Pauliina Peltonen*, Sanna Olkkonen, Magdalena Szyszka and Pekka Lintunen

L2 repair fluency through the lenses of L1 repair fluency, cognitive fluency, and language anxiety

https://doi.org/10.1515/applirev-2023-0011
Received January 13, 2023; accepted April 9, 2024; published online May 6, 2024

Abstract: Repairs (including false starts, repetitions, and different types of self-corrections) have been examined in second language (L2) speech fluency research as one dimension of (dis)fluent speech. However, in contrast to other dimensions of L2 speech fluency (speed and breakdown), repair fluency is not equally well understood: the results are mixed, and more research investigating the factors behind L2 repair fluency is needed. While some previous studies suggest links between first language (L1) and L2 repair fluency, to what extent L2 repairs are connected with cognitive and affective factors is less understood. To achieve a comprehensive view of the factors behind L2 repair fluency, we combine perspectives of L1 repair fluency, attention control, and language anxiety (LA) that have individually been shown to potentially affect L2 repairs but have rarely been examined together. We analyzed data from L1 Finnish and L2 English monologue speech tasks, a Stroop task in L1 and L2, and surveys for general and task-specific LA from 59 advanced users of English to investigate how L1 repair fluency, cognitive fluency, and LA are related to L2 repair fluency. Correlational analyses revealed that task-specific LA and certain Stroop measures were connected with L2 repair measures, while correlations between L1 and L2 repair fluency measures were weak. An analysis of repair profiles of participants displaying the highest levels of L2 repair fluency revealed that, overall, repairs are more common in the L2 than in the L1, but patterns regarding preferences for repair types vary across individuals. The study has methodological implications for psycholinguistic and SLA research into L2 repair fluency and broader implications for L2 classrooms and assessment.

***Corresponding author: Pauliina Peltonen**, Department of English, University of Turku, Arcanuminkuja 1, 20014 Turku, Finland, E-mail: paupelt@utu.fi
Sanna Olkkonen and Pekka Lintunen, Department of English, University of Turku, Turku, Finland
Magdalena Szyszka, Department of English, University of Turku, Turku, Finland; and Department of Linguistics, University of Opole, Opole, Poland

Keywords: second language speech; fluency; repair; attention control; language anxiety

1 Introduction

Fluency is widely studied as a central aspect of more general oral proficiency in the second language (L2). According to Lennon's (1990) widely cited distinction, a 'broad sense' of fluency as the level of overall (oral) proficiency can be distinguished from the 'narrow sense' of fluency. In research contexts, the narrow sense is traditionally applied and fluency is approached as one dimension of oral proficiency related to smoothness and effortlessness (Chambers 1997). SLA studies on fluency have chiefly been concerned with examining 'utterance fluency' (Segalowitz 2010), which refers to fluency-related features in a speech sample and is roughly comparable with the narrow sense in Lennon's (1990) terminology. The present study focuses on one dimension of utterance fluency, 'repair fluency' (Skehan 2009), which typically includes false starts (FSs), repetitions, and different types of self-corrections. The efficiency of the processing underlying fluent speech production, i.e., 'cognitive fluency' (Segalowitz 2010), is also examined as a potential factor influencing L2 repair fluency from the perspectives of automaticity and attention control. As the processing can also be influenced by affective factors, we include 'language anxiety' (LA) as another potential factor linked with L2 repair fluency.

Levelt's (1989) model for speech production, with its modifications for L2 oral production (de Bot 1992; Kormos 2006), provides a theoretical underpinning for better understanding the role of the aforementioned factors in this study. According to the model (Levelt 1989), the act of speaking starts at conceptual macro- and microplanning stages ('conceptualisation'), resulting in the formulation of a preverbal message, which is further encoded into a grammatical structure supported by the mental lexicon ('formulation'). As soon as the preverbal message is converted into a surface structure, the phonological/phonetic system is activated to form an overt speech ('articulation'; Levelt 1989). Limited L2 resources may pose challenges to speech fluidity at microplanning, formulation, and articulation stages (de Bot 1992). Additionally, regardless of language, speech is monitored by a speaker at various stages of oral production (Kormos 2006). This self-monitoring allows to detect errors in speech, which are either self-corrected, resulting in repairs, or ignored. Self-monitoring, thus, requires attentional resources and can be more cognitively demanding in the L2 than in the first language (L1) (Segalowitz 2010). Moreover, anxiety exhibited in the course of L2 speaking involves attentional control processes (Eysenck et al. 2007), which may further interplay with self-monitoring and repairs.

While previous research has explored links between cognitive and utterance fluency (e.g., Kahng 2020) and utterance fluency and affective factors (e.g., Aubrey 2022), our study is among the first to examine the three aspects together from the perspective of repairs. Previous studies have suggested individual variation in L2 repair use, but these findings have mostly been related to L1 repair fluency, whereas more research is needed on how cognitive processing and affective factors, such as LA, influence L2 repair fluency patterns. Furthermore, repair fluency itself is not equally well understood compared to the other dimensions of fluency, namely *speed fluency* and *breakdown fluency* (Skehan 2009), underscoring the need for further research in this area.

In the present study, we combine perspectives of L1 repair fluency, cognitive fluency, and language anxiety (LA) to achieve a comprehensive view of the factors behind L2 repair fluency. L1 Finnish and L2 English repair fluency were analyzed from the same participants based on monologue speech samples from 59 advanced users of English. Cognitive fluency was measured with a Stroop task in the L1 and L2, and LA is approached both from general and task-specific perspectives with questionnaire data. Our analyses focused on the connections between L1 and L2 repair fluency, cognitive fluency, and LA and the individual profiles of participants displaying the highest levels of L2 repair. Bringing together these perspectives in an integrative approach, the study increases understanding of L2 repair fluency and the factors behind it.

2 Theoretical background

In this section, we present the theoretical background for our study. First, we discuss L1 and L2 repair fluency based on utterance fluency research (Section 2.1), followed by the cognitive perspective on L2 repair fluency (Section 2.2) and links between language anxiety and repair fluency (Section 2.3).

2.1 L1 and L2 repair fluency

In L2 fluency studies, utterance fluency is often examined based on three dimensions of fluency proposed by Skehan (e.g., 2009): speed (e.g., articulation rate), breakdown (frequency, duration, and location of pauses), and repair fluency (different types of adjustments to speech). The first two have been extensively examined and are widely documented to be linked with learners' L2 proficiency levels and to correlate with listeners' assessments of L2 fluency (e.g., Kahng 2014; Kormos and Dénes 2004; Tavakoli et al. 2020). However, the present study focuses on the third, repair fluency dimension, as it can be regarded as the least understood dimension of the three (see also, e.g., Bosker et al. 2013; Kahng 2014). While repair fluency has been operationalized with slightly different frameworks, three aspects are commonly included: FSs (cut-off or rejected words), repetitions, and self-corrections (e.g., Skehan 2009). In the present study, we examine all of these aspects and make a distinction between single-word self-corrections (replacements) and more extensive self-corrections (reformulations) following Foster and Skehan's (1999) classification (see Section 4.2.3).

