Nordand

Universitetsforlaget

VITENSKAPELIG PUBLIKASJON

Årgang 19, nr. 2-2024, s. 73–90
ISSN online: 2535-3381
DOI: https://doi.org/10.18261/nordand.19.2.1

How Does Access to L1 Speech Affect L2 and L3 Perceived Fluency?

Evidence from Finnish-Speaking and Finnish–Swedish Bilingual Speakers

Elina Lehtilä
Doctoral Researcher, Department of English, University of Turku
anelleh@utu.fi
https://orcid.org/0000-0002-8236-7978

Pauliina Peltonen
Academy Research Fellow, Department of English, University of Turku
paupelt@utu.fi
https://orcid.org/0000-0002-8743-0032

Pekka Lintunen
Professor, Department of English, University of Turku
pekka.lintunen@utu.fi
https://orcid.org/0000-0003-3448-9310

Abstract
While first language (L1) fluency and cross-linguistic influences have emerged as factors that contribute to second language (L2) speech fluency, the potential effects of L1 speaking style and cross-linguistic differences on target language (L2, L3…) perceived fluency remain largely unexplored. This study reports the results from two experiments that examined how access to speakers' L1 Finnish speech influences the perceived fluency of L2 English (Experiment I) and L3 Swedish (Experiment II). L2 English and L3 Swedish speech elicited from the same L1 Finnish speakers (n = 20) were used in both experiments, and Experiment I also included L2 English speech samples from 10 Finnish–Swedish bilingual speakers to enable comparisons across different L1 backgrounds. Altogether 72 listeners participated in the two experiments and rated the speakers' L2/L3 fluency either based solely on L2/L3 speech (control condition), or after listening to the speakers' L1 Finnish speech (experimental condition). The results of Experiment I showed no significant effect of L1 background on perceived L2 English fluency, but the listeners with access to the speakers' L1 speech assigned higher speed and repair fluency ratings for both speaker groups. In contrast, in Experiment II, the same L1 Finnish speakers received significantly lower ratings for all aspects of their L3 Swedish fluency by the listeners who had access to their L1 speech. The implications of the findings for fluency research, teaching, and assessment are discussed.

Keywords
perceived fluency, multilingualism, second language acquisition, individual differences

Introduction

Speech fluency is an essential part of second language (L2) proficiency and assessment. In second language acquisition (SLA) research, fluency is usually approached from one of two perspectives: in the *broad* sense (i.e., higher-order fluency), fluency is often used as a synonym for overall oral proficiency, while in the *narrow* sense (i.e., lower-order fluency), fluency refers to the temporal aspects of speech (Lennon, 1990, 2000). In the present study, we approach fluency from the narrow sense and regard fluency as one of the three main components of (L2) oral proficiency within the Complexity-Accuracy-Fluency (CAF) framework. Research indicates that the three CAF components both uniquely contribute to and interact with one another in L2 proficiency and performance, and that the components and their interrelationships are dynamic in nature, often evolving in unpredictable ways over time (Norris & Ortega, 2009; for an overview of the framework, see, e.g., Housen et al., 2012).

Following Segalowitz's (2010) threefold framework of fluency, the present study focuses on *perceived fluency*, referring to the listener's interpretation of the speaker's fluency based on their spoken performance (typically studied by means of numeric fluency ratings from listeners), as opposed to *cognitive* (i.e., the psycholinguistic processes underlying speech production) or *utterance fluency* (i.e., temporal fluency measures based on speech data, such as speech rate and mean length of run). However, perceived fluency is closely connected with both utterance and cognitive fluency, since listeners' perceptions of the temporal aspects of speech are considered to reflect their interpretation of the ease and efficiency of the speaker's speech production processes (Lennon, 1990; Segalowitz, 2010). Furthermore, we consider the fluency features that affect listeners' perceptions of fluency to reflect the three aspects of *speed* (features related to speech tempo), *breakdown* (features related to pauses), and *repair fluency* (features related to corrections and hesitations; e.g., Skehan, 2009).

Recent fluency studies have provided increasing evidence that speakers' first (L1) and target language (L2, L3…) utterance fluency are connected (e.g., De Jong et al., 2015; Duran-Karaoz & Tavakoli, 2020; Peltonen & Lintunen, 2022). Some studies also indicate that the strength of the L1–L2 connections is potentially mediated by language proficiency (Huensch & Tracy-Ventura, 2017; Peltonen, 2018). Due to this overlap in L1 and L2 fluency patterns, L1 fluency should be acknowledged in L2 fluency assessment contexts (De Jong, 2018; Gao & Sun, 2024). Although the connections between L2 utterance fluency measures and perceived fluency ratings have been widely studied (see, e.g., Suzuki et al., 2021), no empirical studies to date have examined how information about the characteristics of the speaker's L1 speech (i.e., speaking style) might influence their perceived L2 fluency. In addition, previous utterance fluency studies involving cross-linguistic comparisons have indicated that the typological closeness between the L1 and L2 might affect both L2 utterance fluency (e.g., Lehtonen, 1979; Peltonen & Lintunen, 2016) and the strength of the connections between the speakers' L1 and L2 utterance fluency (e.g., Huensch & Tracy-Ventura, 2017). However, cross-linguistic differences have received little attention in perceived fluency research (for exceptions, see Derwing et al., 2009; Derwing & Munro, 2013). Furthermore, most perceived fluency studies have focused on major Indo-European languages such as English (Kormos & Dénes, 2004), French (Préfontaine et al., 2016), or Dutch (Bosker et al., 2013), and research on spoken L2/L3 Swedish from a perceived fluency perspective has, thus far, been scarce (but see Kallio et al., 2023).

Considering the context of the present study, Finland offers an ideal setting for the investigation of cross-linguistic differences. Finnish and Swedish are the two official languages in Finland, and both Finnish- and Swedish-language schools follow the national

core curriculum for basic education. Thus, differences in L2 English between Finnish- and Swedish-speaking Finns are not likely to arise from differences in their educational or cultural backgrounds. Furthermore, due to the status of both Swedish and Finnish as the official languages in Finland, all L1 Finnish and L1 Swedish speakers are required to study the other official language at school in addition to at least one additional foreign language. In Finnish basic education, instruction of the first compulsory language (A1) currently begins in grade 1 (in grade 3 until 2020), while the other compulsory language (B1) is introduced in grade 6 at the latest, and the vast majority of Finnish-speaking pupils choose English as their A1 and Swedish as their B1 language (Finnish National Agency for Education, 2019).

