adult learners (De Jong *et al.*, 2015; Huensch & Tracy-Ventura, 2017a), the participants in the present study included 15–16-year-olds (ninth graders) and 17–18-year-olds (upper secondary school students); it is thus possible that differences in overall maturity across the groups explain some of the findings related to L1 differences.

6.1.3 Interactional fluency

The third research aim (*to examine how fluency can be analyzed based on dialogue data from the assessment and measurement perspectives*) was explored in Articles II and IV, which used dialogue data. As indicated in Figure 5, the main results related to this research aim include the following: turn pause duration can be considered a key indicator of interactional fluency from both the measurement and assessment perspectives, both individual and interactional fluency can be analyzed based on dialogue data, and interactional fluency can be viewed as consisting of two main components—temporal (between-turn pauses) and lexical (cohesive devices).

Comparing the results of Articles II and IV, the mean length of TPs seems to be a particularly promising measure of interactional fluency from both the measurement (utterance fluency, Article II) and assessment (perceived fluency, Article IV) perspectives. In Article II, when differences across ninth graders and upper secondary school subjects were compared with the problem-solving framework, medium effect sizes were found for some of the individual fluency measures, indicating upper secondary school subjects' higher fluency; however, the largest effects were found for TP duration (see also Section 6.1.2). Thus, in addition to corroborating previous results of (individual) temporal fluency measures distinguishing learners at different proficiency levels, the results of Article II highlighted the importance of interactional fluency measures (especially mean length of TPs) in demonstrating the differences between learner groups in an interactional context. The present study thus complemented previous studies comparing monologue and dialogue L2 fluency (Tavakoli, 2016; Witton-Davies, 2014) or approaching dialogue fluency from a more theoretical, exploratory perspective (McCarthy, 2010) by being among the first to operationalize interaction-specific measures of fluency in the context of L2 peer interaction. The results of the present study demonstrated the usefulness of these measures in distinguishing learner groups from both the assessment and measurement perspectives, especially regarding between-turn pause duration.

The present study also highlighted that both individual and interactional aspects can be measured and assessed based on interactional data. In other words, while it is important to complement the individual-focused approach with an analysis of the collaborative aspects and joint performance,³⁶ it is also useful to approach dialogic data from an individual learner's perspective by using measures that capture within-turn fluency. In Article IV, the strongest correlations between fluency ratings and measures were observed for individual fluency (measures SR, AR, and MLT) and

³⁶ Other solutions may need to be adopted in studies comparing fluency in monologues and dialogues (e.g., Tavakoli, 2016, 2018; Witton-Davies, 2014).

interactional fluency (the mean length of TPs). These findings are in line with Van Os *et al.*'s (2020) results from a more controlled experimental setting, in which raters combined cues from speech rate and turn-taking behavior (gaps and overlaps) to form their fluency rating. The findings further suggest that temporal aspects of speech related to individual fluency, which have previously been shown to be connected to raters' assessments in monologue contexts (e.g., Bosker *et al.*, 2013; Cucchiari *et al.*, 2002; Kormos & Dénes, 2004; Magne *et al.*, 2019; Préfontaine *et al.*, 2016; Rossiter, 2009), are also linked to individual fluency ratings in the dialogue context. These measures can thus also be applied to interactional data to capture the fluency of individual learners' contributions along with interaction-oriented measures.

While (between-)turn pauses were strongly correlated with the assessments of interactional fluency, regarding individual fluency, speed fluency measures (AR) or measures that combine aspects of speed and pausing (SR and MLT), but not within-turn (individual) pauses themselves, were strongly correlated with raters' mean scores (cf. Bosker *et al.*, 2013; Rossiter, 2009). Previous studies based on monologue data assessments have shown raters to be sensitive to pause location in controlled experimental contexts (Boomer & Dittmann, 1962; Kahng, 2018), but it is possible that in a dialogue context, "interactional" pauses, or pauses between turns, are more salient to raters than within-turn pauses (see also Sato, 2014). This could be further explored in future studies focusing specifically on the role of pausing in L2 interaction from the assessment and measurement perspectives, incorporating analyses of both within-turn and between-turn pausing.

The qualitative analysis of the raters' comments also provided further insights into how the raters approached individual and interactional fluency when assessing the samples. Individual fluency assessment, which operationalized fluency in the narrow sense (Lennon, 1990), seemed to be relatively straightforward: The raters based their ratings largely on the temporal aspects of fluency that they were also instructed to assess. This finding is comparable to qualitative analyses of comments based on monologue data assessment (Préfontaine & Kormos, 2016). Interactional fluency seemed to be somewhat more difficult to assess, potentially due to the lack of separate instructions for assessing this particular dimension. The raters' comments revealed that they focused on a relatively broad range of interactional phenomena, perhaps equating the construct to some extent with interactional competence (see, e.g., Borger, 2019). However, it should be noted that some aspects that are usually regarded as dimensions of IC, such as topic management (Galaczi, 2014; Galaczi & Taylor, 2018), were rarely referred to in the comments. Nevertheless, in future studies, to ensure that interactional fluency is assessed from a narrower perspective than IC (e.g., Salaberry & Kunitz, 2019) and as one dimension of interaction, specific instructions guiding the judges toward a relatively narrow sense of interactional

fluency could be devised (cf. broad and narrow senses of fluency; Lennon, 1990). For instance, concrete aspects of interactional fluency, such as temporal features (TPs) and cohesive devices (e.g., other-repetitions; cf. checklists in May *et al.*, 2020), could be incorporated into the instructions. This could also help with distinguishing the different dimensions, including interactional and individual fluency, as the judges' comments revealed that the dimensions were, to some extent, perceived as interrelated (see also Van Batenburg *et al.*, 2018). These findings also corroborated the correlational analyses.

Finally, the exploratory operationalizations of interactional fluency in Articles II and IV suggest that it is possible to distinguish at least two main components of interactional fluency that reflect different means of establishing flow across turns: a temporal (between-turn pauses) and a lexical component (cohesive devices: other-repetitions and collaborative completions; see also Peltonen, 2017). In particular, the differences in the use of other-repetitions as a means to acknowledge the other speaker's contribution demonstrated a medium effect size in distinguishing ninth graders and upper secondary school students, thus being a potential candidate of interactional fluency to be further explored in future studies. Potential indicators of interactional fluency will be discussed in more detail in the following section.

6.1.4 A classification of L2 speech fluency components from a problem-solving perspective

Based on the discussion of the main results, the present study can be viewed as having contributed to existing L2 fluency research in three main ways: by providing a more comprehensive picture of fluency by complementing a quantitative analysis with a qualitative analysis (i.e., through triangulating methods of data analysis), by applying a problem-solving perspective to fluency analysis, and by analyzing L2 dialogue data (cf. Figure 1). In the previous sections, these contributions were discussed somewhat separately; therefore, the purpose of the present section is to draw these perspectives together in the form of a classification of L2 speech fluency components from a problem-solving perspective. This classification combines the main theoretical and methodological contributions of the present study into a single framework.

The classification is based on the Fluency Resources Framework applied in Articles II–IV. The classification mainly concerns utterance fluency measures in Segalowitz's (2010) terminology: Under each main component, examples of central measures for capturing the category in question are listed.³⁷ The classification draws

³⁷ From the perspective of perceived fluency (raters' assessments of L2 fluency), the distinction between the main components may be less clear-cut (cf. e.g., the overlap

on previous accounts of fluency—notably, De Jong’s (2016b) and Kormos’s (2006) compilations of central fluency measures and previous empirical studies on (monologue) fluency (see especially Section 2.2). The idea for grouping speed and breakdown measures of fluency into a single temporal component is based on Lennon’s (1990) early twofold distinction between temporal and vocal (including FPs, repetitions, and self-corrections) fluency. The problem-solving perspective, including measures for stalling mechanisms and CSs, is based on Dörnyei and Kormos (1998) and is combined with fluency analysis in the present classification.

The classification differs from previous frameworks of fluency in three ways. First, it adopts a somewhat broader view of fluency; previous studies, especially in recent years, have mostly applied Skehan’s (2003, 2009, 2014; Tavakoli & Skehan, 2005) threefold framework. While Skehan’s framework consists of temporal fluency dimensions (speed and breakdown) and a repair dimension, the current classification also incorporates aspects of problem-solving to provide insights into learners’ resources for maintaining fluency. Second, related to the previous point, a notable difference lies in the treatment of so-called disfluency markers or different types of hesitations: These are considered potentially fluency-enhancing devices, or stalling mechanisms (following Dörnyei & Kormos, 1998), and are primarily viewed in terms of their role in maintaining the flow of speech and reducing time spent in silence. For these devices, it is recommended that a quantitative analysis is complemented with a qualitative analysis to examine their functions and contexts of use in detail. In terms of individual measures for stalling mechanisms, FPs are not considered as belonging under the same breakdown dimension as SPs (for further discussion of the status of FPs, see, e.g., Clark & Fox Tree, 2002 and Section 2.2), and repetitions are not considered along with other repair mechanisms.³⁸ Two additional measures—drawls and fillers—which have been less commonly used in recent fluency research (but see Möhle, 1984; Raupach, 1980 for early studies on

between interactional fluency and strategic competence detected in Article IV). Further research is needed for adapting the classification to incorporate a perceived fluency perspective.

³⁸ Only repetitions were included in the classification; false starts, reformulations, and replacements (Foster & Skehan, 1999), which are commonly examined under the repair dimension, were not (see Dörnyei & Kormos, 1998). Article I demonstrated that different repairs can have different functions, underscoring the importance of analyzing different types of repair separately; the distinction between repetitions vs. false starts and reformulations proved particularly important because the former relates more to filling time with talk and maintaining fluency (considered as one type of stalling mechanism in the classification), while the latter can be viewed as indicators of self-monitoring and indicate an orientation toward accuracy (see Ellis & Barkhuizen, 2005, pp. 149–151; Gilabert, 2007; Kormos, 1999, 2000; Levelt, 1989; Olkkonen, 2017; Olkkonen & Peltonen, 2017; see also Sections 2.2 and 4.3).

drawls; Götz, 2013; Hasselgren, 2002; Wolk *et al.*, 2020 for fillers; Dumont, 2018; Riegenbach, 1991 for both) are also adopted as measures for stalling mechanisms (following Dörnyei & Kormos, 1998). The other type of problem-solving mechanisms included in the classification is CSs, which are used to circumvent lexis-related problems and can thus contribute to maintaining the flow of speech (Dörnyei & Kormos, 1998).

Third, the classification accounts for both monologue and dialogue contexts for studying L2 speech fluency, while previous studies have mostly focused on individual fluency in monologue conditions. The classification does not, however, suggest a purely individual-focused approach to fluency analysis in dialogue conditions; that is, while temporal fluency measures, stalling mechanisms, and CSs can also be examined in the dialogue context, additional measures are needed to capture the interactional dimension of fluency. Simply applying monologue-based utterance fluency measures to the dialogue context would provide only a partial picture of fluency in interactional settings. Thus, the classification reflects a socio-cognitive approach to L2 fluency (see Segalowitz, 2016; see also Galaczi & Taylor, 2018; Roever & Kasper, 2018, for L2 assessment and IC; e.g., Atkinson, 2019; Douglas Fir Group, 2016 for SLA more broadly): Aspects related to individual performance and the joint construction of fluency (McCarthy, 2010) can and should be examined in dialogue data to form a comprehensive picture of fluency.

The classification is presented in Figure 6 on the following page.