Compared to the speed and breakdown dimensions, studies have not found an equally straightforward relationship between repair fluency and L2 proficiency (Kormos and Dénes 2004; Lennon 1990; Tavakoli et al. 2020) or L2 fluency assessments (e.g., Bosker et al. 2013; Dumont 2018). The findings might be influenced by the fact that repair fluency features tend to be relatively infrequent in L2 speech data sets (e.g., Götz 2013; Kahng 2014) and are subject to relatively strong within-group variation (e.g., Dumont 2018; Götz 2013; Kahng 2014; Tavakoli et al. 2020), which underscores the need to examine speaker profiles (see also Olkkonen et al. in print). In addition to repair features being susceptible to individual differences, they can be multifunctional. In particular, FSs and self-corrections differ from repetitions, which can be used strategically as 'stalling mechanisms' to cope with processing time pressure during planning and to avoid lengthy silences to maintain speech fluency (e.g., Dörnyei and Kormos 1998; Götz 2013; Peltonen 2020). To uncover learners' individual patterns regarding the use of repair features, the present study includes an analysis of the participants' repair fluency profiles in their L1 and L2 and relates this behavior to their cognitive fluency and language anxiety.

The approach to repair fluency in the present study is particularly novel, as L1 and L2 repair fluency are analyzed from the same participants. Recently, there has been increasing interest in exploring how L1 speaking style is related to L2 speech fluency with various L1–L2 combinations and L2 proficiency levels (e.g., De Jong et al. 2015; Duran-Karaoz and Tavakoli 2020; Huensch and Tracy-Ventura 2017; Kahng 2020; Peltonen 2018). Overall, these studies indicate links between L1 and L2 fluency, but the strength of the associations varies to some extent across fluency dimensions. The differing findings might also be related to the languages and populations examined in the studies. However, the repair fluency dimension has not yet been fully explored from this perspective: for instance, Duran-Karaoz and Tavakoli (2020) only examined one repair fluency measure, number of repair. De Jong et al. (2015), Huensch and Tracy-Ventura (2017), and Kahng (2020) included two measures of repair fluency, repetitions per second and corrections per second. In contrast, (Peltonen 2018) only examined repetitions per minute. The correlations between L1 and L2 repair fluency features in these studies ranged from mostly small (Huensch and Tracy-Ventura 2017; Peltonen 2018) to medium (Duran-Karaoz and Tavakoli 2020) or both (Kahng 2020), one study reporting strong correlations (De Jong et al.

2015). Therefore, the results are relatively mixed. Our study adds to this body of research by being the first one to examine the connections between L1 and L2 repair fluency with four repair measures, thus providing a particularly comprehensive perspective on repair fluency.

2.2 Cognitive fluency and repair fluency

The observable utterance fluency is, in part, determined by the fluency of cognitive processing. Cognitive fluency is explained by Segalowitz (2010: 165) as "the efficiency of operation of the underlying processes" behind speaking, and this efficiency depends, in part, on the amount of available cognitive resources. When, for example, aspects of grammar and phonology of a language automatize, cognitive resources are released for the more controlled operations, such as monitoring and planning (see also Levelt 1989). Thus, while cognitive fluency is often studied from the viewpoint of linguistic resources and processing speed (e.g., Suzuki and Kormos 2023), following Segalowitz (2010), we define it as an interplay of automatic and controlled processing (see also Olkkonen 2017b). Regarding repairs, while automaticity in language processing should be related to fewer repairs, the role of control of attention is less clear.

Control of attention is often studied based on neurocognitive control mechanisms called executive functions, which typically include monitoring, inhibition, and switching (Miyake et al. 2000). However, only a few studies have examined connections between repairs in the L1 or L2 speech and these aspects of cognitive processing. The amount of available cognitive resources is crucial for monitoring (Kormos 2006; Levelt 1989), for example, as more challenging tasks even in the L1 make the speech monitoring less accurate, leading to more errors and fewer repairs (Oomen and Postma 2002). As the resources are usually more limited in the L2, speakers seem to correct a smaller proportion of their errors than in the L1, and the types of repairs differ somewhat between the L1 and L2 (Kahng 2014; Kormos 2000). Less advanced language users have also been shown to correct themselves less than more advanced ones during lexical access tasks, both in the L1 and L2, due to both resource and proficiency limitations (Olkkonen 2017a). Therefore, while more resources for monitoring help to ascertain the accuracy of output, it may also increase the number of repairs.

Furthermore, as previously noted, individual differences influence repair fluency. Zuniga and Simard (2019) showed that individual L1 repair behavior and a general control of attention skill (measured with a trail making test) explained 40 % of L2 repairs (Zuniga and Simard 2019; see also Kahng 2020; Nguyen et al. 2020). In the present study, we used a Stroop task (Stroop 1935), which measures, in addition to automaticity, the individual efficiency of inhibiting irrelevant material. The task consists of neutral and color words, written in different color inks, and one has to name the ink instead of the word. The RTs especially in the incongruent condition, where the color word and ink do not match, reveal both the automaticity of the color word activation, as well as the ability to control that activation. Inability to inhibit the wrong answer, thus, would lead to more repairs. Engelhardt et al. (2013), for example, found that fewer repetitions and repairs in a sentence construction task correlated significantly with the inhibition ability in the Stroop task. The study was, however, concerned only with L1 processing. As the activation of vocabulary is one of the points where the L2 performance greatly differs from that of the L1 (e.g., Levelt 1989), we were interested in how similar this pattern is in the L2. It has been previously demonstrated that the interference from an involuntarily activating material starts to resemble that of the L1 only slowly with developing proficiency (Braet et al. 2011; Olkkonen et al. in review). Therefore, in the present study, the Stroop task was conducted both in the L1 and L2 to better understand if there are differences in the patterns of control of attention that affect L2 repair behavior. As language anxiety levels interplay with cognitive processes and, in particular, control of attention (e.g., Eysenck et al. 2007), we also measured language anxiety to examine the extent to which affective factors interplay with L2 repairs.

2.3 Language anxiety and repair fluency

In SLA research, language anxiety (LA) has frequently been conceptualized as "the worry and negative emotional reaction aroused when learning or using a second language" (MacIntyre 1999: 27). This individual affective factor is thought to obstruct cognitive resources, hindering language acquisition and speech performance in particular (Gregersen and MacIntyre 2014; Horwitz 2010). More specifically, LA may pose limitations on all stages of Levelt's (1989) model of speech production: conceptualisation, formulation, articulation, and self-monitoring (Kormos 2006; Pérez Castillejo 2019). The levels of LA experienced at the earlier stages may be transferred to the following stage, producing constantly increasing LA levels, which are accumulated at the level of articulation and self-monitoring (Piechurska-Kuciel 2008). At this final stage of language production, anxiety may resonate in repair fluency features, which are related to the outcomes of inefficient processing of L2 speech (Kormos 2006). Despite this potential link, to our knowledge, very few studies have addressed the link between LA and repair fluency specifically (cf. Zuniga and Simard 2022).