Our study aims to fill the gaps regarding L1–L2 connections and cross-linguistic differences in perceived fluency research by examining how access to speakers' L1 Finnish speech influences the perceived fluency of L2 English and L3 Swedish at different proficiency levels. The study employed a methodologically novel research design where the listeners (n = 72) rated the L2/L3 samples either after listening to the speakers' L1 Finnish speech (experimental condition), or without hearing the speakers' L1 speech (control condition). The statistical analyses focused on whether the listeners' fluency ratings of L2 English and L3 Swedish were affected by the listeners' access to the speakers' L1 Finnish speech (effect of listening condition). We also included the speakers' L1 (Finnish vs. Finnish–Swedish bilingual) as a between-subject variable to examine the potential effect of the speakers' L1 background on their perceived L2 English fluency. By exploring how listeners' access to the speakers' L1 speech and cross-linguistic differences affect L2/L3 perceived fluency, the present study provides a new perspective to the study of connections between L1 and L2/L3 fluency and has implications for L2 fluency research, assessment, and teaching.

Theoretical Background

To contextualize the present study, the following subsections provide a comprehensive overview of the relationship between L1 and L2 and cross-linguistic differences from the perspectives of speech fluency and the languages under investigation. The first subsection reviews previous relevant research on perceived fluency, with a focus on L1–L2 connections and cross-linguistic comparisons, and the second subsection summarizes research findings on the L1–L2 utterance fluency relationship. The final subsection discusses research on cross-linguistic influences and fluency across the first and target languages in the present study, i.e., Finnish, Swedish, and English.

Previous Studies on Perceived Fluency

Considering Lennon's (1990, 2000) narrow and broad senses of fluency, perceived fluency studies have usually focused either on lower-order fluency by instructing the listeners to base their ratings on specific temporal features of speech (e.g., Bosker et al., 2013; Derwing & Munro, 2013) or on listeners' intuitive ratings of fluency by not providing detailed definitions of fluency (typically corresponding with higher-order fluency; e.g., Kormos & Dénes, 2004; Suzuki & Kormos, 2020). While the latter approach is suitable for examining listeners' intuitive and potentially varied notions of fluency, the former approach was more appropriate for the present study to ensure that all listeners had a common, narrow understanding of fluency and based their ratings on similar criteria.

As discussed in the Introduction, both cross-linguistic differences and L1–L2 connections have received little attention in prior perceived fluency studies. A notable exception is Derwing et al.'s (2009) longitudinal study on the relationship between L1 and L2 fluency

development in 16 Slavic- (Russian and Ukrainian) and 16 Mandarin-speaking learners of English. In the study, speech samples in the speakers' L1 and L2 English were collected at four occasions over the course of two years. The L1 samples from the first data collection time were rated for fluency by linguistically trained L1 speakers of Mandarin (n = 8) or Russian (n = 8), and the L2 English samples from the three subsequent time points were rated by eight L1 speakers of English. Significant positive correlations were found between both L1 and L2 fluency ratings and L1 and L2 utterance fluency measures for both groups at the first time point, but there were more significant and stronger correlations for the Slavic speakers. The results suggest that the L1–L2 fluency relationship may vary across L1 backgrounds and shift over time in terms of both perceived and utterance fluency.

In another study that considered L1–L2 connections in perceived fluency, Pinget et al. (2014) examined how residuals (i.e., L2-specific measures corrected for L1 fluency behavior; Segalowitz, 2010) correlated with L2 fluency ratings and whether they were better predictors of fluency ratings than traditional L2 utterance fluency measures. The speakers were 15 English and 15 Turkish learners of Dutch, whose L2 productions were rated for fluency by 20 untrained L1 Dutch speakers. The results showed that while the residuals successfully predicted a large percentage of perceived fluency variance, they were not better predictors than the traditional measures. However, the listeners were not presented with the speakers' L1 speech, which could explain why adjusting the L2 fluency measures for L1 fluency did not result in better predictions of L2 fluency ratings.

Connections Between L1 and L2 Utterance Fluency

Although previous L2 perceived fluency studies have rarely involved cross-linguistic comparisons or the speakers' L1 fluency profiles, both L1–L2 connections and cross-linguistic influences have received more attention in utterance fluency research. In their seminal study on the connections between L1 and L2 utterance fluency, De Jong et al. (2015) examined L2 Dutch fluency across speaker groups with L1s typologically close to (English, n = 29) and typologically distant from (Turkish, n = 24) the L2. The results revealed moderate to strong, significant positive correlations between L1 and L2 fluency measures when the L1 groups were combined. Regression analyses also showed that all L2 fluency measures could be predicted from L1 measures to varying extents, and the inclusion of language group (English vs. Turkish) as a predictor variable significantly improved the regression model for some fluency measures. However, the interaction between language group and L1 fluency measures did not significantly improve the models, indicating that the nature of the connections between L1 and L2 fluency was the same regardless of the typological distance between the L1 and L2.

Using similar research design and methods, Huensch and Tracy-Ventura (2017) studied L1–L2 fluency relationships among English-speaking university students of Spanish (n = 24) and French (n = 25) before and after studying abroad. Regression analyses revealed that the relative predictive power of L1 fluency, target language, and proficiency for L2 fluency changed during study abroad, indicating that variation in L2 utterance fluency is affected by L1 fluency behavior, cross-linguistic differences, and proficiency, but how these factors contribute to L2 fluency may change over time. However, it should be noted that the role of L2 proficiency in the connections between L1 and L2 utterance fluency is somewhat unclear, with some findings indicating that the strength of L1–L2 correlations increases with developing L2 proficiency (Peltonen, 2018), while others have found no mediating effect of proficiency level on the relationship between L1 and L2 fluency measures (Duran-Karaoz & Tavakoli, 2020).