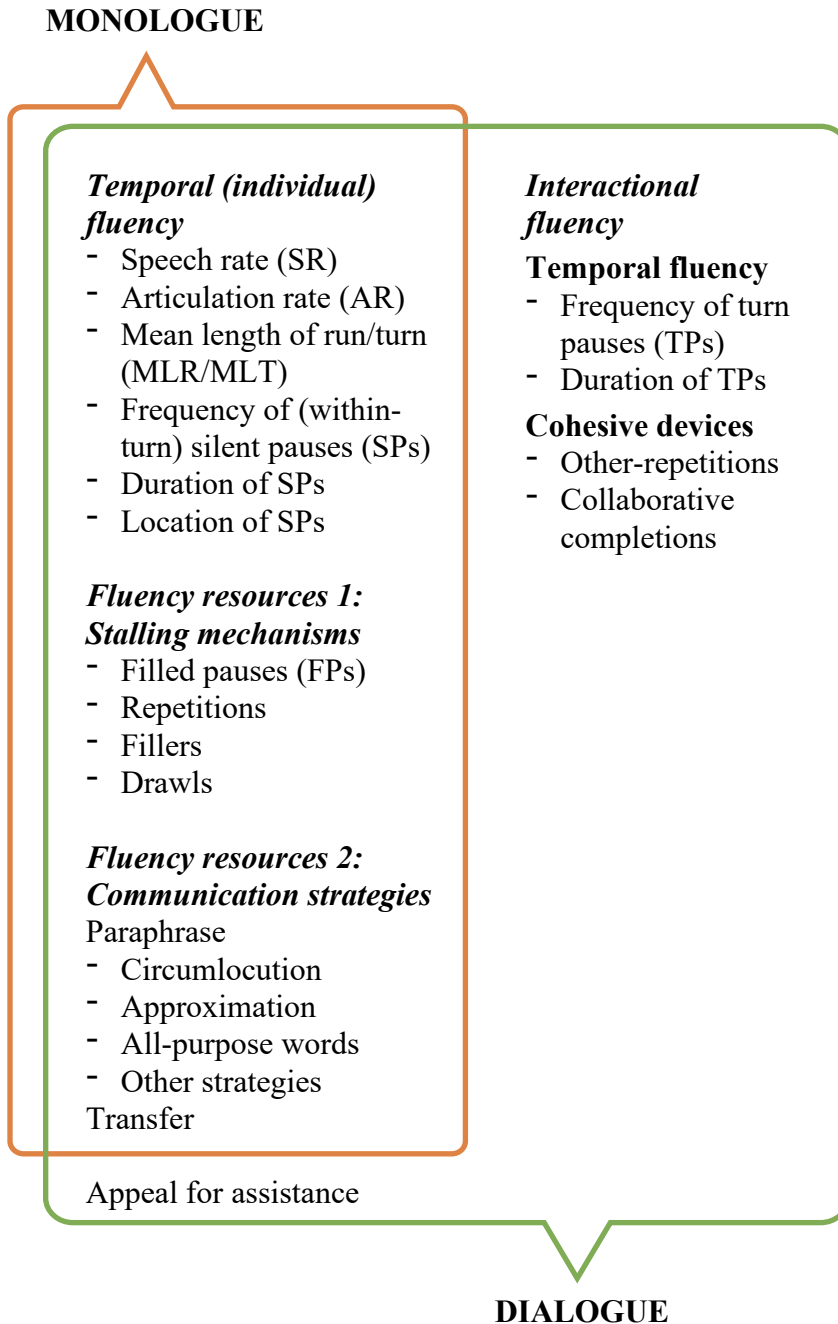


Figure 6. Components and measures of L2 speech fluency from a problem-solving perspective in monologue and dialogue.

As can be seen in Figure 6, the main division is made between two types of data: monologue (orange box) and dialogue (green box). Both types were used in the present study (monologues in Articles I and III and dialogues in Articles II and IV). Despite the basic distinction between monologue and dialogue, most measures can be used in both contexts (as illustrated with overlapping boxes in Figure 6). Three main aspects that can be examined using both monologue and dialogue data are included: temporal fluency, stalling mechanisms, and communication strategies. Temporal fluency consists of speed (AR) and breakdown (different aspects of SPs) dimensions in Skehan's (2003, 2009, 2014; Tavakoli & Skehan, 2005) sense, also incorporating composite measures (SR; MLR in monologue and MLT in dialogue). The composite measure MLR/MLT differs across monologue and dialogue contexts: How long speakers can keep the flow of talk going is operationalized as MLR (the average number of syllables between SPs) in monologue and as MLT (syllables per number of turns) in dialogue. Overall, the temporal measures have been well established in L2 speech monologue fluency research (see Sections 2.2 and 2.3) and have been shown to be suitable for quantitative, group-level comparisons of learners' monologue and dialogue productions in the present study.

To capture resources for maintaining fluency, the classification also includes stalling mechanisms: (the frequencies of) FPs, repetitions, fillers, and drawls. Following the Fluency Resources Framework, these form the first type of problem-solving mechanisms to be studied as contributing to fluency (*fluency resources 1*), the other being communication strategies (*fluency resources 2*).³⁹ Three main types of CSs can be distinguished: paraphrase, transfer (e.g., code-switching), and appeal for assistance. The main categories of paraphrases include circumlocutions, approximations, all-purpose words, and other strategies (see, e.g., Dörnyei & Scott, 1997; for details, see Section 4.3). From a quantitative perspective, the frequencies of CSs can be examined, but as has been shown in the present study, the examination should be complemented with a qualitative analysis of the functions and contexts for CSs (see also Section 6.2). Both stalling mechanisms and CSs can be examined in monologue and dialogue contexts; however, appeals for assistance appear only in the dialogue condition.

In addition to these aspects that are shared across monologue and dialogue contexts, dialogue-specific measures are included in the classification. *Interactional fluency* can be divided into two main components: temporal fluency (including measures for between-turn pause duration and frequency) and cohesive devices

³⁹ In the present study, the frequencies of CSs were relatively low in the monologue condition, and they were thus examined based on only the dialogue data. While the frequencies of CSs may generally be lower in monologic data, they can be studied based on both monologic and dialogic task types.

(including other-repetitions and collaborative completions). These features were chosen to be explored as indicators of interactional fluency in the present study based in part on previous research (e.g., Dings, 2014; Hüttner, 2009; Lerner, 1996; Riggenbach, 1991; Taguchi, 2014) and partly due to being identified as prominent in the data (see also Sections 3.3 and 4.3). It should be noted, however, that because previous studies applying interactional fluency measures to analyze dialogue samples have been rare (but see, e.g., Riggenbach, 1991; Sato, 2014), these measures should be considered tentative, to be confirmed in subsequent studies and possibly complemented by other measures. Yet, this present suggestion can be viewed as providing the basis for operationalizing the interactional dimension of L2 speech fluency in a dialogue context.

As a whole, the classification brings together the problem-solving approach and the interactional and individual fluency components applied in the present study. The measures suggested in the classification can form the basis for a quantitative fluency analysis, with complementary qualitative analyses especially recommended for stalling mechanisms and CSs. Within the context of the present study, the classification was applied to two task types: picture description (monologue task) and a problem-solving task (dialogue task). In future studies, the applicability of the classification to other types of data, including, for instance, less structured dialogue tasks, could be tested and potentially modified for the purposes of analyzing different types of tasks. While temporal measures and stalling mechanisms are likely to occur across a range of task types, the use of CSs might vary across task types (see also Dobao, 2001; Poulisse, 1990). Additional potential indicators of interactional fluency could also be identified based on analyses of different interactional task types, as suggested above.

6.2 Limitations and future directions

In this section, the limitations and generalizability of the results of the present study will be discussed, starting with specific limitations related to the individual substudies and followed by broader considerations regarding the scope of the dissertation as a whole. Then, suggestions for future research will be presented.

The present study aimed to approach L2 speech fluency comprehensively, focusing in particular on under-researched aspects, such as problem-solving mechanisms and indicators of interactional fluency. However, some limitations regarding the individual substudies should be acknowledged. First, while the sample size (varying between 42 and 50 subjects) in the individual substudies examining utterance fluency (Articles I–III) was generally in line with other current L2 speech fluency studies (e.g., De Jong *et al.*, 2015; Kahng, 2014), it can be regarded as relatively small, especially from the perspective of dialogue analysis (21 pairs in

Article II). In particular, the features that are relatively infrequently occurring (e.g., repetitions and fillers) could have revealed more prominent group-level patterns if larger sample sizes had been used (cf. learner corpus studies, e.g., Götz, 2019a, 2019b; Wolk *et al.*, 2020). Yet, in general, the sample sizes can be regarded as sufficient for the purposes of the present study, allowing group-level, quantitative comparisons using a variety of statistical methods. In Article IV, which focused on perceived fluency, the samples (including 15 raters and six rated samples⁴⁰) were also rather small due to its exploratory nature; the study can thus be viewed as providing the basis for further research, exploring the topic, for instance, with other rater populations (teachers and NSs) and other types of L2 samples. For the qualitative analyses of L2 samples in Articles I–III, only a subset of subjects in each study was examined to enable detailed analyses. Thus, the generalizability of the findings regarding the qualitative analysis is relatively limited, and the findings should be considered tentative. However, as the purpose of the qualitative analysis was to illustrate individual variation and analyze certain subjects' productions in detail, the number of subjects can be regarded as appropriate for the aims of the analysis.

Other limitations that apply to the individual substudies include the lack of proficiency-level testing in Article I, which used the CAF corpus that did not include proficiency-level information for the participants. A proficiency-level test (LexTALE) was, nevertheless, included in the data collection for Articles II–IV, and proficiency-level information was reported in the articles. It should be pointed out that because intact school-level groups were maintained in the studies to examine how students at different school levels differed in their fluency, some within-group variation in the vocabulary scores was observed, potentially contributing to the lack of between-group differences. However, especially for correlational analyses, some variation was required. In Article IV, including only one group of raters (university students of English, most of them preservice teachers) can be regarded as a limitation of the study. Raters from other backgrounds, including teachers and/or NSs, could have performed the assessments differently. However, as an exploratory study, the article provided the basis for further studies on perceived interactional fluency in particular. In addition, there is some evidence (Rossiter, 2009) that raters with different backgrounds do not necessarily differ in their ratings. Finally, in relation to

⁴⁰ In particular, the number of rated samples is low. Yet, as mentioned in Section 4.1, the number of rated samples was regarded as large enough to ensure sufficient variation within the sample, but also small enough to avoid rater fatigue (because in addition to numeric ratings, the raters also commented on each sample for each rated dimension). However, it should be acknowledged that the chance of false negatives increases with small sample sizes; ideally, future studies with larger sample sizes should confirm the exploratory findings presented in Article IV.

Articles II–IV, only participants from one L1, Finnish, were included to narrow the scope of the present study (cf. Swedish speakers also included in Article I) and to enable a focus on aspects besides cross-linguistic influence on L2 fluency. However, as Finland provides an optimal context for comparing Swedish speakers and Finnish speakers, who share a similar cultural and educational background but speak typologically different L1s (for details, see Ringbom, 2007), in future studies, Swedish-speaking students of English could be compared with Finnish-speaking students of English. In particular, the links between L1 and L2 fluency (Article III) could be explored among Swedish speakers and Finnish speakers. To my knowledge, Swedish has not yet been studied extensively from an L2 speech fluency perspective (for an early exception, see Lehtonen, 1979).

In addition to the specific limitations discussed above, broader considerations regarding the scope of the present study are discussed in the following paragraphs. While the following aspects can be regarded as limitations of the present study, they also provide aspects to be explored in more detail in future studies. Four central aspects will be focused on to limit the scope of the discussion: multimodality, interactional fluency analysis, strategy analysis, and fluency as an aspect of general oral proficiency.

First, the present study set out to examine L2 speech fluency and resources for maintaining it; consequently, the focus was on utterance fluency analysis (apart from Article IV) and L2 speech analysis, excluding nonverbal aspects, such as gestures and gaze, among other multimodal resources.⁴¹ However, in future studies, especially when L2 speech fluency is applied to dialogue contexts, the importance of nonverbal conduct for fluency should be considered (see also Lintunen *et al.*, 2020b). In L2 fluency studies, multimodal approaches have generally not been employed (but see Guillot, 1999; Götz's 2013 theoretical framework incorporating nonverbal aspects). Similarly, L2-testing researchers studying interactional competence have recently called for more research on the role of nonverbal aspects in IC (Plough, Banerjee, & Iwashita, 2018). For instance, L2 performance may be assessed differently depending on whether the assessments are based on video or audio only (e.g., Gullberg, 1998). Considering CA-SLA studies on L2 interaction in particular, multimodal analyses, in which verbal resources are only one aspect analyzed among other, nonverbal resources (e.g., gestures, gaze, and body posture), are becoming increasingly common, following a broader trend in CA research⁴² (e.g.,

⁴¹ Yet, the study incorporated an analysis of nonverbal conduct in a very specific context—that is, to distinguish appeals for assistance from solitary word searches by analyzing gaze direction (Goodwin & Goodwin, 1986; see also Section 4.3).