Since the seminal study of Horwitz et al. (1986), LA has been approached from general and task-specific perspectives. The former addresses LA as a relatively stable and solidified emotional reaction manifested in the setting of L2 learning and use (Horwitz et al. 1986), while the latter defines LA as a more dynamic and temporary condition (MacIntyre 2017). Evidence from research on general LA has confirmed its relationship with a number of learner internal and external factors, such as L2 proficiency (e.g., Piechurska-Kuciel 2008), perceived competence, and willingness to communicate (Mystkowska-Wiertelak and Pawlak 2017), and its negative effect on L2 speech performance (see Horwitz 2010; MacIntyre 2017). Research on task-specific LA, still in its infancy, aims to explain the role of interpersonal and social contexts as well as communicative situations in anxiety fluctuations (MacIntyre 2017: 26).

So far, research exploring the link between speech fluency and LA from general and task-specific perspectives has generated inconsistent outcomes (Aubrey 2022; Bielak 2022; Kormos and Préfontaine 2017; Pérez Castillejo 2019; Zuniga and Simard 2022). Strong positive correlations between breakdown fluency (average length of pauses) and LA were reported by Aubrey (2022), who used the idiodynamic approach, where four participants rated their fluctuations of anxiety levels during monologue tasks. Momentary breakdowns in conceptualisation and formulation speech processes were identified among the factors that impinge on the emotion-fluency relationship. Following a similar approach, Bielak (2022) measured breakdown fluency and speed fluency based on, among others, an L2 English monologue task ($N = 43$). His findings disclosed that task-specific LA was positively correlated ($r = 0.59$) only with one measure of breakdown fluency: the ratio of Analysis of Speech unit boundary pauses. However, no links between task-specific LA and breakdown or speed fluency measures were found in Kormos and Préfontaine's (2017) study. Yet, Pérez Castillejo (2019) confirmed that general LA was a strong predictor of several breakdown and speed fluency indices. In a similar vein, Zuniga and Simard (2022) found a moderate positive relationship between Self-Initiated Self-Repair and LA, and a high positive correlation between LA and self-repairs for slow-attention speakers ($r = 0.78$). The inconclusive outcomes call for further investigations in this area, particularly in the domain of repair fluency. The present study incorporates both general and task-specific LA measures to explore their relationships with L2 repair fluency.

3 Rationale and research questions

Based on our review of previous research on repair fluency from the perspectives of L1 and L2 utterance fluency, automaticity and attention control as indicators of cognitive fluency, and LA in Section 2, the findings of some utterance fluency studies suggest that L1 and L2 repair fluency are linked. However, the results are somewhat mixed and highlight individual differences in repair use. Similarly, cognitive approaches to repair fluency highlight their individual nature; yet this approach has only rarely been combined with an analysis of repairs in L1 and L2 speech (but see Kahng 2020). Furthermore, the studies conducted on the links between LA and repair fluency provide somewhat mixed results, highlighting the need for further research in the area. Apart from Zuniga and Simard's (2022) study, to our knowledge, previous studies have not brought together the three perspectives of utterance fluency, cognitive fluency, and affective factors in understanding L2 repair fluency. Thus, following from this gap in research, our study aims to explore L2 repair fluency and its relation to L1 repair fluency, cognitive fluency, and LA based on the following research questions (RQs):

1) To what extent do L1 repair fluency, cognitive fluency, and general and task-specific LA correlate with L2 repair fluency?
2) What kinds of profiles do participants with the highest levels of L2 repair display regarding L1 repair fluency, cognitive fluency, and general and task-specific LA?

4 Methodology

In this section, we discuss the methodology of the present study, starting with Participants (Section 4.1) and followed by Data and procedure (Section 4.2) and Analysis (Section 4.3).

4.1 Participants

The participants of the present study were L1 Finnish university students, with L2 English (N = 59, age M = 22.78). The participants were advanced users of English: the majority of them studied English as their major (n = 44) or minor (n = 9) subject. Based on the participants' LexTale (see Lemhöfer and Broersma 2012) scores, most of them represented the C1/C2 level (n = 42; scores above 80 %), while 17 participants represented the B2 level (scores between 60 and 80 %) in the CEFR (Council of Europe 2001). The participants were chosen from a larger pool of students that participated in the project *Fluency and Disfluency Features in L2 Speech* (*FDF2*; funded by the Research Council of Finland in 2020–2024, decision number 331903). They were screened to have normal or corrected-to-normal vision, and no language-related problems. Furthermore, because of the Stroop task, color blindness was an exclusion criterion. Informed consent was obtained from all participants.

4.2 Data and procedure

4.2.1 Data collection

The data were collected during 2021 as part of the *FDF2* project. The tests were conducted as a part of course work in a group setting during a 90 min session. After being informed about the ethical aspects of the study, including the privacy notice, the participants produced speech samples in their L1 Finnish and L2 English based on cartoon strip prompts in a language laboratory. The participants were instructed to describe the story depicted in the cartoons in their own words. They were given 2 min of individual planning time to have enough time to familiarize themselves with the story and to plan the contents of their description (see also Kormos and Dénes 2004). They were allowed to look at the cartoon while telling the story to prevent the influence of retrieving the story from their memory on the production. Next, the participants performed a set of cognitive fluency tasks in a computer classroom, of which the Stroop task (Stroop 1935) was used in the present study as a measure of both automaticity and attention control. Finally, the participants responded to online questionnaires related to their background information and affective factors, capturing general LA and task-specific LA.

4.2.2 Instruments and elicitation materials

The L1 Finnish and L2 English speech samples were elicited with cartoon strip prompts, as described in Section 4.2.1. Narrations based on picture prompts are a common way to elicit speech data for fluency studies (e.g., Lennon 1990), and the two prompts were designed to be as comparable as possible: all included six frames and a clear storyline. To control for potential order effects, the order of the cartoons and the languages were counterbalanced in the data collection design.

Cognitive fluency was measured with Stroop tasks. The task items (144 in total) included four color words in Finnish and in English (*punainen/red, sininen/blue, keltainen/yellow, vihreä/green*) and three neutral words (*ovi/door, ikkuna/window, tuoli/chair*) written in different color inks (red, blue, yellow, and green ink). There were three conditions in both languages: the congruent condition with the word printed in the corresponding color (*punainen/red* in red); the incongruent condition with a different color ink (*punainen/red* in green); and the neutral condition with a neutral word in one of the four colors (*ovi/door* in red). Moreover, the stimuli words could be either in L1 or L2 (mixed design), but the answering language was blocked (either L1 or L2 first) (see Heidlmayr et al. 2014).