Due to the growing number of studies that have shown connections across speakers' L1 and L2 utterance fluency, Gao and Sun (2024) conducted an analysis of the aggregated effect sizes across 16 studies comparing L1 and L2 utterance fluency from the same speakers. In the meta-analysis, pausing measures were found to display the strongest correlations between the L1 and L2, followed by speed (articulation rate) and composite (speech rate; a measure that combines elements from both speed and breakdown fluency dimensions) fluency measures, while the weakest correlations were observed for repair fluency measures. Although the results suggest that the overall association between L1 and L2 fluency is strongest in pausing and weakest in repair behavior, the extent to which L2 proficiency or the typological distance between the L1 and L2 affects this association remains unclear, as neither L2 proficiency level nor the speakers' L1/L2 were included as moderators in the meta-analysis.

Cross-Linguistic Effects and Speech Fluency Across Finnish, Swedish, and English

Regarding the speakers in the present study, English is typologically similar to Swedish but distant from Finnish, which is why the Swedish-speaking Finns have been hypothesized to have an advantage over Finnish-speaking Finns in learning English (Ringbom, 2007). This hypothesis has received support from national learning outcome evaluations conducted by the Finnish Education Evaluation Centre, which have shown that grade 9 pupils in Swedish-speaking schools exhibit higher English language proficiency than their Finnish-speaking peers (see, e.g., Härmälä & Marjanen, 2022). From the perspective of utterance fluency, Finnish- and Swedish-speaking Finns' L2 English speech was first systematically examined by Lehtonen (1979), who found that Finnish-speaking college students (n = 24) were less fluent in English than Swedish-speaking students in Finland (n = 20) and Sweden (n = 37), with slower speech and articulation rates and a higher percentage of silent pauses. However, according to Ringbom (2007), the differences between Finnish- and Swedish-speaking Finns in their English proficiency are likely to be more prominent at lower educational levels and decrease by university level. Some recent utterance fluency studies seem to support this hypothesis: for instance, Peltonen and Lintunen (2016) found that Swedish-speaking upper secondary school students in Finland spoke English more fluently than their Finnish-speaking peers, but the differences in L2 English fluency between Swedish- and Finnish-speaking university students were minor and non-significant.

Although Finnish and Swedish speakers' perceived English fluency has not been compared previously, Tergujeff (2021) showed that Swedish-speaking upper secondary school students in Finland (n = 30) were perceived to be significantly less accented and more comprehensible in English than the Finnish-speaking students (n = 30) at the same proficiency levels (Common European Framework of Reference for Languages [CEFR] levels B1 and B2; see Council of Europe, 2001). The results demonstrate that even when Finnish- and Swedish-speaking Finns are matched for English proficiency, listeners may rate the two groups differently in terms of different aspects of spoken production, at least at intermediate and upper intermediate proficiency levels. However, to our knowledge, the present study is the first to explore Finnish- and Swedish-speaking Finns' L2 English perceived fluency at an advanced level of proficiency.

Compared to English and Finnish, Swedish has rarely been the focus of fluency research. A noteworthy exception is Kallio et al.'s (2023) study investigating the prosody and fluency of Finland Swedish across native (n = 30) and non-native (L1 Finnish; n = 186) speakers at different proficiency levels. Their results showed that fluency measures differentiated

between both native and non-native speakers of Swedish and non-native speakers at CEFR levels A1/A2 and levels B1/B2. Among these measures, speech rate was the most significant predictor for perceived Swedish fluency. Their findings indicate that fluency measures play a significant role in perceived Swedish fluency, corroborating the language-independence of fluency features and their importance in language teaching and assessment.

In a previous utterance fluency study, which included the same speakers as the present study, Peltonen and Lintunen (2022) compared Finnish, Swedish, and English fluency among 20 L1 Finnish and 10 Finnish–Swedish bilingual university students. The results showed significant, mostly large differences between the groups in their Swedish fluency, but no significant differences between the groups' fluency in Finnish and English, further supporting Ringbom's (2007) hypothesis that the initial gap between Swedish- and Finnish-speaking Finns in their English proficiency gradually narrows and closes by university level. In addition, moderate to strong correlations were found between the fluency measures across all language pairs for the bilingual speakers, but mostly between Finnish and English for the L1 Finnish speakers. Due to the L1 Finnish speakers' higher proficiency level in their L2 English compared to their L3 Swedish, the results were considered to corroborate previous research findings suggesting that connections between L1 and L2 utterance fluency become stronger with developing proficiency (Huensch & Tracy-Ventura, 2017; Peltonen, 2018).

Material and Methods

Research Questions and Hypotheses

Based on gaps in research regarding the potential effect of L1 speaking style and cross-linguistic influences on perceived L2 fluency, we examined how access to L1 Finnish speech affects perceived L2 English and L3 Swedish fluency and whether the speakers' L1 background (Finnish vs. Finnish–Swedish bilingual) influences their perceived L2 English fluency. In line with these objectives, two experiments were conducted to answer the following research question:

RQ1. How is perceived L2 English fluency affected by access to L1 Finnish speech and L1 background among advanced Finnish-speaking (Group 1, G1) and Finnish–Swedish bilingual (Group 2, G2) learners of English? (Experiment I).

RQ2. How is perceived L3 Swedish fluency affected by access to L1 Finnish speech among pre-intermediate Finnish-speaking (G1) learners of Swedish? (Experiment II).

The speech samples used in Experiment I were produced by advanced speakers of L2 English, who were either Finnish L1 speakers (G1; $n = 20$) or Finnish–Swedish bilinguals (G2; $n = 10$). The L3 Swedish speech samples used in Experiment II came from the same L1 Finnish speakers ($n = 20$) with a lower proficiency level in Swedish compared to English. In the present study, 72 linguistically trained listeners with expertise in the target language rated the speakers' speed, breakdown, repair, and overall fluency in L2 English (Experiment I; $n = 30$) and L3 Swedish (Experiment II; $n = 42$). Based on our previous study on the two groups' utterance fluency (Peltonen & Lintunen, 2022; see previous section), for RQ1, we hypothesized no major differences in perceived L2 English fluency between the two speaker groups. As strong connections were found across the two groups' L1 Finnish and L2 English fluency measures, we hypothesized that access to the speakers' L1 speech would influence perceived L2 fluency for both speaker groups. Due to the weaker correlations between the speakers' L1 Finnish and L3 Swedish fluency measures, we hypothesized that access to the

speakers' L1 Finnish speech would have a smaller effect on perceived fluency in L3 Swedish compared to L2 English. However, as a similar research design involving both L1 and L2/L3 speech has not, to our knowledge, been previously implemented in fluency research, we did not have more specific hypotheses regarding the influence of L1 speaking style on perceived L2/L3 fluency.