⁴² On pioneering multimodal CA research, see in particular Goodwin's (e.g., Goodwin, 2013) and Mondada's (e.g., Mondada, 2016) work.

Mondada, 2016; see also Atkinson, 2019, p. 732). From the perspective of the problem-solving approach adopted in the present study, examining how nonverbal resources may help in maintaining fluency from a broader, interactional viewpoint along with verbal resources would be a particularly fruitful starting point for further study. In fact, the results of a small-scale pilot study (Peltonen, 2020) showed that learners use gestures in conjunction with CSs to maintain the flow of interaction, yet the reliance on compensatory gestures varied across the learners (see also Gullberg, 1998). Importantly, the study showed how a multimodal, video-based analysis can enrich utterance fluency analysis by providing opportunities to analyze, for instance, what the learners do during silences. While this pilot study was based on only two pair-task samples from the dialogue data set used in the present study, it demonstrated the feasibility of approaching fluency from a multimodal perspective and provided a starting point for further studies exploring nonverbal resources along with verbal resources for maintaining fluency.

Second, expanding the scope of L2 speech fluency research also relates to the use of dialogue data and the analysis of interactional fluency along with individual fluency. While the present study was among the first to operationalize interactional fluency and thus provided suggestions for analyzing interactional fluency in a dialogue context with four exploratory measures, the scope for analyzing interactional fluency in the present study can be regarded as relatively narrow. The analysis was mostly quantitative in nature; as the qualitative analyses focused on resources for maintaining fluency, a detailed analysis of the contexts and functions of other-repetitions and collaborative completions was left out of the scope of the present study (see Peltonen, 2017).

As the dialogue task in the present study was a problem-solving task, applying interactional fluency to other task types might help in identifying potential additional indicators of interactional fluency. When other potential indicators of interactional fluency are considered, related lines of study, including studies of IC (e.g., Galaczi, 2014; May *et al.*, 2020; Pekarek Doehler & Pochon-Berger, 2015) and alignment (Costa *et al.*, 2008; Pickering & Garrod, 2004), could provide potential candidates for interactional fluency indicators for further study (see also Section 3.3). Furthermore, because the participants have to alternate between listener and speaker roles in dialogue contexts, more attention could also be paid to the listener's perspective (see Lam, 2018, 2019), perhaps even by adopting the term listening fluency (Anckar & Veivo, 2020).

Third, the analysis of strategic aspects and their relationship to fluency could be studied further in the future. The results of the quantitative and qualitative analyses provided complementary insights into the complex relationship between the two: While a quantitative correlational approach mainly showed a correlation with disfluencies and strategies, the qualitative analysis showed that learners employ

stalling mechanisms to mitigate the effects of these disfluencies. Furthermore, Article IV showed that strategic competence from an assessment perspective was quite often understood in terms of collaborative effort; yet, in the present study, the focus was mainly on examining CSs from an individual learner's perspective. While this approach revealed connections between strategic language use and L2 speech fluency, in future studies, a more interactional analysis that is based on a detailed, qualitative examination could be more fruitful, for instance, in revealing differences across students, as the frequency-based, group-level analysis did not distinguish ninth graders and upper secondary school students in the present study. Studies in the *negotiation of meaning* tradition (Long, 1996; see also e.g., Foster, 1998; Nakahama *et al.*, 2001) could provide potential candidates for interactional strategies that learners use cooperatively in maintaining interactional flow (see also Dörnyei & Kormos, 1998; Van Batenburg *et al.*, 2018).

Finally, as the focus of this dissertation was on L2 speech fluency, it provides insights into one, albeit central, aspect of oral proficiency. Other aspects of L2 proficiency, including complexity and accuracy (following the CAF framework), were left out of the scope of the present study. As the CAF aspects can be viewed as interrelated, some researchers argue that they should be examined together to form a comprehensive picture of L2 proficiency (e.g., Larsen-Freeman, 2009). However, concentrating only on a single dimension of CAF allows for a more thorough examination of that particular dimension, as all aspects of CAF can be viewed as consisting of multiple dimensions (e.g., Norris & Ortega, 2009). If all aspects are analyzed in a single study, the danger is that the different dimensions are operationalized relatively narrowly with few measures. Fluency is a particularly multifaceted concept; especially because fluency was operationalized from a rather broad problem-solving perspective and applied to both monologue and dialogue data, other CAF dimensions did not fit into the scope of the present study. In future studies, in addition to examining the CAF aspects together, task accomplishment or functional/communicative adequacy (Pallotti, 2009; Révész, Ekiert, & Torgersen, 2016) could be taken into account. This would provide more holistic insights into the level of task completion according to the instructions independently of the CAF levels (cf. Article IV results, which revealed that some judges referred to task appropriateness as part of their interactional fluency comments).

6.3 Implications for L2 assessment and teaching

In this section, implications for L2 assessment based on the results of the present study are discussed, followed by a discussion of pedagogical implications. While there are many potential implications for L2 assessment and teaching, to limit the scope of the discussion, three broad themes for both aspects will be highlighted. The

main implications for L2 assessment relate to the interrelated nature of L1 and L2 speech fluency, the joint nature of interactional fluency, and the challenges regarding automatized L2 fluency assessment. For L2 teaching, the main implications relate to the teaching of fluency resources, the teaching of interactional fluency, and the acknowledgment of L1 speaking style in the teaching of L2 fluency.

Starting with the assessment of L2 speech fluency, the results of Article III suggest that L1 fluency influences at least some aspects of L2 speech fluency to a certain extent in monologue production. Because the goal of L2 proficiency assessment is to assess learners based on their skills in the L2, the fact that L1-related skills may influence L2 performance complicates the assessment. While De Jong *et al.* (2015) “partialed out” L1 influence from a measurement perspective in their study, in practice, controlling for L1 influence on L2 performance does not seem feasible, especially in high-stakes testing (see also De Jong, 2018, p. 245). Yet, considering low-stakes testing, such as final course assessments, L1 speaking style could be taken into account in an informal way in the assessment. For instance, if the teachers are familiar with their students and the manner of speaking in their L1 (e.g., lengthy pauses), they could acknowledge this when awarding fluency scores. Comparable L1 samples could also be obtained from the learners to allow for more reliable comparisons across their performances in the L1 and L2. Overall, as pointed out by De Jong (2018, p. 245), it is important for raters to be aware of the potential influence of L1 speaking style on fluency assessments. As other aspects of L2 proficiency are often also assessed along with fluency, the potential for construct-irrelevant variation resulting from speaking-style influence may not be as large a factor with regard to other aspects of proficiency, such as accuracy; however, this is a question to be further examined empirically.

Another challenge for L2 assessment relates to the use of dialogic tasks and the joint nature of interactional fluency. While the goal of L2 assessment is often to evaluate an individual’s competence in the target language (for discussion, see, e.g., Roever & Kasper, 2018), the question of whether some skills should be considered collaborative in nature becomes relevant in conjunction with interactional fluency. Article IV thus explored the possibility of awarding joint scores (see, e.g., May, 2009; Taylor & Wigglesworth, 2009) for interactional fluency along with individual scores for individual fluency. While a thorough discussion of the matter is outside the scope of the present study, a less traditional approach to scoring, involving joint scores, could better reflect the collaborative nature of interactional fluency than an individual score (cf. discussion related to IC in Salaberry & Kunitz, 2019); individual fluency could still be evaluated from an individual’s perspective with an individual score (see also Roever & Kasper, 2018, on combining psycholinguistic–individualist and sociolinguistic–interactional approaches in L2 assessment). Naturally, it should be acknowledged that organizing high-stakes assessment in particular is an

immensely complex endeavor, for which resources, aims, and stakeholders' needs, among others, need to be acknowledged and balanced; thus, joint scoring may not be feasible nor applicable in these contexts.

The final aspect of the implications for L2 assessment relates to the results of the qualitative analyses performed in the present study and the computer-assisted, (semi)-automatized assessment of L2 speaking skills, especially fluency (see also Huhta, Kallio, Ohranen, & Ullakonoja, 2020). The issue is discussed here due to its timeliness in the Finnish context: a speaking component is currently being planned for the upper secondary school Matriculation Examination—a high-stakes test. As the analyses of the present study demonstrated, many of the features that are related to maintaining fluency are multifunctional and require analyses that consider the immediate context of the feature (see also Degand *et al.*, 2019; Dumont, 2018; Lennon, 2000). While many temporal speed and breakdown measures can be considered relatively straightforward indicators of (dis)fluency and can be extracted from speech samples relatively automatically, other features, especially stalling mechanisms, are related to fluency in a more complex manner. Thus, while an automatic assessment of certain temporal features can provide some insights into the learner's fluency and L2 proficiency, the risk in a purely automated assessment of fluency is a relatively narrow operationalization of fluency and limitation to relatively controlled (monologic) tasks. To achieve a comprehensive picture of the learners' competence, (semi)-automated analyses of temporal aspects should be complemented with human raters' assessments, which involve interpreting how multifunctional features in particular are used by the learner and how these features together provide information about the learner's level of L2 fluency. This type of assessment, which combines automatic assessment and human ratings, is being considered for the Matriculation Examination and seems to be a feasible solution in light of the findings of the present study (see also Huhta *et al.*, 2020).

The present study also has implications for the teaching of L2 speaking skills, especially fluency. While the teaching of fluency was not the specific focus in any of the substudies, some of the results in Article II and III in particular could be used in L2 classrooms for pedagogical purposes (for more suggestions on how to support fluency development in L2 teaching, see, e.g., Derwing, 2017; Guillot, 1999; Lintunen *et al.*, 2020b, pp. 196–199; Pakula, 2019; Rossiter, Derwing, Manimtim, & Thomson, 2010). Regarding the Article II results and the use of stalling mechanisms as fluency-enhancing resources, different types of awareness-raising activities could be employed in classrooms to draw learners' attention to the role of filler words, among other resources, in enhancing fluency (see also Dörnyei, 1995; Guillot, 1999). To avoid remaining silent when “getting stuck” during word searches, for instance, learners could be encouraged to use a variety of resources, including stalling mechanisms and CSs, to reduce time spent in silence and to

provide more processing time and attentional resources. Along with the resources examined in the present study, teaching *formulaic sequences* (e.g., Wray, 2005) is likely to contribute to fluency development, because these prefabricated word sequences are retrieved as wholes from memory and can thus facilitate fluency (Pawley & Syder, 1983; Wood, 2006) and automatization from a psycholinguistic perspective (Kormos, 2006).

From the perspective of interactional fluency, acknowledging the collaborative nature of fluency in dialogue contexts is not only important in L2 assessment but should also be reflected in L2 teaching when practicing interactional skills. In addition to practicing strategies for maintaining within-turn fluency, as discussed above, learners could also benefit from instruction that focuses on different ways of maintaining fluency across turns—for instance, by indicating acknowledgment of the interlocutor’s contributions (e.g., other-repetitions and collaborative completions). Practicing the use of other-repetitions and other means for incorporating vocabulary used by the interlocutor in your own turn could help in providing more time and processing resources for planning the rest of the turn (Dörnyei & Kormos, 1998; Peltonen, 2017). Overall, providing opportunities to practice speaking skills, especially with dialogue tasks requiring collaboration (such as the problem-solving task used in the present study), is likely to contribute to both individual and interactional fluency development through automatization. Ideally, ample speaking practice would be combined with focused instruction on fluency-enhancing strategies, including awareness-raising activities adapted to the learners’ level, to facilitate noticing and to draw learners’ attention to both individual and collaborative ways of maintaining fluency (see also Pakula, 2019). For instance, the interactional functions of certain stalling mechanisms (e.g., keeping the floor and signaling to the interlocutor that one intends to continue speaking) could be highlighted and practiced with different collaborative tasks.