LA was approached from two perspectives in the present study: general LA levels were examined based on the Foreign Language Classroom Anxiety Scale (the FLCAS; Horwitz et al. 1986), and the Post-Session Survey on Anxiety (the PSSA) determined task-specific LA exhibited after the monologue task performance. The FLCAS was used due to its "consistently high reliability" (MacIntyre 2017: 16) and its wide application in research on LA. It comprises 33 Likert scale items (from *1 – I strongly disagree* to *5 – I strongly agree*) that refer to language learning situations evoking anxiety, such as *I am usually at ease during English language tests*. The PSSA contained three questions regarding the L2 monologue task. The first, quantitative one (*How anxious were you while performing the monologue task in English?*) was answered on a scale from *1 – not at all, 2 – somewhat, 3 – quite anxious, 4 – very anxious* and *5 – extremely anxious*. Qualitative data were collected from two open-ended questions: *What factors impinged on your anxiety in this session?* and *How comfortable did you feel while speaking English in the monologue task today?*

4.2.3 Data treatment

The monologue speech samples were used as the basis for measuring L1 and L2 repair fluency in the present study (L1 sample duration M = 63.80 s, SD = 19.40; L2 sample duration M = 68.75 s, SD = 23.96). The samples were first transcribed and double-checked by a group of research assistants as part of course work under the supervision of the first and second authors. The research assistants also annotated the samples manually with the speech analysis software Praat (Boersma and Weenink 2022) for four repair fluency features (based on Foster and Skehan 1999; see Table 1). Twelve percent of the annotations were double-checked by the first author. This was regarded as sufficient to ensure the reliability of the annotations due to the detailed instructions received by the assistants and the opportunity to discuss potentially difficult cases as part of the course. Foster and Skehan's (1999) classification was chosen, as it contains the most commonly examined repair categories in L2 fluency research and provides clear definitions for each category.

To calculate the repair fluency measures, the frequencies of the annotated features were extracted with a script (Lennes 2002). The frequencies of repair fluency features were standardized per minute of speaking time (total time excluding silent pause time), following a common practice in utterance fluency research (based on De Jong 2016).

Regarding the analysis of the Stroop task performances, in the present study, only the mean reaction times (RTs, see, e.g., Braet et al. 2011) in the incongruent condition were included, as these are most clearly related to the control of attention. RTs were measured with Praat by research assistants (10 % double-checked by the

Table 1: Repair fluency features with operationalizations and examples in L1 Finnish and L2 English.

Repair fluency feature	Operationalization	Example from the L1 Finnish data set (repair feature in bold)	Example from the L2 English data set (repair feature in bold)
False start (FS)	Rejected, cut-off sound, word, or longer utterance	tässä lapset ovat **rakenta-** rakentaneet lumiukon (P-65)	**the ma-** uh boy came back to the tree (P-38)
Repetition	Word or longer utterance repeated without modification	jonka sitte (.) öö takia tuo (0.30) **tuo** lakki (.) lentää (P-29)	then the: (0.48) **the** flower grows into a tree (P-65)
Replacement	Word replaced with another word without additional modifications	nää (.) **niil** on tämmönen (0.34) ta:imi tässä (P-54)	the girl is: **has** a watering can (P-6)
Reformulation	Utterance longer than one word repeated with some modification	ja sitten ne (3.08) **sitten siel** on omenoita (P-54)	with both of thee- **both the girl and the boy** standing next to it (P-29)

Note: The reformulation examples also contain false starts ("sitten ne" and "with both thee-").

second author), starting from the beginning of pronouncing the correct answer and excluding the wrong answers and late responses (0.6 % of all answers).

LA was analyzed based on the FLCAS and the PSSA responses. The aggregated sum of scores in the FLCAS indicated the level of LA, ranging from 33 (Min.) to 165 (Max.). A higher score revealed a higher level of LA. A Cronbach alpha coefficient of 0.91 demonstrated a high internal reliability of the instrument.

4.3 Analysis

The statistical analyses were performed in SPSS (version 27). Due to non-normal distributions in the data, non-parametric Spearman's rank-order correlation coefficients were calculated to answer RQ1 (p-values < 0.05 treated as statistically significant). Plonsky and Oswald's (2014: 88) recommendations for interpreting effect sizes were followed: r = 0.25 small effect, r = 0.40 medium effect, and r = 0.60 large effect. Regarding RQ2, the participants with the highest levels of L2 repair fluency were selected for a detailed analysis (n = 7). The chosen participants (high L2 repair group; HRG) used 10 or more L2 repairs in total based on the raw frequencies (Participants 6, 29, 68) or the standardized measure L2 repairs per minute of speaking time (Participants 38, 53, 54, 65, 68). In addition to an analysis of the participants' individual L1 and L2 repair fluency, cognitive fluency, and LA measures, the information about their profiles was complemented with an analysis of their open-ended answers to the two PSSA questions.

5 Results

Before discussing the results for the two RQs, the descriptive statistics for all measures for the whole sample and for the HRG are presented in Table 2.

As seen from Table 2, regarding the L1 and L2 repair fluency measures, the patterns were relatively similar in the two languages based on the averages for the whole sample (*N* = 59). Overall, the average frequencies were quite low for both L1 and L2 repair fluency measures. Of the four measures, FSs were the most common type, on average, both in the L1 and in the L2, while replacements were the least frequent type. Repetitions were somewhat more frequently used, on average, in the L2 than in the L1. However, based on Wilcoxon signed ranks tests (*N* = 59) between the L1 and L2 repair fluency measures, the differences were not statistically

Table 2: Descriptive statistics for the whole sample and the HRG (repair totals in bold).

Measure	Whole sample (*N* = 59)		HRG (*n* = 7)	
	M	*SD*	*M*	*SD*
L1 FSs	2.63	2.58	3.28	2.78
L1 repetitions	0.64	1.19	1.49	1.86
L1 replacements	0.31	0.80	1.08	1.71
L1 reformulations	0.59	1.10	0.50	0.87
L1 repair total	**4.17**	**3.39**	**6.35**	**4.10**
L2 FSs	2.52	2.32	5.95	2.19
L2 repetitions	1.21	1.94	3.09	3.80
L2 replacements	0.34	0.90	1.96	1.77
L2 reformulations	0.66	1.21	2.36	2.28
L2 repair total	**4.73**	**4.39**	**13.36**	**5.07**
Stroop L1/L1 (ms)	883.73	142.48	875.42	79.67
Stroop L1/L2 (ms)	846.94	132.55	841.12	107.61
Stroop L2/L1 (ms)	917.31	147.68	961.78	121.00
Stroop L2/L2 (ms)	917.80	136.01	960.30	117.82
General LA	76.24	20.00	78.43	20.94
Task-specific LA	1.97	0.98	2.00	0.58

Note: L1/L1 Stroop = answer and stimulus in L1; L1/L2 Stroop = answer in L1, stimulus in L2; L2/L2 Stroop = answer and stimulus in L2; L2/L1 Stroop = answer in L2, stimulus in L1. All repair measures are reported per minute (min.) of speaking time.

significant.[1] It should also be noted that the standard deviations were comparably high for all L1 and L2 repair measures, indicating important variation between participants within each group in terms of repair behavior. While the standard deviations were also relatively high for the four Stroop measures, the results regarding the differences in the mean values were slightly different compared to the L1 and L2 repair fluency measures. The overall RTs, on average, were faster for the L1 than for the L2; that is, the participants were slower to answer in their L2. RTs were also slower when answering in L1 with a stimulus in L1 compared to having the stimulus in the L2. In comparison, when answering in the L2, the language of stimulus did not affect the mean RTs. Finally, the general and task-specific LA reached the mean values of 76.24 and 1.97, respectively, for the whole sample. The first value is below the mid-value between the minimum and maximum score on the FLCAS, which is 99. The second also indicates a relatively low task-specific LA below the average of 3 on this scale.