Speech Samples

The speech samples used in the present study were selected from a larger data set collected in the project "Fluency across Multilingual Speakers" (*MultiFluency*; funded by the Swedish Cultural Foundation in Finland) based on our previous research on these speakers' utterance fluency in Finnish, Swedish, and English (Peltonen & Lintunen, 2022). All speakers participated in the project voluntarily and signed an informed consent form before participating. The speaker groups represented two L1 backgrounds: Group 1 (G1) included 20 Finnish-speaking Finns (17 females, 2 males, 1 other) and Group 2 (G2) included 10 Finnish–Swedish bilingual Finns (8 females, 2 males; for more information regarding the group division based on the participants' language background, see Peltonen & Lintunen, 2022, p. 52). Most speakers studied English either as their major (*nG1* = 16; *nG2* = 5) or minor subject (*nG1* = 4; *nG2* = 1), while four speakers in G2 studied other language subjects at university. Three target language experts and one advanced student of the target languages working as a research assistant assessed the speakers' L2 English and L3 Swedish speech using the CEFR descriptors (Council of Europe, 2001, pp. 28–29). The oral proficiency assessments indicated that both groups represented level C1 on average in spoken English, while G1 represented level A2 on average in spoken Swedish, in accordance with the speakers' self-assessments (see Table 1).

Table 1. Speakers' Background Characteristics by Speaker Group (Finnish vs. Finnish–Swedish Bilingual)

	Finnish speakers (Group 1, G1; *n* = 20)		**Finnish–Swedish bilinguals (Group 2, G2; *n* = 10)**	
Background variable	***M***	***SD***	***M***	***SD***
Age (years)	20.60	2.04	27.00	8.38
Duration of English studies at school (years)	10.18	1.09	10.33	2.40
Self-rated English spoken production (1–5[a])	4.00	0.65	4.10	0.74
Duration of Swedish studies at school (years)	6.63	1.56		
Self-rated Swedish spoken production (1–5[a])	1.85	1.04		

[a]Scale: 1 = weak, 2 = moderate, 3 = good, 4 = very good, 5 = excellent.

In the speaking tasks, all speakers were asked to tell a short story in Finnish, Swedish, and English based on a cartoon strip (see, e.g., Derwing & Munro, 2013). Three comparable cartoon strips were used (see Peltonen & Lintunen, 2022): each cartoon strip featured a linear storyline depicted in six frames and consisted only of pictures. The order of the cartoons and languages was counterbalanced across participants, with the exception that the first task was always done in one of the speakers' L1s (Finnish for G1, Finnish or Swedish for G2). The speakers had two minutes of planning time before each task and they could look at the pictures while talking, but they were not allowed to take notes.

To standardize the volume across the speech samples for the rating tasks, all samples were resampled to a sampling frequency of 44100 Hz and scaled to an intensity of

70 dB (see, e.g., Pinget et al., 2014). To increase the ecological validity of the ratings, the entire L2/L3 samples were presented in both listening conditions of both experiments (see, e.g., Préfontaine et al., 2016; Suzuki & Kormos, 2020). The mean duration of the L2 English samples was 64.60 s (*SD* = 29.34) and the mean duration of the L3 Swedish samples was 56.75 s (*SD* = 19.02). For the experimental condition, short excerpts of approximately 20 s were extracted from the beginning of the original L1 Finnish recordings, excluding any initial disfluencies and ending at a natural break in speech. Excerpts rather than entire L1 speech samples were used to minimize the potential effect of listener fatigue. Furthermore, as the L1 sample was only used as reference for the L2/L3 sample, a shorter speech sample was deemed sufficient to get an impression of the speaker's personal speaking style.

Listeners

A variety of types of listeners have been used in previous perceived fluency studies, ranging from experienced raters, such as teachers or linguists (e.g., Préfontaine et al., 2016), to untrained L1 speakers of the target language (e.g., Bosker et al., 2013; Pinget et al., 2014) and proficient L2 learners (e.g., Derwing & Munro, 2013; Rossiter, 2009). To ensure that all listeners were both highly familiar with the speakers' L1 (Finnish) and proficient in the L2/L3 (English in Experiment I, Swedish in Experiment II), linguistically trained L1 or L2 speakers of the target languages (see also Derwing et al., 2009) were chosen as listeners in the present study and recruited among students of language subjects from six Finnish universities. In total, 72 university students participated in the two experiments either as a part of an MA-level target language course at university (n = 39) or as volunteers (n = 33). The participants in both experiments were assigned to one of the two listening conditions (control vs. experimental). In the control condition, the listeners heard only the L2/L3 speech samples, while in the experimental condition, the listeners were also presented with the short excerpts from the participants' L1 Finnish speech samples. Participation in the study was voluntary and informed consent was obtained from all participants prior to the experiment.

In Experiment I, all listeners (n = 30) were advanced university students of English who reported speaking Finnish as their L1. All listeners in the control condition (n = 15; 14 females, 1 male) and all except one listener in the experimental condition (n = 15; 11 females, 3 males, 1 preferred not to say) studied English as a major subject. In Experiment II, all listeners (n = 42) had a background in linguistics and expertise in Swedish, either through studying Swedish at university (n = 35), speaking Swedish as their L1 (n = 3), or both (n = 4). Most listeners reported speaking Finnish as their L1 (n = 34; two reporting Swedish as an additional language spoken at home), while six listeners reported Swedish as their L1 (two listeners reporting Finnish as an additional language spoken at home). One listener reported both Finnish and Swedish as their L1s and one listener reported Russian as their L1 and Finnish and Swedish as additional languages spoken at home. The listeners in the experimental condition (n = 22; all female) studied Swedish at university either as a major (n = 16) or a minor (n = 6) subject, and the listeners in the control condition (n = 20; 14 female, 4 male, 2 other) studied Swedish either as a major (n = 9) or minor (n = 8) subject at university, with three L1 speakers of Swedish studying other language subjects besides Swedish at university. Most listeners in both experiments had participated in teacher training, and all listeners reported high familiarity with Finnish-accented L2/L3 speech (see Table 2). All listeners in both experiments reported normal hearing.

