

ARTICLE

Pause behavior within reformulations and the proficiency level of second language learners of English

Simon A. Williams^{1*}  and Malgorzata Korco²

¹University of Sussex and ²University of Middlesex

*Corresponding author. Email: s.a.williams@sussex.ac.uk

(Received 26 October 2017; revised 07 November 2018; accepted 07 November 2018)

Abstract

This research reports on a quantitative analysis of the combination of two types of disfluency, reformulations and pauses, in the speech of lower intermediate and advanced speakers of English as a second language (L2). The present study distinguishes between corrections and false starts within the category of reformulations as well as between silent and filled pauses. It focuses on the extent to which pause behavior within reformulations varies according to the stage of L2 development and the type of reformulation used. An analysis was made of 56 L2 speakers' 2-min monologues. The results showed that lower intermediate and advanced speakers differed on the frequency of silent pauses inserted in corrections but not on their frequency in false starts. This suggests that false starts depend less on proficiency level, and may reflect temporary problems with conceptual encoding or extralinguistic factors that contribute to the efficacy of L2 production rather than difficulties with linguistic processing per se. The frequency of silent pauses rather than silent pause duration or the frequency and duration of filled pauses appeared to be the only marker to differentiate between false starts and corrections across the two proficiency groups.

Keywords: bilingual speech production; pause; reformulation; repair; second language proficiency

Despite extensive, often decades-long speaking practice, adult production of speech, especially when longer stretches of discourse are required, is far from a seamless exercise. The speech of both first language (L1) and (to a greater extent) second language (L2) users is punctuated with pauses, fillers, hesitations, repetitions, reformulations, vowel prolongations, and errors (or slips of the tongue in the case of L1 speech; Brennan & Schober, 2001; Fox Tree, 1995). Yet, while disfluencies and deviations from the original speech plan may be cumbersome to speakers and listeners alike, to psycholinguists such verbal aberrations are a useful source of information. Each reflects disruption to a speech production process, and each can thus be used to explicate models of normal speech processing. This research investigates how the fluidity of L2 learners' speech is affected by their L2 proficiency level and how it

Table 1. Definitions and examples of reformulation types

Term	Definition	Example ¹
False start	the abandonment of an utterance followed by its immediate revision with the intention of improving coherence	every moment i(s) yeah ... the perception of every moment is individual
Correction	the attempted replacement of perceived nonstandard output (e.g., of syntax, lexis, or pronunciation) with a form that a fluent speaker would recognize as standard	it was happened er ... i ... it happened in my bathroom

depends on the type of reformulation used. Specifically, it reports on a quantitative analysis of two types of disfluency, reformulations and pauses, in the speech of lower intermediate and advanced speakers of English as an L2. It focuses on the extent to which pause behavior within reformulations varies according to the stage of L2 development and the type of reformulation used.

Reformulations in this paper are understood as a form of speaker self-repair in real time, and we adapt the coding scheme used by Levelt (1983) to distinguish between two subtypes that Levelt called appropriacy repairs and error repairs, respectively, but that in this paper, for the sake of clarity, we call false starts and corrections. We define a false start as the abandonment of a linguistically standard utterance followed by its immediate revision, and a correction as the attempted replacement of perceived nonstandard output (e.g., of syntax, lexis, or pronunciation) with a form that a fluent speaker would recognize as standard (examples of each type are provided in Table 1). Pause behavior within reformulations refers to the frequency and length of filled or silent hesitations that speakers insert in false starts and corrections. Because the literature identifies different functions for silent and filled pauses (Corley & Stewart, 2008), in this study we treat them separately. Here, silent pauses are defined as stretches of silence within the stream of speech with a duration of no less than 0.250 s after Cucchiarini, Strik, and Boves (2012). For example,

she quit and find another job [0.605] found another job.

Filled pauses are defined as lexical or nonlexical voiced utterances (e.g., *ehr*, *erm*, *like*, *actually*; Riggensbach, 1991) and prosodic markers like laughter and sighs that interrupt the stream of speech (Schuller et al., 2013). They are often preceded, and may be followed, by a silent pause (Beattie, 1977). In the transcription extracts, filled pauses are indicated by the type of filler in superscript after the time in seconds:

and it's quite [0.298^{yeah}] it was quite

In their work on L2 fluency, a number of authors refer to Lennon's (1990) distinction between fluency in the broad sense of global speaking proficiency and the narrow sense of the ease and smoothness with which speech is delivered, a component of oral proficiency (Baker-Smemeo, Dewey, Bown, & Martinsen, 2014;

Bosker, Pinget, Quené, Sanders, & de Jong, 2012; Chambers, 1997; de Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2013; Derwing, Munro, Thomson, & Rossiter, 2009; Ginther, Dimova, & Yang, 2010; Housen & Kuiken, 2009; Kahng, 2014; Rossiter, 2009; Schmidt, 1992). It is the narrow sense of the term, operationalized as “the best use of time constraints when speaking” to produce “an uninterrupted stream of smooth and hesitation-free speech” (Tavakoli, 2011, p. 72) that is used here. The present study seeks to confirm whether two kinds of reformulation, false starts and corrections, together with the pauses they contain, depend on L2 proficiency level. Two subprocesses of speech production, pauses and reformulations, will thus be analyzed at two different levels of L2 proficiency. The next section reviews work on reformulations, pauses, and proficiency level.