Regarding the HRG (n = 7), the means for each L2 repair measure were higher than for the whole sample. Similarly, regarding L1 repair measures, the HRG differed from the whole sample to some extent, as the L1 repair fluency averages were higher based on the averages (with the exception of reformulations). The L1 repair total average was also higher in the HRG than in the whole sample. In the Stroop task, the average L2 RTs of the HRG were markedly slower than those of the whole sample. Finally, the mean levels of both general and task-specific LA in the HRG were minimally higher compared to the whole sample.

5.1 Connections between L1 repair fluency, cognitive fluency, LA, and L2 repair fluency

The correlations between L1 repair fluency, Stroop measures, general and task-specific LA, and L2 repair fluency (RQ1) are presented in Table 3.

As seen in Table 3, overall, the correlations between L1 repair fluency measures and L2 repair fluency measures were positive, weak, and statistically non-significant. The strength of the association was the strongest between L1 and L2 repairs (r = 0.231), but still small in effect and not statistically significant. Stroop RTs correlated statistically significantly and positively with FSs (L1/L2 r = 0.281; L2/L1 r = 0.360; and L2/L2 r = 0.367).[2] Other connections did not reach significance. The correlations

---

1 FSs per min. $Z = -0.350$, $p = 0.734$; repetitions per min. $Z = -1.589$, $p = 0.112$; replacements per min. $Z = -0.065$, $p = 0.948$; reformulations per min. $Z = -0.392$, $p = 0.695$; repair total $Z = -0.871$, $p = 0.384$.

2 Complementary simple regression analyses suggested that of the three Stroop measures, Stroop L2/L1 had the strongest predictive power in explaining L2 FSs (8.5 %; $B = 0.006$, Std. error = 0.002, $\beta = 0.317$,

Table 3: Correlations between L1 repair, Stroop measures, LA, and L2 repair measures (N = 59).

Measure	L2 FSs	L2 repetitions	L2 replacements	L2 reformulations
L1 FSs	0.090	0.041	0.028	−0.002
L1 repetitions	0.065	0.231	0.103	0.126
L1 replacements	−0.106	−0.030	0.079	0.193
L1 reformulations	−0.152	−0.134	0.132	0.007
Stroop L1/L1 (ms)	0.233	0.250	−0.030	0.019
Stroop L1/L2 (ms)	0.281[a]	0.182	−0.030	0.063
Stroop L2/L1 (ms)	0.360[b]	0.217	0.167	−0.048
Stroop L2/L2 (ms)	0.367[b]	0.225	0.127	0.026
General LA	0.141	0.019	0.013	−0.073
Task-specific LA	0.078	0.090	0.186	0.282[a]

Note: [a]$p < 0.05$, [b]$p < 0.01$.

between general LA and all L2 repair measures were close to zero, suggesting a lack of relationship between the measures. Similarly, task-specific LA levels were weakly and non-significantly related to three L2 repair measures, whereas a statistically significant positive relationship was found between task-specific LA and L2 reformulations (r = 0.282).

5.2 Profiles of high L2 repair participants regarding L1 repair fluency, cognitive fluency, and LA

The L1 and L2 repair fluency, Stroop RTs, and LA measures for the HRG (RQ2) are presented in Table 4.

As seen in Table 4, all HRG participants repaired more in their L2 than in the L1, yet displayed relatively varied profiles regarding their L1 and L2 repair fluency patterns. P-6 only used FSs and replacements in the L1, whereas their L2 production also contained one repetition and one reformulation. Overall, they mostly relied on short corrections (replacements) in the L2 sample. For P-6, the difference between the L1 and L2 Stroop RTs was the largest of the HRG, with very fast L1 performance and markedly slower L2. P-29 displayed a more varied repair profile in their L2 speech compared to L1 production, which only included repetitions. Interestingly, however, repetitions were the most common repair type in both their L1 and L2 production. In both languages, the repetitions mostly consisted of short function

F (1, 57) = 6.364, p = 0.014, Adj. R^2 = 0.085), followed by Stroop L2/L2 (7.5 %; B = 0.006, Std. error = 0.003, β = 0.301, F (1, 57) = 5.695, p = 0.020, Adj. R^2 = 0.075). The model for Stroop L1/L2 was not statistically significant (B = 0.003, Std. error = 0.003, β = 0.153, F (1, 57) = 1.363, p = 0.248, Adj. R^2 = 0.006).

Table 4: Measures for the HRG (repair totals in bold).

Measure	Participant (P) number						
	P-6	**P-29**	**P-38**	**P-53**	**P-54**	**P-65**	**P-68**
L1 FSs	3.05	0	2.28	1.37	6.17	7.90	2.19
L1 repetitions	0	5.01	0	2.75	1.54	0	1.10
L1 replacements	1.53	0	0	1.37	4.63	0	0
L1 reformulations	0	0	0	0	1.54	1.97	0
L1 repair total	**4.58**	**5.01**	**2.28**	**5.49**	**13.89**	**9.87**	**3.29**
L2 FSs	3.90	2.35	7.28	6.86	7.50	5.32	8.43
L2 repetitions	0.78	3.19	0	0	1.87	5.32	10.54
L2 replacements	3.12	0.78	1.82	0	3.75	0	4.22
L2 reformulations	0.78	1.60	3.64	6.86	1.86	1.77	0
L2 repair total	**8.59**	**7.85**	**12.74**	**13.72**	**15.00**	**12.42**	**23.19**
Stroop L1/L1 (ms)	745.92	823.92	981.33	850.00	863.70	945.00	918.09
Stroop L1/L2 (ms)	689.67	739.73	1,004.50	914.82	807.40	836.09	895.67
Stroop L2/L1 (ms)	987.42	740.90	1,081.00	911.56	899.55	1,052.73	1,059.33
Stroop L2/L2 (ms)	991.25	771.58	1,077.58	913.92	853.33	1,074.33	1,040.10
General LA	65	78	66	105	69	56	110
Task-specific LA	2	2	2	2	2	1	3

Note: All repair measures are reported per minute (min.) of speaking time.

words (e.g., "where the two: (.) two kids"). The Stroop results for P-29 in the L1 and L2 resembled each other very closely, with quicker cross-language performance in both languages, and they were the only participant with the L2 responses as fast or faster than the L1. In the open responses to the PSSA,[3] P-6 and P-29 mentioned that the background noise had impinged on their LA in the speaking tasks, but while P-6 perceived the noise as a protective curtain ("During the monologue task it was not as quiet, so I did not get as anxious as during other tasks"), P-29 viewed it as a distractor ("Hearing the other students influenced [my anxiety] a bit").