Table 2. Listeners' Background Characteristics by Experiment (I vs. II) and Listening Condition (Control vs. Experimental)

	Experiment I (L2 English)				Experiment II (L3 Swedish)			
	Control (*n* = 15)		Experimental (*n* = 15)		Control (*n* = 20)		Experimental (*n* = 22)	
Background variable	***n***	***%***	***n***	***%***	***n***	***%***	***n***	***%***
Teacher training	10	67	12	80	13	65	22	100
Language assessment experience	4	27	7	53	4	20	9	41
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age (years)	27.13	5.24	27.07	5.56	25.65	7.81	23.86	2.40
Duration of L2/L3 studies at school (years)	10.67	1.18	10.13	0.64	6.62	1.75	6.41	1.40
Duration of L2/L3 studies at university (years)	4.37	1.61	4.17	0.92	3.86	0.79	2.82	1.97
Finnish accent familiarity (1–6[a])	5.20	1.15	4.93	1.22	4.90	1.07	4.91	1.06

[a]Scale: 1 = not familiar at all, 6 = extremely familiar.

Rating Procedure

All rating data were collected online using the REDCap software hosted at the University of Turku (Harris et al., 2019). The task instructions were given in the target language (English in Experiment I, Swedish in Experiment II), but otherwise the rating procedure was identical in both experiments. Both versions of the rating task in both languages were piloted before data collection. All listeners completed the rating task remotely and unsupervised on their own computers at a time most convenient to them. To gain a comprehensive picture of L2/L3 perceived fluency, both analytic ratings of the different aspects of fluency and a holistic rating of overall fluency were included in the task. The listeners were instructed to first rate the L2/L3 speech samples on 9-point scales (see, e.g., Bosker et al., 2013; Kahng, 2018; Suzuki & Kormos, 2020) for speed, breakdown, and repair fluency based on the speed of speech (1 = very slow, 9 = very fast), pauses (1 = very many and/or very long disruptive pauses, 9 = very few and/or very short disruptive pauses), and corrections (1 = very many and/or very long corrections, 9 = very few and/or very short corrections). They were then asked to rate the overall fluency of the speaker based on their ratings of speed, breakdown, and repair fluency (1 = extremely disfluent, 9 = extremely fluent). The participants were instructed to listen to each speech sample only once before rating it, but they could spend as much time on rating each sample as they wanted. To avoid familiarity bias, the participants were shown the cartoon strips used in the speaking tasks prior to the rating task (see, e.g., Kahng, 2018; Rossiter, 2009; Suzuki & Kormos, 2020). Following the instructions, the participants completed two practice trials to familiarize themselves with the rating procedure. After the practice trials, the L2/L3 speech samples were presented in one of three pseudo-randomized orders. To avoid listener fatigue, the participants were instructed to take a short break halfway through the experiment.

In both experiments, all listeners rated all target language samples. In the experimental condition of both experiments, the rating procedure was identical to that for the control condition, but the listeners were first presented with the excerpts from the participants' L1 Finnish speech samples. The listeners were instructed to listen to the L1 sample before listening to the L2/L3 sample, but the instructions specified that they should not rate the L1 sample. The listeners were told that the L1 speech sample was presented in order for them to get an impression of the speakers' personal speaking style in their L1, but no further

instructions were provided to ensure that the listeners utilized the L1 sample intuitively when rating the target language sample.

Analyses

The statistical analyses were performed using R 4.0.4 (R Core Team, 2021). First, to examine interrater reliability, Cronbach's alpha coefficients were calculated separately by experiment and listening condition via the irr R package (v0.84.1; Gamer et al., 2019). The Cronbach's alpha values were well above acceptable (>.83; Plonsky & Derrick, 2016) for all the rated constructs (see Appendix A), indicating that the level of interrater agreement was sufficient and comparable across conditions in both experiments. The fluency scores were then averaged for each speaker across the listeners in both conditions and subjected to further statistical analyses using the rstatix R package (v0.7.2; Kassambara, 2023).

Prior to conducting the analyses, the data were examined to confirm that the assumptions for parametric tests were sufficiently met in both experiments (through visual examination of the QQ-plots, Shapiro-Wilk normality tests, and the additional Levene's tests in Experiment I; see Appendix B). In Experiment I (RQ1), the mean ratings for each speaker were submitted to a series of mixed 2 × 2 between/within analyses of variance (ANOVAs) with the speakers' L1 (Finnish vs. Finnish–Swedish bilingual) as the between-subject variable and listening condition (control vs. experimental) as the within-subject variable. Both eta-squared (η^2) and partial eta-squared (η^2_p) were used as estimates of effect sizes (Norouzian & Plonsky, 2018). In Experiment II (RQ2), paired *t*-tests comparing the ratings in the control and experimental conditions were performed to determine whether access to the speakers' L1 Finnish speech affected their perceived L3 Swedish fluency. Cohen's *d* effect sizes were calculated for each rated variable and interpreted according to Plonsky and Oswald's (2014) guidelines for L2 research (d = .60 as small, d = 1.00 as medium, and d = 1.40 as large). Following the recommendations of Larson-Hall (2015, p. 287), the False Discovery Rate (FDR) method was used to control for potential type I errors resulting from multiple comparisons in both experiments, as the FDR offers more power over family-wise error rate (FWER) controlling procedures (albeit with an increased risk of false positives compared to the more conservative FWER procedures).

Results

Results of Experiment I

Before presenting the results of the statistical analyses, the means and standard deviations of the average speed, breakdown, repair, and overall L2 English fluency ratings and their difference values by listening condition in both speaker groups are provided in Table 3.

As can be seen from Table 3, both speaker groups received the lowest mean ratings for breakdown fluency and the highest mean ratings for repair fluency in both listening conditions. On average, the fluency ratings were very similar across the Finnish speakers (G1) and the Finnish–Swedish bilinguals (G2) for all the rated constructs in both listening conditions. Moreover, both speaker groups received higher mean ratings for speed, breakdown, and repair fluency from the listeners in the experimental condition compared to the control condition, but slightly lower ratings for overall fluency. Overall, however, the descriptive statistics did not point to any major differences in the mean fluency ratings across the speaker groups or listening conditions. The results of the mixed between/within ANOVAs for all the rated variables are shown in Table 4, with accompanying effect sizes (η^2 and η^2_p).