The results related to the connections between L1 and L2 fluency in Article III could also inform L2 teaching pedagogy in at least two ways: by influencing the role of native speech and NS models in the classroom, as well as by encouraging more widespread use of learners’ own L1 speech as a pedagogical resource (see also Lintunen *et al.*, 2020b, pp. 196–199). First, to raise awareness of variation in NSs’ fluency, learners could be guided to identify differences across speakers’ fluency, ranging from different speech rates to pausing patterns and the use of particular stalling mechanisms. Noticing that there are different profiles of fluency even in the L1 could contribute to alleviating the learners’ anxiety and help in setting realistic goals for L2 learning (Lintunen *et al.*, 2020b, p. 198). In the next step, learners could compare their own speech in the L1 and L2 (or even compare this to the additional languages in their repertoire). Learners’ recordings of their own speech in the L1 and L2 could be analyzed for fluency-related features. For instance, learners could be

guided to pay attention to idiosyncratic aspects of their speech, such as specific fillers or lengthy pauses. If the learners are relatively advanced, the transcription or annotation of speech with speech analysis software (such as Praat) could be used as a tool for drawing learners' attention to the micro-level details that contribute to fluency in their production (cf. the application of CA to L2 teaching, involving the analyses of transcribed interaction; e.g., Waring, 2019; see also Guillot, 1999). Generally speaking, because audio- and video-recording options are available for every computer or smartphone user, recording and analyzing learners' speech from a fluency perspective could be more widely employed in classrooms.

6.4 Concluding remarks

The present study set out to examine how learners maintain speech fluency in L2 English in monologue and dialogue from a mixed-methods perspective. The general aim was approached from the perspective of three specific research aims filling gaps in L2 fluency research, as outlined in the introduction to this dissertation. The specific research aims related to examining the use of qualitative analysis to complement a quantitative approach and to illustrate individual variation among L2 speakers (aim 1); the application of a novel, problem-solving framework to the analysis of L2 speech fluency and the resources for maintaining it to complement previous analyses focusing on temporal fluency (aim 2); and the use of dialogue data along with the more often studied monologue data (aim 3).

The results related to the research aims highlighted the complex nature of maintaining L2 speech fluency in monologue and dialogue contexts from the quantitative and qualitative perspectives. The qualitative analysis underscored the importance of acknowledging individual differences in maintaining fluency and paying particular attention to the context and functions of fluency-related features. In Articles I–III, triangulating the methods of data analysis contributed to providing a comprehensive picture of fluency and fluency resources: Complementing a quantitative, frequency-based analysis of fluency-related features with a detailed qualitative analysis helped to explain the tendencies demonstrated in the quantitative analyses and showed how learners differed in the resources they employed for maintaining fluency. The Fluency Resources Framework employed in Articles II–IV offered a novel, problem-solving perspective of fluency and the means to maintain it. Applying the framework to monologue and dialogue fluency analysis revealed differences in learners' fluency and the means to maintain fluency at different school levels: More advanced students were generally more fluent. While individual, temporal fluency measures demonstrated the most prominent differences in monologue data, differences in interactional fluency were more prominent in a dialogue context. The L1 fluency measures were also shown to be connected to L2

monologue fluency. Overall, for stalling mechanisms and CSs, despite some group-level patterns being detected in the learners' productions, complementary qualitative analyses focusing on how the features were used, especially in conjunction with other features, were needed to more fully demonstrate their role in maintaining fluent speech. Regarding the dialogue context, the overall results demonstrated the importance of acknowledging aspects of interactional fluency in addition to measures capturing individual fluency. Individual and interactional fluency seem to be interrelated to some extent, especially from a perceived fluency point of view, but can also be studied separately. Both are needed to gain a comprehensive view of fluency in dialogue contexts.

The results related to the three research aims can also be viewed as having broader theoretical–methodological implications for the study of L2 speech fluency. By complementing a quantitative analysis of fluency-related features with a qualitative analysis of their functions and contexts in all the substudies, the present study has aimed to show the benefits of triangulating methods of data analysis by broadening the methodological approaches from purely quantitative toward mixed-methods designs in L2 speech fluency studies. Furthermore, a problem-solving approach was employed to extend the theoretical framework for studying L2 speech fluency to strategic aspects and to enable the analysis of learners' resources for maintaining fluency. This Fluency Resources Framework was successfully applied in Articles II–IV to analyze monologue and dialogue data from both a quantitative and a qualitative perspective and can be used as a basis for studying L2 speech fluency from a problem-solving perspective in future studies.

The analysis of L2 dialogue data in Articles II and IV extended L2 fluency research to a new type of data that enabled the study of a novel, interactional dimension of L2 fluency, capturing learners' collaboration and co-construction of interactional flow along with the means to maintain fluency within individual turns. Taking into account the co-constructed nature of fluency in an interactional setting (following McCarthy's 2010 notion of confluence), the present study was among the first to operationalize interactional fluency along with adapting temporal individual fluency measures to a dialogue context. In the classification presented in Section 6.1.4, it was suggested that interactional fluency can be viewed in terms of two main components: temporal fluency and cohesive devices. Two measures for both components were suggested: frequency and duration of turn pauses for the former and other-repetitions and collaborative completions for the latter. The co-constructed nature of fluency was also explored from a perceived fluency point of view in Article IV, which incorporated judges' assessments of interactional fluency in addition to individual fluency. In both Articles II and IV, the duration of turn pauses in particular emerged as a prominent factor in distinguishing learners. In future studies,

interactional fluency could be studied further in other task types and with other potential measures capturing the two main components (see also Section 6.2).

Considering the recent developments in SLA highlighting social perspectives on L2 learning and proficiency (Atkinson, 2019; Douglas Fir Group, 2016; Ortega, 2019), along with the acknowledgment of the communicative and social aspects of fluency in L2 speech fluency research (Segalowitz, 2016; Wright & Tavakoli, 2016; see also Chapter 1), the present study has provided suggestions for how the scope of L2 speech analysis could be expanded beyond the traditional monologue-based, individual fluency approach. As part of the recent trend of extending L2 fluency research from monologue data to dialogue data (e.g., Sato, 2014; Tavakoli, 2016, 2018; Van Os *et al.*, 2020; Witton-Davies, 2014), the present study has aimed to provide insights into the possibilities of a) analyzing an interactional dimension of fluency and b) integrating both individual and interactional approaches into dialogue data analysis. While traditionally, the psycholinguistic/cognitive approach has been more influential in L2 speech fluency studies, the accumulating knowledge from L2 dialogue fluency analyses and the exploration of the interactional aspects of fluency will hopefully contribute to a shift from purely individualistic approaches toward integrated approaches that acknowledge both the social and cognitive dimensions of L2 speech fluency. While extending fluency research to new avenues brings with it new challenges, especially regarding the implications for teaching and assessment, the benefits of considering fluency from a broader, communicative perspective seem to outweigh the drawbacks. Considering individual learners' contributions to interaction and viewing language use as joint action (Clark, 1996) or the nature of interaction as collaboratively constructed (e.g., He & Young, 1998; Salaberry & Kunitz, 2019) should not be viewed as opposing or mutually exclusive perspectives on L2 speech fluency in interactional contexts. Instead, merging psycholinguistic–individualist and sociolinguistic–interactional approaches (Roever & Kasper, 2018) in L2 speech fluency research could provide the foundation for future fluency studies.

List of References

- Allwood, J., Nivre, J., & Ahlsén, E. (1990). Speech management—On the non-written life of speech. *Nordic Journal of Linguistics*, *13*, 3–48. <http://dx.doi.org/10.1017/S0332586500002092>.
- Anckar, J., & Veivo, O. (2020). Fluency in L2 listening. In P. Lintunen, M. Mutta, & P. Peltonen (Eds.), *Fluency in L2 learning and use* (pp. 49–62). Bristol: Multilingual Matters.
- Atkinson, D. (2019). Beyond the brain: Intercorporeality and co-operative action for SLA studies. *The Modern Language Journal*, *103*, 724–738. <http://dx.doi.org/10.1111/modl.12595>.
- Auer, P., Couper-Kuhlen, E., & Müller, F. (1999). *Language in time: The rhythm and tempo of spoken interaction*. Oxford: Oxford University Press.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Baker-Smemeo, W., Dewey, D. P., Bown, J., & Martinsen, R. A. (2014). Does measuring L2 utterance fluency equal measuring overall L2 proficiency? Evidence from five languages. *Foreign Language Annals*, *47*, 707–728. <http://dx.doi.org/10.1111/flan.12110>.
- Bialystok, E. (1990). *Communication strategies: A psychological analysis of second-language use*. Oxford: Basil Blackwell.
- Boersma, P., & Weenink, D. (2019). *Praat: Doing phonetics by computer* (version 6.0.5) [computer program]. Retrieved from <http://www.praat.org/>.
- Boomer, D. S., & Dittmann, A. T. (1962). Hesitation pauses and juncture pauses in speech. *Language and Speech*, *5*, 215–220. <http://dx.doi.org/10.1177/002383096200500404>.
- Borger, L. (2019). Assessing interactional skills in a paired speaking test: Raters' interpretation of the construct. *Apples – Journal of Applied Language Studies*, *13*, 151–174. <http://dx.doi.org/10.17011/apples/urn.201903011694>.
- Borràs, J., & Llanes, À. (2019). Re-examining the impact of study abroad on L2 development: A critical overview. *The Language Learning Journal*. Advance online publication. <http://dx.doi.org/10.1080/09571736.2019.1642941>.
- Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech*, *44*, 123–147. <http://dx.doi.org/10.1177/00238309010440020101>.
- Bosker, H. R., Pinget, A.-F., Quené, H., Sanders, T., & De Jong, N. H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, *30*, 159–175. <http://dx.doi.org/10.1177/0265532212455394>.
- Bradlow, A. R., Kim, M., & Blasingame, M. (2017). Language-independent talker-specificity in first-language and second-language speech production by bilingual talkers: L1 speaking rate predicts L2 speaking rate. *Journal of the Acoustical Society of America*, *141*, 886–899. <http://dx.doi.org/10.1121/1.4976044>.
- Brand, C., & Götz, S. (2011). Fluency versus accuracy in advanced spoken learner language: A multi-method approach. *International Journal of Corpus Linguistics*, *16*, 255–275. <http://dx.doi.org/10.1075/ijcl.16.2.05bra>.

- Campione, E., & Véronis, J. (2002). A large-scale multilingual study of silent pause duration. In B. Bel & I. Marlien (Eds.), *Proceedings of the Speech Prosody 2002 Conference* (pp. 199–202). Aix-en-Provence: Laboratoire Parole et Langage.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, *1*, 1–47. <http://dx.doi.org/10.1093/applin/I.1.1>.
- Carroll, D. (2004). Restarts in novice turn beginnings: Disfluencies or interactional achievements? In R. Gardner & J. Wagner (Eds.), *Second language conversations* (pp. 201–220). London: Continuum.
- Celce-Murcia, M., Dörnyei, Z., & Thurrell, S. (1995). Communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied Linguistics*, *6*, 5–35. Retrieved from <https://escholarship.org/uc/item/2928w4zj>.
- Chambers, F. (1997). What do we mean by fluency? *System*, *25*, 535–544. [http://dx.doi.org/10.1016/S0346-251X\(97\)00046-8](http://dx.doi.org/10.1016/S0346-251X(97)00046-8).
- Chen, S.-Q. (1990). A study of communication strategies in interlanguage production by Chinese EFL learners. *Language Learning*, *40*, 155–187. <http://dx.doi.org/10.1111/j.1467-1770.1990.tb01332.x>.
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Clark, H. H. (2002). Speaking in time. *Speech Communication*, *36*, 5–13. [http://dx.doi.org/10.1016/S0167-6393\(01\)00022-X](http://dx.doi.org/10.1016/S0167-6393(01)00022-X).
- Clark, H. H., & Fox Tree, J. E. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition*, *84*, 73–111. [http://dx.doi.org/10.1016/S0010-0277\(02\)00017-3](http://dx.doi.org/10.1016/S0010-0277(02)00017-3).
- Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, *37*, 201–242. <http://dx.doi.org/10.1006/cogp.1998.0693>.
- Corley, M., & Stewart, O. W. (2008). Hesitation disfluencies in spontaneous speech: The meaning of *um*. *Language and Linguistics Compass*, *2*, 589–602. <http://dx.doi.org/10.1111/j.1749-818x.2008.00068.x>.
- Costa, A., Pickering, M. J., & Sorace, A. (2008). Alignment in second language dialogue. *Language and Cognitive Processes*, *23*, 528–556. <http://dx.doi.org/10.1080/01690960801920545>.
- Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Csépes, I. (2009). *Measuring oral proficiency through paired-task performance*. Frankfurt am Main: Peter Lang.
- Cucchiari, C., Strik, H., & Boves, L. (2002). Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *Journal of the Acoustical Society of America*, *111*, 2862–2873. <http://dx.doi.org/10.1121/1.1471894>.
- Davis, L. (2009). The influence of interlocutor proficiency in a paired oral assessment. *Language Testing*, *26*, 367–396. <http://dx.doi.org/10.1177/0265532209104667>.
- De Jong, N. H. (2016a). Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching*, *54*, 113–132. <http://dx.doi.org/10.1515/iral-2016-9993>.
- De Jong, N. H. (2016b). Fluency in second language assessment. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 203–218). Berlin: De Gruyter.
- De Jong, N. H. (2018). Fluency in second language testing: Insights from different disciplines. *Language Assessment Quarterly*, *15*, 237–254. <http://dx.doi.org/10.1080/15434303.2018.1477780>.
- De Jong, N. H., & Bosker, H. R. (2013). Choosing a threshold for silent pauses to measure second language fluency. In R. Eklund (Ed.), *Proceedings of DiSS 2013: The 6th workshop on disfluency in spontaneous speech* (pp. 17–20). Stockholm: Royal Institute of Technology (KTH). Retrieved from https://www.isca-speech.org/archive/diss_2013/dis6_017.html.
- De Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. J. (2015). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for