P-38 and P-53 displayed somewhat opposite patterns regarding L1 and L2 repair: while P-38 produced a relatively short L1 sample, only containing FSs, P-53 produced a short L2 sample with FSs and reformulations. Both produced longer samples in the other language and included three of the four repair types in the production. P-38 did not use repetitions in either language. Their Stroop results, on the other hand, were very similar as they were the only ones with L1/L2 RT slower than L1/L1; that is, the cross-language stimuli interfered with their production more than the within-language one. P-38 also had the slowest RTs across all conditions. P-53 (along with P-68) displayed a comparatively higher level of general LA than the others in the

3 The original responses in Finnish have been translated to English by the authors.

group. In the PSSA, P-53 openly stated, “I got a bit blocked during the monologue task” and added that LA was affected by the “possibility of making mistakes.”

P-54 was among the few participants displaying similar patterns of repair fluency across their L1 and L2 productions. They frequently used FSs and preferred short corrections (replacements) over longer reformulations in both languages. The Stroop RTs were quite similar between the languages, but still showing somewhat more interference in and from L1 than L2 and slower L2 RTs. In the PSSA, P-54 attributed their LA and disfluency to the general bodily conditions: “The tasks didn’t cause much anxiety, except for some ulterior reasons – I have slept badly in recent days so I am not in my sharpest […] (language is not as fluent as it would be in a normal situation).”

P-65 repaired relatively frequently in both languages, but interestingly used repetitions only in the L2, potentially functioning as a stalling mechanism for buying more planning time (see Dörnyei and Kormos 1998). Their L1 speaking style could be characterized as accuracy-oriented, as illustrated in (1) with their explicit marking of a correction related to a detail in the comic.

(1) mies kävelee lumiukkoa kohti silinterihattu päässä (1.74) korjaan silinterihattu kädessä
a man walks towards the snowman with a tophat on his head (1.74) I correct with a tophat in his hand

Similar behavior was not observed in the participant’s L2 production. The Stroop task results revealed that, along with P-29, P-65 had the same pattern in the L2 RTs as in the L1 (slower within-language performance), which is considered a sign of advanced language skills. Despite the high level of L2 competence (LexTale score 81.25 %, indicating CEFR level C1/C2), P-65 associated LA level with limited vocabulary size: “When speaking in the L1, anxiety is due to fear of underperforming, whereas in English due to relatively smaller vocabulary size compared to the L1.”

Finally, P-68 displayed the highest frequencies of L2 repair in the whole sample, producing particularly high frequencies of FSs and repetitions. Their L1 production also contained the same repair types, but lower frequencies. Interestingly, P-68’s scores on both task-specific and general LA were the highest in the group. In the PSSA, P-68 directly reported the relationship between mistakes and LA, and how they resulted in L2 speech disfluency: “I like to speak English, but recording and the number of mistakes made me a bit anxious and made me stumble on words I usually don’t struggle with.” P-68’s comparably frequent use of repairs in the L2 might also be linked to the level of L2 proficiency: while the other participants represented either the higher range of B2 or C1/C2 level in the CEFR, P-68 represented the lower range of B2 (LexTale score 63.75 %). The Stroop RTs are also in line with this interpretation, with slower L2 and more interference in L1.

6 Discussion

RQ1 examined the extent to which L1 repair fluency, cognitive fluency, and general and task-specific LA were correlated with L2 repair fluency. The correlations showed, to some extent, differing patterns across the different constructs. The findings of weak links across L1 and L2 repair fluency are consistent with some of the previous fluency studies (Huensch and Tracy-Ventura 2017; Peltonen 2018) but contradict the findings of studies that have found medium to strong correlations (De Jong et al. 2015; Duran-Karaoz and Tavakoli 2020; Kahng 2020). Of the four repair measures examined in the present study, L1 and L2 repetitions correlated most strongly, but the strength remained weak and statistically non-significant. The lack of statistically significant correlations can be attributed, at least to some extent, to the low frequencies of repair features in the sample. Compared to some previous studies, the present study mostly included students at C1/C2 level in the CEFR, potentially explaining the relatively infrequent repair use especially in the L2 (see also Tavakoli et al. 2020). Interestingly, however, the L1 and L2 repair measure averages did not differ on the level of the whole sample, corroborating some previous findings of non-significant differences in corrections across L1 and L2 (Kahng 2020) or repair fluency overall (Duran-Karaoz and Tavakoli 2020).

For the cognitive fluency measurements, the slower L2 RTs in the incongruent Stroop items, as compared to the L1 RTs, confirm that even for the cohort of advanced language users, the L2 activation was not as strong as that in the L1. This was also confirmed by the finding that the L2 performance was not influenced by interference to the same extent as the L1 (see, e.g., Braet et al. 2011; Olkkonen et al. in review); that is, it was easier to suppress the interference from L2, reflecting less automatic processing, which is in line with the cognitive fluency framework applied in the current study. Concerning RQ1, the significant correlations between the number of FSs and the measures involving L2 items point to an interpretation that these were related more to language proficiency, even at an advanced level, than to general inhibition skill (Stroop L1/L1). This was further confirmed by the results of the complementary regression analyses, especially when the answering language was the L2. In Engelhardt and colleagues' study (2013), the L1 repairs (mostly FSs), which correlated significantly with Stroop measures, were explained as instances where inhibition fails or is too slow to stop producing the wrong item. This is in line with the cognitive fluency framework used in the present study, as these findings can be explained with the more taxed cognitive resources when an L2 is included, to efficiently apply control of attention in either inhibiting or pre-monitoring the FSs. The lack of correlation between the other L2 repairs and the cognitive fluency measures suggests that, instead of relating to processing capacity or skills, they are more related to conscious aspects and strategic choices, further confirmed by their connection to affective factors in the present study. For the overall framework of automaticity and control, the results highlight that the different types of repairs seem to be related to different processes, which may help to understand the lack of straightforward connection with L2 proficiency.

While general LA did not correlate with L2 repair fluency, task-specific LA was found to be connected to the number of L2 reformulations. The higher the levels of task-specific LA were reported, the more reformulations the participants used. This finding may be linked to the data collection setting: although the participants performed a monologue task with a relatively low cognitive demand (cf. problem-solving tasks, e.g., Robinson and Gilabert 2007), their performance was recorded with other people around and might have resulted in elevated transient states of anxiety for some participants, as seen in some comments to the PSSA. Furthermore, the link between reformulations and task-specific LA can be explained from a cognitive perspective, as anxiety can be manifested in disruptions in the information retrieval (MacIntyre 2017). During retrieval problems, highly proficient L2 speakers may resort to reformulations rather than shorter repetitions or replacements, because these longer utterances provide ample time for message conceptualization and formulation. Therefore, the use of reformulations is potentially associated with optimal trade-off coping strategies, which compensate for advanced L2 speakers' limitations in speech production processes. Overall, our findings suggest that, in line with recent approaches to LA (Gkonou et al. 2017), the task-specific perspective on LA might be more suitable for examining the relationships between L2 speech fluency indices and LA, because it generates more calibrated data of individuals' feelings during L2 task performance compared to the traditional general LA measures.