Table 3. Summary Statistics for L2 English Fluency Ratings (Min. 1 – Max. 9) by Speaker Group (Finnish vs. Finnish–Swedish Bilingual) and Listening Condition (Control vs. Experimental)

	Finnish speakers (Group 1, G1)			Finnish–Swedish bilinguals (Group 2, G2)		
	Control	Experimental	Difference	Control	Experimental	Difference
Rated variable	*M* (*SD*)	*M* (*SD*)	*M*[a] (*SD*)	*M* (*SD*)	*M* (*SD*)	*M*[a] (*SD*)
Speed fluency	6.24 (0.72)	6.42 (0.68)	0.18 (0.23)	6.33 (0.83)	6.55 (0.84)	0.21 (0.23)
Breakdown fluency	6.05 (1.13)	6.16 (1.14)	0.11 (0.37)	6.06 (0.88)	6.29 (0.81)	0.23 (0.37)
Repair fluency	6.92 (0.86)	7.45 (0.74)	0.53 (0.32)	7.17 (0.74)	7.71 (0.95)	0.54 (0.33)
Overall fluency	6.80 (0.74)	6.67 (0.72)	−0.13 (0.28)	6.94 (0.74)	6.84 (0.85)	−0.10 (0.27)

[a]Slight discrepancies in mean difference values are due to rounding.

Table 4. Results of Mixed ANOVAs with the Speakers' L1 (Finnish vs. Finnish–Swedish Bilingual) and Listening Condition (Control vs. Experimental) as Factors

	Speaker L1				Listening Condition					Interaction			
df (1, 28)	*F*	*p*	η^2	η^2_p	*F*	*p*	p_{FDR}	η^2	η^2_p	*F*	*p*	η^2	η^2_p
Speed fluency	0.15	.700	.005	.005	**19.47**	**<.001**	**<.001**	**.016**	**.410**	0.11	.741	<.001	.004
Breakdown fluency	0.03	.862	.001	.001	5.55	.026	.104	.006	.165	0.62	.439	<.001	.022
Repair fluency	0.67	.421	.022	.023	**72.55**	**<.001**	**<.001**	**.093**	**.722**	0.00	.958	<.001	<.001
Overall fluency	0.30	.589	.010	.011	4.38	.045	.135	.005	.135	0.06	.807	<.001	.002

Note. Statistically significant differences (after the FDR corrections) marked in bold.

As shown in Table 4, the effect of the speakers' L1 background and the interaction between L1 and listening condition were non-significant for all the rated dimensions. For speed and repair fluency, a main effect of listening condition was found, showing that both speaker groups received significantly higher speed and repair fluency ratings from the listeners in the experimental condition (with access to the speakers' L1 speech) compared to the listeners in the control condition (no access to the speakers' L1 speech), with a stronger effect for repair fluency compared to speed fluency. After the FDR adjustments, the effect of listening condition for breakdown and overall fluency ratings fell below significance, indicating that access to L1 speech did not significantly influence perceived breakdown or overall fluency.

Results of Experiment II

The means and standard deviations of the average speed, breakdown, repair, and overall L3 Swedish fluency scores and their difference values by listening condition and the results of the paired *t*-tests are presented in Table 5.

Table 5. Results of the Paired t-tests Comparing L3 Swedish Fluency Ratings (Min. 1 – Max. 9) Between Listening Conditions (Control vs. Experimental)

	Control	Experimental	Difference				
	M	*M*	*M*[a]				
Rated variable	*SD*	*SD*	*(SD)*	*t*(19)	*p*	p_{FDR}	Cohen's *d*
Speed fluency	5.35	4.97	–0.39	**4.68**	**<.001**	**<.001**	**1.05**
	(1.18)	(0.99)	(0.37)				
Breakdown fluency	4.36	4.01	–0.36	**4.73**	**<.001**	**<.001**	**1.06**
	(1.55)	(1.50)	(0.34)				
Repair fluency	6.11	5.88	–0.23	**2.53**	**.020**	**.020**	**0.57**
	(0.92)	(0.86)	(0.40)				
Overall fluency	5.06	4.84	–0.22	**3.66**	**.002**	**.002**	**0.82**
	(1.35)	(1.33)	(0.27)				

Note. Statistically significant differences (after the FDR corrections) marked in bold.

[a]Slight discrepancies in mean difference values are due to rounding.

Similar to Experiment I, the mean scores for the rated variables showed that, on average, the speakers received the lowest ratings for perceived breakdown fluency and the highest ratings for perceived repair fluency in both listening conditions (see Table 5). However, in contrast to Experiment I, the results of the paired *t*-tests indicated that the speakers were perceived as significantly more fluent across all rated variables in the control condition (listeners who only heard the speakers' L3 Swedish speech samples) compared to the experimental condition, with effect sizes ranging from small for repair (d = 0.57) and overall fluency (d = 0.82) to moderate for speed (d = 1.05) and breakdown fluency (d = 1.06).

Discussion

The aim of our study was to examine how listeners' access to speakers' L1 Finnish speech influences perceived L2 English and L3 Swedish fluency across different proficiency levels and L1 backgrounds. In Experiment I, a series of mixed ANOVAs revealed a significant main effect of listening condition for L2 English speed and repair fluency, but no significant main effects of speakers' L1 background (Finnish vs. Finnish–Swedish bilingual) or interactions between L1 and listening condition for any of the rated variables. In other words, the fluency ratings did not differ between L1 Finnish and Finnish–Swedish bilingual speakers of English, and there was no difference between the two speaker groups in the way that access to L1 speaking style affected their fluency ratings. As the listeners with access to L1 speech assigned significantly higher speed and repair fluency ratings for both speaker groups compared to the listeners without access to L1 speech, the results indicate that the knowledge of a speaker's speaking style in their L1 has a positive effect on perceived L2 repair and speed fluency at advanced proficiency levels. However, in Experiment II, the results of the paired *t*-tests showed that the same L1 Finnish speakers whose L2 English fluency was rated in Experiment I received significantly lower ratings for all the rated variables in the experimental condition in L3 Swedish, in which their proficiency level was lower. These findings were somewhat in contrast to our hypotheses, as we expected L1 speech to have a greater effect on perceived fluency in L2 English than in L3 Swedish, due to the stronger connections in utterance fluency between the speakers' L1 and L2 compared to their L1 and L3 (Peltonen & Lintunen, 2022). A potential explanation for the findings is that, in the present study, the L1 speech samples may have set individual

standards for each speaker against which their L2/L3 productions were compared by the listeners in the experimental condition. Thus, in Experiment I, the similarities between the speakers' L1 Finnish and L2 English utterance fluency features led to higher repair and speed fluency ratings, whereas, in Experiment II, the greater distance between the speakers' L1 Finnish and L3 Swedish fluency led to lower fluency ratings compared to the listeners who only heard the L3 samples. To summarize, the results suggest that information about the speaker's L1 speech may affect perceived L2/L3 fluency differently depending on the strength of the associations between L1 and L2/L3 fluency features, which are potentially connected with the speaker's proficiency level in the L2/L3 (see also Derwing et al., 2009; Huensch & Tracy-Ventura, 2017; Peltonen, 2018).