- first language behavior. *Applied Psycholinguistics*, 36, 223–243. <http://dx.doi.org/10.1017/S0142716413000210>.
- De Jong, N. H., & Mora, J. (2019). Does having good articulatory skills lead to more fluent speech in first and second languages? *Studies in Second Language Acquisition*, 41, 227–239. <http://dx.doi.org/10.1017/S0272263117000389>.
- De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, 41, 385–390. <http://dx.doi.org/10.3758/BRM.41.2.385>.
- Degand, L., Gilquin, G., Meurant, & Simon, A. C. (2019). Introduction. In L. Degand, G. Gilquin, L. Meurant, & A. C. Simon (Eds.), *Fluency and disfluency across languages and language varieties* (pp. 9–14). Louvain-la-neuve: Presses Universitaires de Louvain.
- Derwing, T. M. (2017). L2 fluency development. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 246–259). New York: Routledge.
- Derwing, T. M., Munro, M. J., & Thomson, R. I. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, 29, 359–380. <http://dx.doi.org/10.1093/applin/amm041>.
- Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, 31, 533–557. <http://dx.doi.org/10.1017/S0272263109990015>.
- Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language Learning*, 54, 655–679. <http://dx.doi.org/10.1111/j.1467-9922.2004.00282.x>.
- Dings, A. (2014). Interactional competence and the development of alignment activity. *The Modern Language Journal*, 98, 742–756. <http://dx.doi.org/10.1111/j.1540-4781.2014.12120.x>.
- Dobao, A. M. F. (2001). Communication strategies in the interlanguage of Galician students of English: The influence of learner- and task related factors. *Atlantis*, 23(1), 41–62. Retrieved from <https://www.atlantisjournal.org/index.php/atlantis/issue/archive?issue=23-1>.
- Douglas Fir Group (2016). A transdisciplinary framework for SLA in a multilingual world. *The Modern Language Journal*, 100, 19–47. <http://dx.doi.org/10.1111/modl.12301>.
- Ducasse, A. M., & Brown, A. (2009). Assessing paired orals: Raters' orientation to interaction. *Language Testing*, 26, 423–443. <http://dx.doi.org/10.1177/0265532209104669>.
- Dumont, A. (2018). *Fluency and disfluency: A corpus study of non-native and native speaker (dis)fluency profiles* (Unpublished doctoral dissertation). Université catholique de Louvain, Louvain-la-Neuve.
- Duran-Karaoz, Z., & Tavakoli, P. (2020). Predicting L2 fluency from L1 fluency behavior: The case of L1 Turkish and L2 English speakers. *Studies in Second Language Acquisition*. Advance online publication. <http://dx.doi.org/10.1017/S0272263119000755>.
- Dörnyei, Z. (1995). On the teachability of communication strategies. *TESOL Quarterly*, 29, 55–85. <http://dx.doi.org/10.2307/3587805>.
- Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.
- Dörnyei, Z., & Kormos, J. (1998). Problem-solving mechanisms in L2 communication: A psycholinguistic perspective. *Studies in Second Language Acquisition*, 20, 349–385. <http://dx.doi.org/10.1017/S0272263198003039>.
- Dörnyei, Z., & Scott, M. L. (1997). Communication strategies in a second language: Definitions and taxonomies. *Language Learning*, 47, 173–210. <http://dx.doi.org/10.1111/0023-8333.51997005>.
- Ejzenberg, R. (2000). The juggling act of oral fluency: A psycho-sociolinguistic metaphor. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 287–313). Ann Arbor: University of Michigan.
- Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford: Oxford University Press.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (revised edition). Cambridge, MA: MIT Press.

- Fehringer, C., & Fry, C. (2007). Hesitation phenomena in the language production of bilingual speakers: The role of working memory. *Folia Linguistica*, 41, 37–72. <http://dx.doi.org/10.1515/flin.41.1-2.37>.
- Fillmore, C. J. (1979). On fluency. In C. J. Fillmore, D. Kempler, & W. S-Y. Wang (Eds.), *Individual differences in language ability and language behavior* (pp. 85–102). New York: Academic Press.
- Finnish National Agency for Education (2019). Lukion opetussuunnitelman perusteet 2019 [National core curriculum for upper secondary schools 2019]. Retrieved from https://www.oph.fi/sites/default/files/documents/lukion_opetussuunnitelman_perusteet_2019.pdf.
- Firth, A., & Wagner, J. (1997). On discourse, communication, and (some) fundamental concepts in SLA research. *The Modern Language Journal*, 81, 285–300. <http://dx.doi.org/10.2307/329302>.
- Foster, P. (1998). A classroom perspective on the negotiation of meaning. *Applied Linguistics*, 19, 1–23. <http://dx.doi.org/10.1093/applin/19.1.1>.
- Foster, P. (2020). Oral fluency in a second language: A research agenda for the next ten years. *Language Teaching*. Advance online publication. <http://dx.doi.org/10.1017/S026144482000018X>.
- Foster, P., & Skehan, P. (1999). The influence of source of planning and focus of planning on task-based performance. *Language Teaching Research*, 3(3), 215–247. <http://dx.doi.org/10.1177/136216889900300303>.
- Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21, 354–375. <http://dx.doi.org/10.1093/applin/21.3.354>.
- Freed, B. F. (1995). What makes us think that students who study abroad become fluent? In B. F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 123–148). Amsterdam: John Benjamins.
- Fulcher, G. (2003). *Testing second language speaking*. Harlow: Longman.
- Færch, C., & Kasper, G. (1983). Plans and strategies in foreign language communication. In C. Færch & K. Gabriele (Eds.), *Strategies in interlanguage communication* (pp. 20–60). London: Longman.
- Galaczi, E. D. (2014). Interactional competence across proficiency levels: How do learners manage interaction in paired speaking tests? *Applied Linguistics*, 35, 553–574. <http://dx.doi.org/10.1093/applin/amt017>.
- Galaczi, E. D., & Taylor, L. (2018). Interactional competence: Conceptualisations, operationalisations, and outstanding questions. *Language Assessment Quarterly*, 15, 219–236. <http://dx.doi.org/10.1080/15434303.2018.1453816>.
- Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Gilabert, R. (2007). Effects of manipulating task complexity on self-repairs during L2 oral production. *International Review of Applied Linguistics in Language Teaching*, 45, 215–240. <http://dx.doi.org/10.1515/IRAL.2007.010>.
- Gilabert, R., Barón, J., & Levkina, M. (2011). Manipulating task complexity across task types and modes. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 105–138). Amsterdam: John Benjamins.
- Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1–68). Cambridge: Cambridge University Press.
- Ginther, A., Dimova, S., & Yang, R. (2010). Conceptual and empirical relationships between temporal measures of fluency and oral English proficiency with implications for automated scoring. *Language Testing*, 27, 379–399. <http://dx.doi.org/10.1177/0265532210364407>.
- Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. New York: Academic Press.
- Goodwin, C. (2013). The co-operative, transformative organization of human action and knowledge. *Journal of Pragmatics*, 46, 8–23. <http://dx.doi.org/10.1016/j.pragma.2012.09.003>.
- Goodwin, M. H., & Goodwin, C. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica*, 62, 51–76. <http://dx.doi.org/10.1515/semi.1986.62.1-2.51>.

- Gráf, T. & Huang, L.-F. (2019). Repeats in native and learner English. In L. Degand, G. Gilquin, L. Meurant, & A. C. Simon (Eds.), *Fluency and disfluency across languages and language varieties* (pp. 219–241). Louvain-la-neuve: Presses Universitaires de Louvain.
- Grosjean, F. (1980). Temporal variables within and between languages. In H. W. Dechert & M. Raupach (Eds.), *Towards a cross-linguistic assessment of speech production* (pp. 39–53). Frankfurt am Main, Bern & Cirencester: Peter D. Lang.
- Guillot, M.-N. (1999). *Fluency and its teaching*. Clevedon: Multilingual Matters.
- Gullberg, M. (1998). *Gesture as a communication strategy in second language discourse: A study of learners of French and Swedish*. Lund: Lund University Press.
- Guz, E. (2015). Establishing the fluency gap between native and non-native speech. *Research in Language*, 13, 230–247. <http://dx.doi.org/10.1515/rela-2015-0021>.
- Götz, S. (2013). *Fluency in native and nonnative English speech*. Amsterdam: John Benjamins.
- Götz, S. (2019a). Do learning context variables have an effect on learners' (dis)fluency? Language-specific vs. universal patterns in advanced learners' use of filled pauses. In L. Degand, G. Gilquin, L. Meurant, & A. C. Simon (Eds.), *Fluency and disfluency across languages and language varieties* (pp. 177–196). Louvain-la-neuve: Presses Universitaires de Louvain.
- Götz, S. (2019b). Filled pauses across proficiency levels, L1s and learning context variables: A multivariate exploration of the *Trinity Lancaster Corpus Sample*. *International Journal of Learner Corpus Research*, 5, 159–180. <http://dx.doi.org/10.1075/ijlcr.17018.got>.
- Hall, J. K., & Pekarek Doehler, S. (2011). L2 interactional competence and development. In J. K. Hall, J. Hellermann, & S. Pekarek Doehler (Eds.), *L2 interactional competence and development* (pp. 1–15). Bristol: Multilingual Matters.
- Hasselgren, A. (2002). Learner corpora and language testing: Smallwords as markers of learner fluency. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 143–173). Amsterdam: John Benjamins.
- Hawkins, P. R. (1971). The syntactic location of hesitation pauses. *Language and Speech*, 14, 277–288. <http://dx.doi.org/10.1177/002383097101400308>.
- He, A. W., & Young, R. (1998). Language proficiency interviews: A discourse approach. In R. Young & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 1–24). Amsterdam: John Benjamins.
- Hildén, R. (2000). *Att tala bra, bättre och bäst: Suomenkielisten abiturienttien ruotsin kielen taito testisuoritusten valossa* [To talk well, better, the best: Finnish-speaking final-year upper secondary school students' skills in Swedish in light of test performances]. Helsinki: University of Helsinki.
- Hilton, H. (2008). The link between vocabulary knowledge and spoken L2 fluency. *Language Learning Journal*, 36, 153–166. <http://dx.doi.org/10.1080/09571730802389983>.
- Hilton, H. (2009). Annotation and analyses of temporal aspects of spoken fluency. *Calico Journal*, 26, 644–661. <http://dx.doi.org/10.1558/cj.v26i3.644-661>.
- Hilton, H. (2014). Oral fluency and spoken proficiency: Considerations for research and testing. In P. Leclercq, A. Edmonds, & H. Hilton (Eds.), *Measuring L2 proficiency: Perspectives from SLA* (pp. 27–53). Bristol: Multilingual Matters.
- Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30, 461–473. <http://dx.doi.org/10.1093/applin/amp048>.
- Housen, A., Kuiken, F., & Vedder, I. (Eds.) (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. Amsterdam: John Benjamins.
- Huensch, A., & Tracy-Ventura, N. (2017a). Understanding second language fluency behavior: The effects of individual differences in first language fluency, cross-linguistic differences, and proficiency over time. *Applied Psycholinguistics*, 38, 755–785. <http://dx.doi.org/10.1017/S0142716416000424>.
- Huensch, A., & Tracy-Ventura, N. (2017b). L2 utterance fluency development before, during, and after residence abroad: A multidimensional investigation. *The Modern Language Journal*, 101, 275–293. <http://dx.doi.org/10.1111/modl.12395>.