RQ2 addressed the HRG's profiles regarding L1 and L2 repair fluency, cognitive fluency, and general and task-specific LA. Based on the analysis, while all HRG participants tended to repair more frequently in their L2 than in their L1, the patterns regarding the use of specific types of repairs in the L1 and L2 varied across the participants and even across L1 and L2, corroborating previous findings of individual differences in repair use (e.g., Dumont 2018; Götz 2013; Kahng 2014; Tavakoli et al. 2020). Our profile analysis, thus, suggests that repair behavior is highly individual and not necessarily connected across a participant's L1 and L2 productions. One exception was P-54, who exhibited remarkably similar repair patterns across their L1 and L2 speech samples; the Stroop RTs, however, showed a stronger interference for L1, suggesting that the automatization of vocabulary was not yet at the same level despite the high level of proficiency. Some evidence for individual preferences across languages could be detected for some of the other participants as well: for instance, for P-29, repetitions were the most common repair type in both languages, echoing some previous findings of strategic uses of repetitions as stalling mechanisms (e.g., Dörnyei and Kormos 1998; Peltonen 2020) and individual preferences in their use (e.g., Dumont 2018; Götz 2013). Overall, for the more proficient participants, the Stroop results resembled each other more between the L1 and L2 than for the less proficient participants, even though this was not reflected in the number of repairs. This is in line with the findings (see, e.g., Braet et al. 2011; Olkkonen et al. in review) demonstrating that, with more advanced L2 skills, the general processing abilities begin to show, as they become less restricted by the L2 processing capacity. The results also serve as a further reminder of how control of attention affects repair fluency in a way that cautions interpreting disfluencies solely as indicators of low proficiency (see also, e.g., Dörnyei and Kormos 1998; Götz 2013; Peltonen 2020).

Finally, the levels of general and task-specific LA in this group ranged to the extent precluding unequivocal conclusions, with the exception of a representative of the lower range of B2 proficiency level, P-68, whose LA levels and L2 repairs, both raw and standardized frequencies, were high. Perhaps, at a lower proficiency level, the interplay between high LA and repairs may be explained with the greater demand of cognitive resources needed to sustain fluent speech than in the case of higher proficiency levels. High LA levels may consume these resources, limiting efficient functioning of monitoring processes. This case draws attention to the role of L2 proficiency in research on the LA and L2 repair relationship.

7 Conclusion

Our study has been among the first to examine a combination of factors, including L1 repair, cognitive fluency (attention control), and affective factors (general and task-specific LA), behind L2 repair fluency. Thus, the present study provided a particularly comprehensive characterization of the complex dynamics related to L2 repair fluency. Our findings showed that while, overall, the links between L1 and L2 repair fluency were weak, task-specific LA and certain measures of cognitive fluency were connected with L2 repair measures. Our analyses underscore the need to analyze the different aspects of repair fluency features separately, as they were connected to cognitive fluency and both LA measures to a different extent. Additionally, the analysis of selected learners' individual profiles showed that examining the different types of repair in a detailed manner is worthwhile, as individual variation in the use of repairs was evident. The Stroop and LA measures showed that some of the participants' repair profiles were linked to their general cognitive processing and, of the LA measures, especially task-specific LA. These tentative results call for further investigations into the contextual factors capturing moment-to-moment fluctuations of individual variables, such as LA.

The integrative approach of the present study provides a starting point for further cross-disciplinary research into L2 fluency. The data set used in the present study, consisting of a large test battery of L1 and L2 speech data, L1 and L2 cognitive data, and affective questionnaire data from the same participants, was particularly well suited for analyzing individual patterns in L2 speech fluency. Collecting and examining data from the same participants in their L1s and L2s is encouraged in future fluency studies as well. While we did not observe a robust group-level connection between L1 and L2 repair fluency, other aspects of L1 and L2 fluency could be connected, as suggested by previous research. In future, the approach could be extended to interactional data, as repairs might have additional functions in dialogic settings compared to the monologic ones examined in the present study. Our study also has implications beyond fluency research: our findings of individual patterns in repair use could inform L2 assessment and teaching practices. As repairs are part of natural language use, along with other disfluencies, acknowledging individual variation in L1 repair could be the way towards more inclusive L2 assessment and teaching practices (see also Lintunen et al. 2020; Peltonen 2023). As our study has shown, observing the use of repairs in both the L1 and L2 provides insights into understanding the intricate underlying cognitive and affective processes of L2 speakers, which can guide future L2 fluency research and inform practical implications.

References

Aubrey, Scott. 2022. The relationship between anxiety, enjoyment, and breakdown fluency during second language speaking tasks: An idiodynamic investigation. *Frontiers in Psychology* 13. 968–946.

Bielak, Jakub. 2022. To what extent are foreign language anxiety and foreign language enjoyment related to L2 fluency? An investigation of task-specific emotions and breakdown and speed fluency in an oral task. *Language Teaching Research* (first view). https://doi.org/10.1177/13621688221079319.

Boersma, Paul & David Weenink. 2022. Praat: Doing phonetics by computer. Version 6.2.16. Available at: http://www.praat.org/.

Bosker, Hans Rutger, Anne-France Pinget, Hugo Quené, Ted Sanders & Nivja H. De Jong. 2013. What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing* 30(2). 159–175.

Braet, Wouter, Nele Noppe, Johan Wagemans & Hans Op de Beeck. 2011. Increased Stroop interference with better second-language reading skill. *Quarterly Journal of Experimental Psychology* 64(3). 596–607.