Regarding the lack of statistically significant effects of L1 background (Finnish vs. Finnish–Swedish bilingual) on perceived L2 English fluency in Experiment I, the findings are in line with our hypotheses and corroborate the results of the previous utterance fluency analysis that demonstrated only minor differences between the two groups in their L2 English fluency (Peltonen & Lintunen, 2022; see also De Jong et al., 2015). However, since Tergujeff's (2021) study showed that Finnish- and Swedish-speaking Finns' L2 English speech was rated differently even at same levels of English proficiency, some differences between the two speaker groups' fluency ratings might have been expected in the present study as well. It is possible that the lack of cross-linguistic effects on perceived L2 English fluency in the present study was due to the participants' higher level of oral proficiency, thus adding further support for Ringbom's (2007) hypothesis that the differences between Finnish- and Swedish-speaking Finns in their L2 English proficiency decrease with developing proficiency. Additionally, if cross-linguistic differences between Finnish- and Swedish-speaking Finns in L2 English persist at advanced proficiency levels, they might be found in other aspects of oral proficiency, such as accentedness or comprehensibility (see, e.g., Tergujeff, 2021).

Our methodologically novel approach to perceived L2/L3 fluency has important implications for the assessment of fluency in both large-scale oral proficiency testing as well as classroom-based speaking tests, as the results suggest that listeners may perceive certain aspects of the speaker's L2/L3 fluency differently if they are familiar with the speaker's L1 speaking style. Especially in terms of repair, which is typically considered a sign of disfluency in L2/L3 testing and pedagogy, the results indicate the need for increasing language teachers' and practitioners' awareness of how individual differences in L1 speaking style may influence intuitive ratings of L2/L3 fluency. However, as the listeners in the present study were not instructed to compare the L1 and L2/L3 speech samples, in future, listeners could be explicitly guided toward rating L2/L3 fluency based on L1 fluency features to shed more light on whether L1 speech subconsciously affected the L2/L3 fluency ratings or whether the listeners knowingly compared some of the L1 and L2/L3 fluency features (see also Lehtilä et al., 2024). Future studies could also examine correlations between fluency ratings and fluency measures to further explore how listeners attend to L2/L3 fluency in different listening conditions.

Conclusion

The present study examined how listeners' access to L1 Finnish speech influences perceived L2 English and L3 Swedish fluency based on two experiments. In Experiment I, we also investigated whether the speakers' L1 background influences their perceived L2 English fluency. While previous studies have found connections between L1 and L2/L3 utterance fluency, this study was the first to incorporate speakers' L1 speech into L2/L3 fluency

ratings. Based on the results, cross-linguistic influences on perceived L2 English fluency appear to be minor for advanced L1 Finnish and Finnish–Swedish bilingual speakers of English. However, access to speakers' L1 Finnish speech led to higher ratings of English speed and repair fluency for both speaker groups, while access to L1 Finnish speech led to lower ratings of L3 Swedish fluency. This indicates that the influence of L1 speaking style on perceived L2/L3 fluency is potentially mediated by the speakers' proficiency level in the L2/L3. As the listeners in the present study were linguistically trained Finnish university students, future studies could apply similar rating procedures to more experienced listeners (e.g., language teachers) and listeners who are less familiar with the speakers' L1 to confirm these findings. In addition, the L1 and L2/L3 in the present study represented typologically distant languages (L1 Finnish vs. L2 English/L3 Swedish). In future, other first and target languages of varying typological distance could be examined. Furthermore, since we employed a between-subjects design where different listeners were assigned to different listening conditions (control vs. experimental), future studies could use within-subjects designs where the same listeners rate L2/L3 samples both with and without access to L1 speech to control for potential individual differences between listeners. Finally, it should be noted that the generalizability of the present study is limited due to the relatively small number of speakers and listeners, and the findings should be confirmed in future studies with larger sample sizes. Larger sample sizes could also help with counteracting the heightened probability for potential false negatives associated with multiple comparisons due to the inclusion of two L1 groups in Experiment I.

Despite these limitations, the present study is unique in its contribution to speech fluency research, as it is the first to explore how access to L1 speech influences perceived L2/L3 fluency. More specifically, by using speech samples from the same speakers to investigate how access to L1 Finnish speech affects the perceived fluency of both L2 English (Experiment I) and L3 Swedish (Experiment II), the study design enabled us to investigate these effects across the speakers' multiple languages at different proficiency levels. The novel approach to perceived L2/L3 fluency introduced in the present study also has implications beyond fluency teaching and assessment, as the findings raise the question whether listeners' familiarity with speakers' L1 speech might also affect the ratings of other aspects of target language oral proficiency. In future, the design of the present study could be extended to the analysis of other aspects of L2/L3 speech production to shed more light on the role of L1 speaking style in the overall assessment of L2/L3 speech. Overall, this study has provided valuable insights regarding the relationship between L1 and L2/L3 speech fluency from the listener's perspective, and the results suggest that acknowledging individual differences in L1 speech and their potential impact on listeners' perceptions of L2/L3 fluency can aid in the development of more reliable fluency assessment and teaching practices.

References

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

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.