- Huhta, A., Kallio, H., Ohanen, S., & Ullakonoja, R. (2020). Fluency in language assessment. In P. Lintunen, M. Mutta, & P. Peltonen (Eds.), *Fluency in L2 learning and use* (pp. 129–145). Bristol: Multilingual Matters.
- Hüttner, J. (2009). Fluent speakers—fluent interactions: On the creation of (co-)fluency in English as a lingua franca. In A. Mauranen & E. Ranta (Eds.), *English as a lingua franca: Studies and findings* (pp. 274–297). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Hynninen, N. (2020). Fluency in English as a lingua franca interaction. In P. Lintunen, M. Mutta, & P. Peltonen (Eds.), *Fluency in L2 learning and use* (pp. 81–95). Bristol: Multilingual Matters.
- Iwashita, N. (1996). The validity of the paired interview format in oral performance assessment. *Melbourne Papers in Language Testing*, 5(2), 51–66.
- Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics*, 29, 24–49. <http://dx.doi.org/10.1093/applin/amm017>.
- Jacewicz, E., Fox, R. A., & Wei, L. (2010). Between-speaker and within-speaker variation in speech tempo of American English. *Journal of the Acoustical Society of America*, 128, 839–850. <http://dx.doi.org/10.1121/1.3459842>.
- Jacoby, S., & Ochs, E. (1995). Co-construction: An introduction. *Research on Language and Social Interaction*, 28, 171–183. http://dx.doi.org/10.1207/s15327973rlsi2803_1.
- Jaffe, J., & Feldstein, S. (1970). *Rhythms of dialogue*. New York: Academic Press.
- Jarvis, S. (2017). Grounding lexical diversity in human judgments. *Language Testing*, 34, 537–553. <http://dx.doi.org/10.1177/0265532217710632>.
- Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning*, 64, 809–854. <http://dx.doi.org/10.1111/lang.12084>.
- Kahng, J. (2018). The effect of pause location on perceived fluency. *Applied Psycholinguistics*, 39, 569–591. <http://dx.doi.org/10.1017/S0142716417000534>.
- Kahng, J. (2020). Explaining second language utterance fluency: Contribution of cognitive fluency and first language utterance fluency. *Applied Psycholinguistics*, 41, 457–480. <http://dx.doi.org/10.1017/S0142716420000065>.
- Kasper, G., & Kellerman, E. (1997). Introduction: Approaches to communication strategies. In G. Kasper & E. Kellerman (Eds.), *Communication strategies: Psycholinguistic and sociolinguistic perspectives* (pp. 1–13). London: Longman.
- Kasper, G., & Wagner, J. (2011). A conversation-analytic approach to second language acquisition. In D. Atkinson (Ed.), *Alternative approaches to second language acquisition* (pp. 117–142). New York: Routledge.
- Kirk, S., & Carter, R. (2010). Fluency and spoken English. In M. Moreno Jaén, F. Serrano Valverde, & M. Calzada Pérez (Eds.), *Exploring new paths in language pedagogy: Lexis and corpus-based language teaching* (pp. 25–38). London: Equinox.
- Klippel, F. (1984). *Keep talking: Communicative fluency activities for language teaching*. Cambridge: Cambridge University Press.
- Kolly, M.-J., Leemann, A., Boula de Mareuil, P., & Dellwo, V. (2015). Speaker idiosyncrasy in pausing behavior: Evidence from a cross-linguistic study. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th international congress of phonetic sciences* (pp. 1–5). Glasgow: The University of Glasgow.
- Koponen, M., & Riggenbach, H. (2000). Overview: Varying perspectives on fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 5–24). Ann Arbor: University of Michigan Press.
- Kormos, J. (1998). Verbal reports in L2 speech production research. *TESOL Quarterly*, 32, 353–358. <http://dx.doi.org/10.2307/3587590>.
- Kormos, J. (1999). The effect of speaker variables on the self-correction behaviour of L2 learners. *System*, 27, 207–221. [http://dx.doi.org/10.1016/S0346-251X\(99\)00017-2](http://dx.doi.org/10.1016/S0346-251X(99)00017-2).

- Kormos, J. (2000). The role of attention in monitoring second language speech production. *Language Learning*, 50, 343–384. <http://dx.doi.org/10.1111/0023-8333.00120>.
- Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32, 145–164. <http://dx.doi.org/10.1016/j.system.2004.01.001>.
- Kosmala, L., & Morgenstern, A. (2019) Should ‘uh’ and ‘um’ be categorized as markers of disfluency? The use of fillers in a challenging conversational context. In L. Degand, G. Gilquin, L. Meurant, & A. C. Simon (Eds.), *Fluency and disfluency across languages and language varieties* (pp. 67–89). Louvain-la-neuve: Presses Universitaires de Louvain.
- Kramsch, C. (1986). From language proficiency to interactional competence. *The Modern Language Journal*, 70, 366–372. <http://dx.doi.org/10.2307/326815>.
- Lam, D. M. K. (2018). What counts as “responding”? Contingency on previous speaker contribution as a feature of interactional competence. *Language Testing*, 35, 377–401. <http://dx.doi.org/10.1177/0265532218758126>.
- Lam, D. M. K. (2019). Interactional competence with and without extended planning time in a group oral assessment. *Language Assessment Quarterly*, 16, 1–20. <http://dx.doi.org/10.1080/15434303.2019.1602627>.
- Lambert, C., Kormos, J., & Minn, D. (2017). Task repetition and second language speech processing. *Studies in Second Language Acquisition*, 39, 167–196. <http://dx.doi.org/10.1017/S0272263116000085>.
- Larsen-Freeman, D. (2009). Adjusting expectations: The study of complexity, accuracy and fluency in second language acquisition. *Applied Linguistics*, 30, 579–589. <http://dx.doi.org/10.1093/applin/amp043>.
- Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York: Routledge.
- Lauranto, Y. (2005). Sujuvuuden mittoja [Measures of fluency]. In L. Kuure, E. Kärkkäinen, & M. Saarenkunnas (Eds.), *Kieli ja sosiaalinen toiminta – Language and social action* (pp. 127–147). AFinLA Yearbook 2005. Jyväskylä: Finnish Association of Applied Linguistics AFinLA.
- Lehtonen, J. (1979). Speech rate and pauses in the English of Finns, Swedish-speaking Finns, and Swedes. In R. Palmberg (Ed.), *Perception and production of English: Papers on interlanguage* (pp. 35–51). Turku: Åbo Akademi University.
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, 44, 325–343. <http://dx.doi.org/10.3758/s13428-011-0146-0>.
- Lennes, M. (2002). Total duration of labeled segments. Praat script. Retrieved from https://github.com/lennes/spect/blob/master/scripts/total_duration_of_labeled_segments.praat.
- Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40, 387–417. <http://dx.doi.org/10.1111/j.1467-1770.1990.tb00669.x>.
- Lennon, P. (2000). The lexical element in spoken second language fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 25–42). Ann Arbor: University of Michigan Press.
- Leppänen, S., Pitkänen-Huhta, A., Nikula, T., Kytölä, S., Törmäkangas, T., Nissinen, K., ... Jousmäki, H. (2011). *National survey on the English language in Finland: Uses, meanings and attitudes*. Studies in Variation, Contacts and Change in English, Volume 5. Retrieved from www.helsinki.fi/varieng/journal/volumes/05/evarieng-vol5.pdf.
- Lerner, G. H. (1996). On the “semi-permeable” character of grammatical units in conversation: Conditional entry into the turn space of another speaker. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and grammar* (pp. 238–276). Cambridge: Cambridge University Press.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

- Lintunen, P., Mutta, M., & Peltonen, P. (2020a). Defining fluency in L2 learning and use. In P. Lintunen, M. Mutta, & P. Peltonen (Eds.), *Fluency in L2 learning and use* (pp. 1–15). Bristol: Multilingual Matters.
- Lintunen, P., Mutta, M., & Peltonen, P. (2020b). Synthesising approaches to second language fluency: Implications and future directions. In P. Lintunen, M. Mutta, & P. Peltonen (Eds.), *Fluency in L2 learning and use* (pp. 186–201). Bristol: Multilingual Matters.
- Lintunen, P., & Peltonen, P. (2020). What is fluency? Learner perceptions of the concept. In P. Lintunen, M. Mutta, & P. Peltonen (Eds.), *Fluency in L2 learning and use* (pp. 16–33). Bristol: Multilingual Matters.
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). New York: Academic Press.
- MacLay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, 15, 19–44. <http://dx.doi.org/10.1080/00437956.1959.11659682>.
- Magne, V., Suzuki, S., Suzukida, Y., Saito, K., Ilkan, M. & Tran, M. (2019). Exploring the dynamic nature of second language listeners' perceived fluency: A mixed-methods approach. *TESOL Quarterly*, 53, 1139–1150. <http://dx.doi.org/10.1002/tesq.528>.
- Marshall, R. C. (2000). Speech fluency and aphasia. In H. Riggensbach (Ed.), *Perspectives on fluency* (pp. 74–88). Ann Arbor: University of Michigan Press.
- May, L. (2009). Co-constructed interaction in a paired speaking test: The rater's perspective. *Language Testing*, 26, 397–421. <http://dx.doi.org/10.1177/0265532209104668>.
- May, L., Nakatsuhara, F., Lam, D., & Galaczi, E. (2020). Developing tools for learning oriented assessment of interactional competence: Bridging theory and practice. *Language Testing*, 37, 165–188. <http://dx.doi.org/10.1177/0265532219879044>.
- McCarthy, M. (1998). *Spoken language and applied linguistics*. Cambridge: Cambridge University Press.
- McCarthy, M. (2010). Spoken fluency revisited. *English Profile Journal*, 1(1), 1–15. <http://dx.doi.org/10.1017/S2041536210000012>.
- McNamara, T. F. (1997). 'Interaction' in second language performance assessment: Whose performance? *Applied Linguistics*, 18, 446–466. <http://dx.doi.org/10.1093/applin/18.4.446>.
- Michel, M. (2011). Effects of task complexity and interaction on L2 performance. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 141–173). Amsterdam: John Benjamins.
- Ministry of Education and Culture (2017). *Gaudeamus igitur - ylioppilastutkinnon kehittäminen* [Gaudeamus igitur - The development of the Matriculation Examination]. Opetus- ja kulttuuriministeriön julkaisuja 2017:16. Helsinki. Retrieved from <http://urn.fi/URN:ISBN:978-952-263-462-7>.
- Ministry of Education and Culture (2018). *Finnish education in a nutshell*. Retrieved from <https://www.oph.fi/en/statistics-and-publications/publications/finnish-education-nutshell>.
- Mondada, L. (2016). Challenges of multimodality: Language and the body in social interaction. *Journal of Sociolinguistics*, 20, 336–366. http://dx.doi.org/10.1111/josl.1_12177.
- Moniz, H., Batista, F., Mata, A. I., & Trancoso, I. (2014). Speaking style effects in the production of disfluencies. *Speech Communication*, 65, 20–35. <http://dx.doi.org/10.1016/j.specom.2014.05.004>.
- Montero, L., Serrano, R., & Llanes, Á. (2017). The influence of learning context and age on the use of L2 communication strategies. *The Language Learning Journal*, 45, 117–132. <http://dx.doi.org/10.1080/09571736.2013.853824>.
- Mora, J. C., & Valls-Ferrer, M. (2012). Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly*, 46, 610–641. <http://dx.doi.org/10.1002/tesq.34>.