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

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

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

De Jong, Nivja H. 2016. Fluency in second language assessment. In Dina Tsagari & Jayanti Banerjee (eds.), *Handbook of second language assessment*, 203–218. Berlin: Mouton de Gruyter.
De Jong, Nivja H., Rachel Groenhout, Rob Schoonen & Jan H. Hulstijn. 2015. Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics* 36(2). 223–243.
Dörnyei, Zoltán & Judit Kormos. 1998. Problem-solving mechanisms in L2 communication: A psycholinguistic perspective. *Studies in Second Language Acquisition* 20(3). 349–385.
Dumont, Amandine. 2018. *Fluency and disfluency: A corpus study of non-native and native speaker (dis)fluency profiles*. Louvain-la-Neuve: Université catholique de Louvain dissertation.
Duran-Karaoz, Zeynep & Parvaneh Tavakoli. 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.
Engelhardt, Paul E., Joel T. Nigg & Fernanda Ferreira. 2013. Is the fluency of language outputs related to individual differences in intelligence and executive function? *Acta Psychologica* 144(2). 424–432.
Eysenck, Michael W., Nazanin Derakshan, Rita Santos & Manuel G. Calvo. 2007. Anxiety and cognitive performance: Attentional control theory. *Emotion* 7(2). 336–353.
Foster, Pauline & Peter Skehan. 1999. The influence of source of planning and focus of planning on task-based performance. *Language Teaching Research* 3(3). 215–247.
Gkonou, Christina, Mark Daubney & Jean-Marc Dewaele (eds.). 2017. *New insights into language anxiety: Theory, research and educational implications*. Bristol: Multilingual Matters.
Götz, Sandra. 2013. *Fluency in native and nonnative English speech*. Amsterdam: John Benjamins.
Gregersen, Tammy & Peter MacIntyre. 2014. *Capitalizing on language learners' individuality: From premise to practice*. Bristol: Multilingual Matters.
Heidlmayr, Karin, Sylvain Moutier, Barbara Hemforth, Cyril Courtin, Robert Tanzmeister & Frédéric Isel. 2014. Successive bilingualism and executive functions: The effect of second language use on inhibitory control in a behavioural Stroop Colour Word task. *Bilingualism: Language and Cognition* 17(3). 630–645.
Horwitz, Elaine K. 2010. Foreign and second language anxiety. *Language Teaching* 43. 154–167.
Horwitz, Elaine K., Michael B. Horwitz & Joann Cope. 1986. Foreign language classroom anxiety. *The Modern Language Journal* 70(2). 125–132.
Huensch, Amanda & Nicole Tracy-Ventura. 2017. Understanding second language fluency behavior: The effects of individual differences in first language fluency, cross-linguistic differences, and proficiency over time. *Applied Psycholinguistics* 38(4). 755–785.
Kahng, Jimin. 2014. Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning* 64(4). 809–854.
Kahng, Jimin. 2020. Explaining second language utterance fluency: Contribution of cognitive fluency and first language utterance fluency. *Applied Psycholinguistics* 41(2). 457–480.
Kormos, Judit. 2000. The role of attention in monitoring second language speech production. *Language Learning* 50(2). 343–384.
Kormos, Judit. 2006. *Speech processing and second language acquisition*. London: Lawrence Erlbaum Associates.
Kormos, Judit & Mariann Dénes. 2004. Exploring measures and perceptions of fluency in the speech of second language learners. *System* 32(2). 145–164.
Kormos, Judit & Yvonne Préfontaine. 2017. Affective factors influencing fluent performance: French learners' appraisals of second language speech tasks. *Language Teaching Research* 21(6). 699–716.
Lemhöfer, Kristin & Mirjam Broersma. 2012. Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods* 44. 325–343.

Lennes, Mietta. 2002. Total duration of labeled segments [Praat script]. Available at: https://github.com/lennes/spect/blob/master/scripts/total_duration_of_labeled_segments.praat.
Lennon, Paul. 1990. Investigating fluency in EFL: A quantitative approach. *Language Learning* 40(3). 387–417.
Levelt, Willem. 1989. *Speaking: From intention to articulation*. Cambridge, CA: MIT Press.
Lintunen, Pekka, Maarit Mutta & Pauliina Peltonen. 2020. Synthesising approaches to second language fluency: Implications and future directions. In Pekka Lintunen, Maarit Mutta & Pauliina Peltonen (eds.), *Fluency in L2 learning and use*, 186–201. Bristol: Multilingual Matters.
MacIntyre, Peter. 1999. Language anxiety: A review of the research for language teachers. In Dolly Jesusita Young (ed.), *Affect in foreign language and second language learning: A practical guide to creating a low-anxiety classroom atmosphere*, 24–45. New York, NY: McGraw-Hill.
MacIntyre, Peter. 2017. An overview of language anxiety research and trends in its development. In Christina Gkonou, Mark Daubney & Jean-Marc Dewaele (eds.), *New insights into language anxiety: Theory, research and educational implications*, 11–30. Bristol: Multilingual Matters.
Miyake, Akira, Naomi P. Friedman, Michael J. Emerson, Alexander H. Witzki, Amy Howerter & Tor D. Wager. 2000. The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology* 41(1). 49–100.
Mystkowska-Wiertelak, Anna & Mirosław Pawlak. 2017. *Willingness to communicate in instructed second language acquisition: Combining a macro- and micro-perspective*. Bristol: Multilingual Matters.
Nguyen, Tin Q., Stephanie N. Del Tufo & Laurie E. Cutting. 2020. Readers recruit executive functions to self-correct miscues during oral reading fluency. *Scientific Studies of Reading* 24(6). 462–483.
Olkkonen, Sanna. 2017a. Processing limitations in L2 fluency: Analysis of inaccuracies in lexical access. *APPLES: Journal of Applied Language Studies* 11. 19–41.
Olkkonen, Sanna. 2017b. *Second and foreign language fluency from cognitive perspective: Inefficiency and control of attention in lexical access*. Jyväskylä: University of Jyväskylä dissertation.
Olkkonen, Sanna, Maarit Mutta & Pekka Lintunen. In print. L2 speaker profiles in monologue production: The effect of cognitive background factors. In Minnaleena Toivola, Pekka Lintunen & Leena Maria Heikkola (eds.), *New directions in speech research – Freely produced speech as data. AFinLA Teema 17*. Jyväskylä: The Finnish Association for Applied Linguistics (AFinLA).
Olkkonen, Sanna, Patrick Snellings, Outi Veivo & Pekka Lintunen. In review. Cognitive fluency in L2: The effect of automatic and controlled lexical processing on speech rate.
Oomen, Claudy C. E. & Albert Postma. 2002. Limitations in processing resources and speech monitoring. *Language and Cognitive Processes* 17(2). 163–184.
Peltonen, Pauliina. 2018. Exploring connections between first and second language fluency: A mixed methods approach. *The Modern Language Journal* 102(4). 676–692.
Peltonen, Pauliina. 2020. *Individual and interactional speech fluency in L2 English from a problem-solving perspective: A mixed-methods approach*. Turku: University of Turku dissertation.
Peltonen, Pauliina. 2023. Fluency revisited. *ELT Journal*. 1–4. https://doi.org/10.1093/elt/ccad047 (Epub ahead of print).
Pérez Castillejo, Susana. 2019. The role of foreign language anxiety on L2 utterance fluency during a final exam. *Language Testing* 36(3). 327–345.
Piechurska-Kuciel, Ewa. 2008. *Language anxiety in secondary grammar school students*. Opole: Opole University Press.
Plonsky, Luke & Frederick L. Oswald. 2014. How big is “big”? Interpreting effect sizes in L2 research. *Language Learning* 64(4). 878–912.

Robinson, Peter & Roger Gilabert. 2007. Task complexity, the cognition hypothesis and second language learning and performance. *International Review of Applied Linguistics in Language Teaching* 45(3). 161–176.

Segalowitz, Norman. 2010. *Cognitive bases of second language fluency*. New York: Routledge.

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

Stroop, John Ridley. 1935. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology* 18. 643–662.

Suzuki, Shungo & Judit Kormos. 2023. The multidimensionality of second language oral fluency: Interfacing cognitive fluency and utterance fluency. *Studies in Second Language Acquisition* 45(1). 38–64.

Tavakoli, Parvaneh, Fumiyo Nakatsuhara & Ann-Marie Hunter. 2020. Aspects of fluency across assessed levels of speaking proficiency. *The Modern Language Journal* 104(1). 169–191.

Zuniga, Michael & Daphnée Simard. 2019. Factors influencing L2 self-repair behavior: The role of L2 proficiency, attentional control and L1 self-repair behavior. *Journal of Psycholinguistic Research* 48. 43–59.

Zuniga, Michael & Daphnée Simard. 2022. Exploring the intricate relationship between foreign language anxiety, attention and self-repairs during L2 speech production. *System* 105. 1–10.