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

De Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2015). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics, 36*(2), 223–243. https://doi.org/10.1017/S0142716413000210

Derwing, T. M., & Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: A 7-year study. *Language Learning, 63*(2), 163–185. https://doi.org/10.1111/lang.12000

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*(4), 533–557. https://doi.org/10.1017/S0272263109990015

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, 42*(4), 671–695. https://doi.org/10.1017/S0272263119000755

Finnish National Agency for Education. (2019). *What languages do pupils study in basic education?* (Facts express 1C/2019). https://www.oph.fi/sites/default/files/documents/factsexpress1c_2019.pdf

Gamer, M., Lemon, J., & Singh, I. F. P. (2019). *irr: Various coefficients of interrater reliability and agreement* (Version 0.84.1) [R package]. The Comprehensive R Archive Network. https://CRAN.R-project.org/package=irr

Gao, J., & Sun, P. P. (2024). How does learners' L2 utterance fluency relate to their L1? A meta-analysis. *International Journal of Applied Linguistics, 34*(1), 276–291. https://doi.org/10.1111/ijal.12493

Härmälä, M., & Marjanen, J. (2022). Englantia koronapandemian aikaan – A-englannin osaaminen 9. luokan lopussa keväällä 2021 [English during the coronavirus pandemic – English proficiency at the end of grade 9 in spring 2021]. Finnish Education Evaluation Centre. https://www.karvi.fi/sites/default/files/sites/default/files/documents/KARVI_2222.pdf

Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., McLeod, L., Delacqua, G., Delacqua, F., Kirby, J., Duda, S. N., & REDCap Consortium. (2019). The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics, 95*, Article 103208. https://doi.org/10.1016/j.jbi.2019.103208

Housen, A., Kuiken, F., & Vedder, I. (Eds.). (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. John Benjamins.

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

Kahng, J. (2018). The effect of pause location on perceived fluency. *Applied Psycholinguistics, 39*(3), 569–591. https://doi.org/10.1017/S0142716417000534

Kallio, H., Kautonen, M., & Kuronen, M. (2023). Prosody and fluency of Finland Swedish as a second language: Investigating global parameters for automated speaking assessment. *Speech Communication, 148*, 66–80. https://doi.org/10.1016/j.specom.2023.02.003

Kassambara, A. (2023). *rstatix: Pipe-friendly framework for basic statistical tests* (Version 0.7.2) [R package]. The Comprehensive R Archive Network. https://CRAN.R-project.org/package=rstatix

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

Larson-Hall, J. (2015). *A guide to doing statistics in second language research using SPSS and R* (2nd ed.). Routledge. https://doi.org/10.4324/9781315775661

Lehtilä, E., Peltonen, P., & Lintunen, P. (2024). Listener perceptions of L1 speech in L2 and L3 fluency assessment contexts: A qualitative approach. In M. Toivola, P. Lintunen, & L. M. Heikkola (Eds.), *Puheen tutkimuksen uusia suuntia – Aineistona vapaasti tuotettu puhe. New directions in speech research – Freely produced speech as data*. (Vol. 17, pp. 191–215). The Finnish Association for Applied Linguistics (AFinLA). https://doi.org/10.30660/afinla.136100

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). Åbo Akademi.

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

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

Norouzian, R., & Plonsky, L. (2018). Eta- and partial eta-squared in L2 research: A cautionary review and guide to more appropriate usage. *Second Language Research*, *34*(2), 257–271. https://doi.org/10.1177/0267658316684904

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

Peltonen, P. (2018). Exploring connections between first and second language fluency: A mixed methods approach. *The Modern Language Journal, 102*(4), 676–692. https://doi.org/10.1111/modl.12516

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*(2), 209–238. https://doi.org/10.1515/eujal-2014-0018

Peltonen, P., & Lintunen, P. (2022). Multilingual speakers' L1, L2, and L3 fluency across languages: A study of Finnish, Swedish, and English. *Nordand, 17*(1), 48–63. https://doi.org/10.18261/nordand.17.1.4

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

Plonsky, L., & Derrick, D. J. (2016). A meta-analysis of reliability coefficients in second language research. *The Modern Language Journal, 100*(2), 538–553. https://doi.org/10.1111/modl.12335

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

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

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. https://www.R-project.org/

Ringbom, H. (2007). *Cross-linguistic similarity in foreign language learning*. Multilingual Matters.

Rossiter, M. J. (2009). Perceptions of L2 fluency by native and non-native speakers of English. *The Canadian Modern Language Review, 65*(3), 395–412. https://doi.org/10.3138/cmlr.65.3.395

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge.

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

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

Suzuki, S., Kormos, J., & Uchihara, T. (2021). The relationship between utterance and perceived fluency: A meta-analysis of correlational studies. *The Modern Language Journal, 105*(2), 435–463. https://doi.org/10.1111/modl.12706

Tergujeff, E. (2021). Second language comprehensibility and accentedness across oral proficiency levels: A comparison of two L1s. *System, 100*, Article 102567. https://doi.org/10.1016/j.system.2021.102567

Appendix A: Cronbach's Alpha Coefficients for the Fluency Ratings by Experiment (I vs. II) and Condition (Control vs. Experimental)

	Experiment I (L2 English)		Experiment II (L3 Swedish)	
Rated variable	**Control**	**Experimental**	**Control**	**Experimental**
Speed fluency	.906	.900	.954	.962
Breakdown fluency	.927	.924	.965	.981
Repair fluency	.888	.906	.898	.905
Overall fluency	.917	.919	.972	.980

Appendix B: Results of the Shapiro-Wilk normality tests and Levene's tests

The data were normally distributed for G1 for all the rated variables in both experiments, and for G2 for speed, breakdown, and overall fluency (Shapiro-Wilk tests: $p > .05$) in Experiment I. For repair fluency, there was one extreme outlier in G2 in both listening conditions, resulting in a negatively skewed distribution (Shapiro-Wilk tests: $p < .05$). As the outlier was the same speaker in both conditions, the outlier did not affect the results of the ANOVAs: when the analyses were conducted without the outlier, the effect of listening condition remained significant ($F = 95.00$, $p < .001$, $\eta^2 = .133$, $\eta^2_p = .779$), and the effect of L1 background ($F = 3.13$, $p = .088$, $\eta^2 = .100$, $\eta^2_p = .104$) and the interaction between L1 and listening condition ($F = 0.56$, $p = .460$, $\eta^2 < .001$, $\eta^2_p = .020$) remained non-significant. No significant violations to the assumption of homogeneity of variances (Levene's tests: $p > .05$) were found for any of the rated variables in Experiment I.