- Möhle, D. (1984). A comparison of the second language speech production of different native speakers. In H. W. Dechert, D. Möhle, & M. Raupach (Eds.), *Second language productions* (pp. 26–49). Tübingen: Gunter Narr Verlag.
- Nakahama, Y., Tyler, A., & van Lier, L. (2001). Negotiation of meaning in conversational and information gap activities: A comparative discourse analysis. *TESOL Quarterly*, 35, 377–405. <http://dx.doi.org/10.2307/3588028>.
- Nakatani, Y., & Goh, C. (2007). A review of oral communication strategies: Focus on interactionist and psycholinguistic perspectives. In A. D. Cohen & E. Macaro (Eds.), *Language learner strategies: Thirty years of research and practice* (pp. 207–227). Oxford: Oxford University Press.
- Nitta, R., & Nakatsuhara, F. (2014). A multifaceted approach to investigating pre-task planning effects on paired oral test performance. *Language Testing*, 31, 147–175. <http://dx.doi.org/10.1177/0265532213514401>.
- Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30, 555–578. <http://dx.doi.org/10.1093/applin/amp044>.
- O’Connell, D. C., & Kowal, S. (2004). The history of research on the filled pause as evidence of *The written language bias in linguistics* (Linell, 1982). *Journal of Psycholinguistic Research*, 33, 459–474. <http://dx.doi.org/10.1007/s10936-004-2666-6>.
- Olkkonen, S. (2017). *Second and foreign language fluency from a cognitive perspective: Inefficiency and control of attention in lexical access*. Jyväskylä: University of Jyväskylä. Retrieved from <http://urn.fi/URN:ISBN:978-951-39-7062-8>.
- Olkkonen, S., & Peltonen, P. (2017). Mitä on toisen kielen sujuvuus? Näkökulmia kognitiivisen ja puheutuoksen sujuvuuden tutkimuksesta. [What is second language fluency? Perspectives from cognitive and utterance fluency research.] In M. Kuronen, P. Lintunen, & T. Nieminen (Eds.), *Näkökulmia toisen kielen puheeseen – Insights into second language speech* (pp. 234–257). AFinLA-e: Soveltavan kielitieteen tutkimuksia 10. Jyväskylä: Finnish Association of Applied Linguistics AFinLA. <http://dx.doi.org/10.30660/afinla.73140>.
- Olynyk, M., d’Anglejan, A., & Sankoff, D. (1990). A quantitative and qualitative analysis of speech markers in the native and second language speech of bilinguals. In R. C. Scarcella, E. S. Andersen, & S. D. Krashen (Eds.), *Developing communicative competence in a second language* (pp. 139–155). Boston, MA: Newbury House.
- Ortega, L. (2019, August). *Orchestrating second language learning in classrooms: Nudging for a sea-change*. Plenary talk presented at the 29th EuroSLA conference, Lund, Sweden.
- Paananen-Porkka, M. M. (2007). *Speech rhythm in an interlanguage perspective: Finnish adolescents speaking English*. Helsinki: University of Helsinki.
- Pakula, H.-M. (2019). Teaching speaking. *Apples – Journal of Applied Language Studies*, 13, 95–111. <http://dx.doi.org/10.17011/apples/urn.201903011691>.
- Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, 30, 590–601. <http://dx.doi.org/10.1093/applin/amp045>.
- Paribakht, T. (1985). Strategic competence and language proficiency. *Applied Linguistics*, 6, 132–146. <http://dx.doi.org/10.1093/applin/6.2.132>.
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191–226). New York: Longman.
- Pekarek Doehler, S. (2019). On the nature and the development of L2 interactional competence: State of the art and implications for praxis. In R. M. Salaberry & S. Kunitz (Eds.), *Teaching and testing L2 interactional competence: Bridging theory and practice* (pp. 25–59). New York: Routledge.
- Pekarek Doehler, S., & Berger, E. (2018). L2 interactional competence as increased ability for context-sensitive conduct: A longitudinal study of story-openings. *Applied Linguistics*, 39, 555–578. <http://dx.doi.org/10.1093/applin/amw021>.

- Pekarek Doehler, S., & Pochon-Berger, E. (2015). The development of L2 interactional competence: Evidence from turn-taking organization, sequence organization, repair organization and preference organization. In T. Cadierno & S. Eskildsen (Eds.), *Usage-based perspectives on second language learning* (pp. 233–267). Berlin: Mouton De Gruyter.
- Peltonen, P. (2017). L2 fluency in spoken interaction: A case study on the use of other-repetitions and collaborative completions. In M. Kuronen, P. Lintunen, & T. Nieminen (Eds.), *Näkökulmia toisen kielen puheeseen – Insights into second language speech* (pp. 118–138). AFinLA-e: Soveltavan kielitieteen tutkimuksia 10. Jyväskylä: Finnish Association of Applied Linguistics AFinLA. <http://dx.doi.org/10.30660/afinla.73130>.
- Peltonen, P. (2020). Gestures as fluency-enhancing resources in L2 interaction: A case study on multimodal fluency. In P. Lintunen, M. Mutta, & P. Peltonen (Eds.), *Fluency in L2 learning and use* (pp. 111–128). Bristol: Multilingual Matters.
- Peltonen, P., & Lintunen, P. (2019, August). *Short-term development of L2 fluency in formal instruction: A mixed methods study*. Paper presented at the 29th EuroSLA conference, Lund, Sweden.
- Penttilä, N. (2019). *Miten puhe sujuu? Tyypillisen ja poikkeavan puheen sujuvuuden piirteitä* [How does speech flow? Fluency features in typical and disordered speech]. Tampere: University of Tampere. Retrieved from <http://urn.fi/URN:ISBN:978-952-03-1132-2>.
- Penttilä, N., Korpjajaakko-Huuhka, A.-M., & Kent, R. (2018). Tavallista sujumattomuutta: aikuisten puheen sujuvuuden kvantitatiivinen analyysi [Typical fluency in Finnish adults]. *Puhe ja kieli*, 38, 153–173. <http://dx.doi.org/10.23997/pk.77384>.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27, 169–226. <http://dx.doi.org/10.1017/S0140525X04000056>.
- Pietilä, P. (1999). *L2 speech: Oral proficiency of students of English at university level*. Anglicana Turkuensia 19. Turku: University of Turku.
- Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64, 878–912. <http://dx.doi.org/10.1111/lang.12079>.
- Plough, I., Banerjee, J., & Iwashita, N. (2018). Interactional competence: Genie out of the bottle. *Language Testing*, 35, 427–445. <http://dx.doi.org/10.1177/0265532218772325>.
- Poulisse, N. (1990). *The use of compensatory strategies by Dutch learners of English*. Dordrecht: Foris.
- Poulisse, N., Bongaerts, T., & Kellerman, E. (1987). The use of retrospective verbal reports in the analysis of compensatory strategies. In C. Færch & G. Kasper (Eds.), *Introspection in second language research* (pp. 213–229). Clevedon: Multilingual Matters.
- Préfontaine, Y. (2013). Perceptions of French fluency in second language speech production. *The Canadian Modern Language Review*, 69, 324–348. <http://dx.doi.org/10.3138/cmlr.1748>.
- Préfontaine, Y., & Kormos, J. (2016). A qualitative analysis of perceptions of fluency in second language French. *International Review of Applied Linguistics in Language Teaching*, 54, 151–169. <http://dx.doi.org/10.1515/iral-2016-9995>.
- Préfontaine, Y., Kormos, J., & Johnson, D. E. (2016). How do utterance measures predict raters’ perceptions of fluency in French as a second language? *Language Testing*, 33, 53–73. <http://dx.doi.org/10.1177/0265532215579530>.
- Quené, H. (2008). Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. *Journal of the Acoustical Society of America*, 123, 1104–1113. <http://dx.doi.org/10.1121/1.2821762>.
- Ramsay, R. W. (1968). Speech patterns and personality. *Language and Speech*, 11, 54–63. <http://dx.doi.org/10.1177/002383096801100108>.
- Raupach, M. (1980). Temporal variables in first and second language speech production. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler* (pp. 263–270). The Hague, Paris & New York: Mouton.

- Révész, A., Ekiert, M., & Torgersen, E. N. (2016). The effects of complexity, accuracy, and fluency on communicative adequacy in oral task performance. *Applied Linguistics*, 37, 828–848. <http://dx.doi.org/10.1093/applin/amu069>.
- Riazantseva, A. (2001). Second language proficiency and pausing: A study of Russian speakers of English. *Studies in Second Language Acquisition*, 23, 497–526. <http://dx.doi.org/10.1017/S027226310100403X>.
- Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes*, 14, 423–441. <http://dx.doi.org/10.1080/01638539109544795>.
- Ringbom, H. (2007). *Cross-linguistic similarity in foreign language learning*. Clevedon: Multilingual Matters.
- Roever, C., & Kasper, G. (2018). Speaking in turns and sequences: Interactional competence as a target construct in testing speaking. *Language Testing*, 35, 331–355. <http://dx.doi.org/10.1177/0265532218758128>.
- Rossiter, M. J. (2009). Perceptions of L2 fluency by native and non-native speakers of English. *The Canadian Modern Language Review*, 65, 395–412. <http://dx.doi.org/10.3138/cmlr.65.3.395>.
- Rossiter, M. J., Derwing, T. M., Manimtim, L. G., & Thomson R. I. (2010). Oral fluency: The neglected component in the communicative language classroom. *The Canadian Modern Language Review*, 66, 583–606. <http://dx.doi.org/10.3138/cmlr.66.4.583>.
- Rühlemann, C. (2006). Coming to terms with conversational grammar: ‘Dislocation’ and ‘dysfluency’. *International Journal of Corpus Linguistics*, 11, 385–409. <http://dx.doi.org/10.1075/ijcl.11.4.03ruh>.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50, 696–735. <http://dx.doi.org/10.1353/lan.1974.0010>.
- Salaberry, R. M., & Kunitz, S. (2019). Introduction. In R. M. Salaberry and S. Kunitz (Eds.), *Teaching and testing L2 interactional competence: Bridging theory and practice* (pp. 1–22). New York: Routledge.
- Sato, M. (2014). Exploring the construct of interactional oral fluency: Second language acquisition and language testing approaches. *System*, 45, 79–91. <http://dx.doi.org/10.1016/j.system.2014.05.004>.
- Schegloff, E. A. (2007). *Sequence organization in interaction*. Cambridge: Cambridge University Press.
- Schegloff, E. A. (2010). Some other “uh(m)”s. *Discourse Processes*, 47, 130–174. <http://dx.doi.org/10.1080/01638530903223380>.
- Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, 14, 357–385. <http://dx.doi.org/10.1017/S0272263100011189>.
- Segalowitz, N. (2000). Automaticity and attentional skill in fluent performance. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 200–219). Ann Arbor: University of Michigan Press.
- Segalowitz, N. (2010). *The cognitive bases of second language fluency*. New York: Routledge.
- Segalowitz, N. (2016). Second language fluency and its underlying cognitive and social determinants. *International Review of Applied Linguistics in Language Teaching*, 54, 79–95. <http://dx.doi.org/10.1515/iral-2016-9991>.
- Segalowitz, N., & Freed, B. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition*, 26, 173–199. <http://dx.doi.org/10.1017/S0272263104262027>.
- Seliger, H. W. (1980). Utterance planning and correction behavior: Its function in the grammar construction process for second language learners. In H. W. Dechert & M. Raupach (Eds.), *Towards a cross-linguistic assessment of speech production* (pp. 87–99). Frankfurt am Main, Bern & Cirencester: Peter D. Lang.
- Shea, C., & Leonard, K. (2019). Evaluating measures of pausing for second language fluency research. *The Canadian Modern Language Review*, 75, 216–235. <http://dx.doi.org/10.3138/cmlr.2018-0258>.
- Skehan, P. (2003). Task-based instruction. *Language Teaching*, 36, 1–14. <http://dx.doi.org/10.1017/S026144480200188X>.

- Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30, 510–532. <http://dx.doi.org/10.1093/applin/amp047>.
- Skehan, P. (2014). The context for researching a processing perspective on task performance. In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 1–26). Amsterdam: John Benjamins.
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., ... Levinson, S. C. (2009). Universals and cultural variation in turn taking in conversation. *PNAS (Proceedings of the National Academy of Sciences)*, 106(26), 10587–10592. <http://dx.doi.org/10.1073/pnas.0903616106>.
- Street, R. L., Jr. (1983). Noncontent speech convergence and divergence in adult–child interactions. In R. N. Bostrom (Ed.), *Communication yearbook 7* (pp. 369–395). New York: Routledge.
- Suzuki, S., & Kormos, J. (2020). Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition*, 42, 143–167. <http://dx.doi.org/10.1017/S0272263119000421>.
- Taguchi, N. (2014). Development of interactional competence in Japanese as a second language: Use of incomplete sentences as interactional resources. *The Modern Language Journal*, 98, 518–535. <http://dx.doi.org/10.1111/j.1540-4781.2014.12087.x>.
- Tannen, D. (1989). *Talking voices: Repetition, dialogue, and imagery in conversational discourse*. Cambridge: Cambridge University Press.
- Tavakoli, P. (2011). Pausing patterns: Differences between L2 learners and native speakers. *ELT Journal*, 65, 71–79. <http://dx.doi.org/10.1093/elt/ccq020>.
- Tavakoli, P. (2016). Fluency in monologic and dialogic task performance: Challenges in defining and measuring L2 fluency. *International Review of Applied Linguistics in Language Teaching*, 54, 133–150. <http://dx.doi.org/10.1515/iral-2016-9994>.
- Tavakoli, P. (2018). L2 development in an intensive study abroad EAP context. *System*, 72, 62–74. <http://dx.doi.org/10.1016/j.system.2017.10.009>.
- Tavakoli, P., Campbell, C., & McCormack, J. (2016). Development of speech fluency over a short period of time: Effects of pedagogic intervention. *TESOL Quarterly*, 50, 447–471. <http://dx.doi.org/10.1002/tesq.244>.
- Tavakoli, P., & Hunter, A.-M. (2018). Is fluency being ‘neglected’ in the classroom? Teacher understanding of fluency and related classroom practices. *Language Teaching Research*, 22, 330–349. <http://dx.doi.org/10.1177/1362168817708462>.
- Tavakoli, P., Nakatsuhara, F., & Hunter, A.-M. (2020). Aspects of fluency across assessed levels of speaking proficiency. *The Modern Language Journal*, 104, 169–191. <http://dx.doi.org/10.1111/modl.12620>.
- Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–273). Amsterdam: John Benjamins.
- Taylor, L., & Wigglesworth, G. (2009). Are two heads better than one? Pair work in L2 assessment contexts. *Language Testing*, 26, 325–339. <http://dx.doi.org/10.1177/0265532209104665>.
- Toivola, M., Lennes, M., & Aho, E. (2009). Speech rate and pauses in non-native Finnish. In *Proceedings of Interspeech 2009: The 10th annual conference of the International Speech Communication Association (ISCA)* (pp. 1707–1710). Brighton: ISCA. Retrieved from https://www.isca-speech.org/archive/interspeech_2009/i09_1707.html.
- Tonkyn, A. (2012). Measuring and perceiving changes in oral complexity, accuracy and fluency: Examining instructed learners’ short-term gains. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 221–246). Amsterdam: John Benjamins.

- Towell, R. (2002). Relative degrees of fluency: A comparative case study of advanced learners of French. *International Review of Applied Linguistics in Language Teaching*, 40, 117–150. <http://dx.doi.org/10.1515/iral.2002.005>.
- Towell, R., & Dewaele, J.-M. (2005). The role of psycholinguistic factors in the development of fluency. In J.-M. Dewaele (Ed.), *Focus on French as a foreign language* (pp. 210–239). Bristol, UK: Multilingual Matters.
- Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, 17, 84–119. <http://dx.doi.org/10.1093/applin/17.1.84>.
- Ullakonoja, R. (2011). *Da. Eto vopros! Prosodic development of Finnish students' read-aloud Russian during study in Russia*. Jyväskylä: University of Jyväskylä. Retrieved from <http://urn.fi/URN:ISBN:978-951-39-4209-0>.
- University of Turku (2019). Research at the Department of English: Complexity, accuracy and fluency (CAF) in L2 English. Retrieved from <https://www.utu.fi/en/university/faculty-of-humanities/english/research>.
- Ur, P. (1990). *Discussions that work: Task-centered fluency practice*. Cambridge: Cambridge University Press.
- Van Batenburg, E. S. L., Oostdam, R. J., Van Gelderen, A. J. S., & De Jong, N. H. (2018). Measuring L2 speakers' interactional ability using interactive speech tasks. *Language Testing*, 35, 75–100. <http://dx.doi.org/10.1177/0265532216679452>.
- Van Os, M., De Jong, N. H., & Bosker, H. R. (2020). Fluency in dialogue: Turn-taking behavior shapes perceived fluency in native and non-native speech. *Language Learning*. Advance online publication. <http://dx.doi.org/10.1111/lang.12416>.
- Waring, H. Z. (2019). Developing interactional competence with limited linguistic resources. In R. M. Salaberry & S. Kunitz (Eds.), *Teaching and testing L2 interactional competence: Bridging theory and practice* (pp. 215–227). New York: Routledge.
- Witton-Davies, G. (2014). *The study of fluency and its development in monologue and dialogue* (Unpublished doctoral dissertation). Lancaster University, Lancaster.
- Wolk, C., Götz, S., & Jäschke, K. (2020). Possibilities and drawbacks of using an online application for semi-automatic corpus analysis to investigate discourse markers and alternative fluency variables. *Corpus Pragmatics*. Advance online publication. <http://dx.doi.org/10.1007/s41701-019-00072-x>.
- Wood, D. (2006). Uses and functions of formulaic sequences in second language speech: An exploration of the foundations of fluency. *The Canadian Modern Language Review*, 63, 13–33. <http://dx.doi.org/10.1353/cml.2006.0051>.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Wright, C., & Tavakoli, P. (2016). New directions and developments in defining, analyzing and measuring L2 speech fluency. *International Review of Applied Linguistics in Language Teaching*, 54, 73–77. <http://dx.doi.org/10.1515/iral-2016-9990>.
- Yuan, J., Liberman, M., & Cieri, C. (2006). Towards an integrated understanding of speaking rate in conversation. In *Proceedings of Interspeech 2006: The ninth international conference on spoken language processing* (pp. 541–544). Pittsburgh, PA: ISCA. Retrieved from https://www.isca-speech.org/archive/interspeech_2006/i06_1795.html.

Appendices

Appendix 1: Monologue picture prompts



Picture prompt for the L1 Finnish monologue task.



Picture prompt for the L2 English monologue task.

Appendix 2: Dialogue task instructions

Note that the instructions have been translated from the Finnish original instructions by the researcher.

Instructions for the pair task

You and your pair are going to complete a problem-solving task.

Your task is to reach a conclusion together by discussing in English.

The more specific instructions are provided below.

This is not a test, the task will not be graded and there are no right or wrong answers.

Try to come up with as many things to talk about as possible.

You have about **2 minutes** to prepare for the task independently.

Then you will have **6 minutes** for the discussion with your partner.

STRANDED ON A DESERT ISLAND

You and your pair have been stranded on a desert island in the Pacific. All you have are the clothes that you are wearing. There is a fresh water spring, banana trees and coconut palms on the island.

The pictures show 16 items you may find useful for survival on the island. Your pair has the same pictures.

Your task is to organize all items in the order of usefulness. During the discussion, you should reach an agreement on the order of importance for all items.

Describe the items, discuss them and justify the order of importance.

You can now start preparing for the task by familiarizing yourself with the pictures.



Appendix 3: Rating sheet

The instructions were translated by the author from the Finnish originals. Only the rating sheet for sample 1 is presented here; the sheets for other samples (2–6) were identical.

Instructions: Rate the six dialogue samples numerically according to the four criteria. Comment on your ratings for fluency, interactional fluency, and strategic competence in the boxes below the numeric scales. For oral proficiency, the numeric rating is enough.

Sample 1 (01-G-004-005)

I) Oral proficiency

Rate the oral proficiency of the speaker on the left on a scale from 1 to 9 (1 = extremely weak oral proficiency, 9 = extremely good oral proficiency). Mark the number in bold.

1 2 3 4 5 6 7 8 9

Rate the oral proficiency of the speaker on the right on a scale from 1 to 9 (1 = extremely weak oral proficiency, 9 = extremely good oral proficiency). Mark the number in bold.

1 2 3 4 5 6 7 8 9

II) Fluency

When rating fluency, pay attention to the temporal aspects of speech: speech rate, pausing (silent pauses and filled pauses, e.g. *um*) and repair (for instance reformulations, repetitions). Accuracy and complexity should not be a part of fluency ratings.

Rate the fluency of the speaker on the left on a scale from 1 to 9 (1 = extremely disfluent, 9 = extremely fluent). Mark the number in bold.

1 2 3 4 5 6 7 8 9

Rate the fluency of the speaker on the right on a scale from 1 to 9 (1 = extremely disfluent, 9 = extremely fluent). Mark the number in bold.

1 2 3 4 5 6 7 8 9

Which aspects did you pay attention to when rating the sample?
Comment on and justify the ratings in the box below (note that you can enlarge the box if needed). The ratings should be based on concrete features of speech.

III) Interactional fluency

Rate the pair's interactional fluency on a scale from 1 to 9 (1 = extremely disfluent, 9 = extremely fluent). Mark the number in bold.

1 2 3 4 5 6 7 8 9

Which aspects did you pay attention to when rating the sample?
Comment on and justify the rating in the box below. The rating should be based on concrete features of speech.



IV) Strategic competence

When rating strategic competence, pay particular attention to the use of communication strategies (for instance paraphrases).

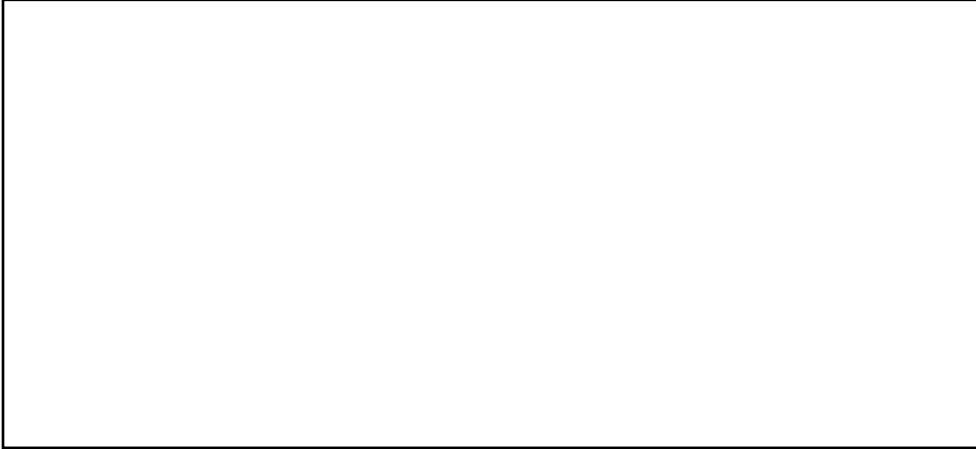
Rate the strategic competence of the speaker on the left on a scale from 1 to 9 (1 = extremely weak strategic competence, 9 = extremely good strategic competence). Mark the number in bold.

1 2 3 4 5 6 7 8 9

Rate the strategic competence of the speaker on the right on a scale from 1 to 9 (1 = extremely weak strategic competence, 9 = extremely good strategic competence). Mark the number in bold.

1 2 3 4 5 6 7 8 9

Which aspects did you pay attention to when rating the sample?
Comment on and justify the ratings in the box below. The ratings
should be based on concrete features of speech.





**UNIVERSITY
OF TURKU**

ISBN 978-951-29-8136-6 (PRINT)
ISBN 978-951-29-8137-3 (PDF)
ISSN 0082-6987 (Print)
ISSN 2343-3191 (Online)