Research background

Precisely what circumstances (delays in message formulation, problems with linguistic encoding, self-monitoring, speaker characteristics, and/or other) produce each and every disfluency type remains debatable (Fraundorf & Watson, 2013; Ginther et al., 2010; Schnadt & Corley, 2006), and as often in L2 research, concepts like disfluency and its exemplars, pauses and reformulations, were initially studied and developed by L1 researchers. Schnadt and Corley (2006), for example, reported an increase in vowel prolongations when L1 speakers were describing objects with low-frequency names, suggesting that lexical accessibility (and hence transient problems during lexical encoding) may lie at the root of these production delays. Reduced lexical access was similarly implicated as a potential explanation for increased pausing and reformulations in the description of objects with low name agreement (where the object is associated with multiple names as opposed to one dominant name; Hartsuiker & Notebaert, 2011). Reformulations (i.e., insertions, substitutions, and deletions), however, were not affected by lexical frequency of the to-be-described items, suggesting that this group of disfluencies represents different underlying production problems. Different patterns of disfluencies in the same study were observed when the number of paths in the visual network described by the speakers was manipulated. Filled pauses and reformulations increased in frequency when multiple paths were included in the visual network, indicating problems with utterance formulation rather than with lexical retrieval. Some of the inconsistencies in the findings on the origin of disfluencies may stem from the fact that reformulations and pauses are often treated holistically as two superordinate categories with no distinction between reformulation or pause subtypes. In contrast, the present study distinguishes between corrections and false starts within the category of reformulations as well as between silent and filled pauses.

Reformulations

In his seminal paper on L1 monitoring and self-repair, Levelt (1983) understood reformulations as repairs, reporting a preponderance of corrections (“error repairs”; 42%) over false starts (“appropriacy repairs”; 30%), the two most ubiquitous repair forms of five described in his data. Levelt’s experimental prompts comprised static

visual networks of lines and colored dots presented on a computer screen (for a description of similar materials, see Declerck & Kormos, 2012; Martin, Weisberg, & Saffran, 1989; Oomen & Postma, 2001, 2002), enabling him, for example, to distinguish between a “difference” repair and an “appropriacy” repair:

We gaan rechtdoor offe . . . We kommen binnen via rood, gaan dan rechtdoor naar groen

We go straight on or . . . We come in via red, go then straight on to green.
 (“difference” repair, Levelt, 1983, p. 51)

We beginnen in het midden met . . . in het midden van het papier met een blauw rondje

We start in the middle with . . . in the middle of the paper with a blue disc.
 (“appropriacy” repair: ambiguity of reference, Levelt, 1983, p. 52)

In the former example, the speaker starts anew; in the latter, the speaker decides to qualify the content of the expression; in both cases, the speaker interrupts his or her speech and makes a change to the original utterance. However, without the evidence of such a highly structured elicitation task, it might be difficult to distinguish the two examples, or to apply some of Levelt’s other analytic categories with any certainty (Blackmer & Mitton, 1991).

Following de Bot’s (1992) adaptation of Levelt’s (1989) speech production model to describe bilingual speech production, researchers such as Brédart (1991) and Kormos (e.g., 1998, 1999, 2000b, 2006) applied Levelt’s (1999) repair taxonomy to the speech of L2 learners and extended it further. On that basis, a number of subsequent studies (Declerck & Kormos, 2012; Kormos, 1999; Van Hest, 1996) have reported statistically significant relationships between false starts and corrections (or their equivalent) and L2 proficiency level. Both O’Connor (1988) and Lennon (1990) had considered that the nature and location of self-repair might indicate the level of L2 performance. O’Connor (1988) found that her beginners’ reformulations more often comprised corrections compared to her advanced learners, who she claimed had more recourse to “anticipatory,” that is, discourse-related, insertions, such as false starts. Van Hest (1996) similarly reported that low-intermediate and intermediate speakers produced more corrections and fewer false starts than advanced speakers. Other studies that have investigated the influence of L2 proficiency on reformulations suggest that advanced learners become more attentive to discourse-level problems than surface errors such as lexis, grammar, and phonology (e.g., Kormos, 1998; Kovac & Milatovic, 2012; O’Connor, 1988; Van Hest, 1996), and that advanced speakers’ surface errors might be resolved or avoided by recourse to discourse strategies before false starts occur. As implied by the association of Levelt’s repair categories with the guided nature of his elicitation task, L2 studies report significant variation in the relative occurrence of speakers’ repair categories in response to different tasks (e.g., Foster & Skehan, 1996; Kormos, 2000a; Skehan & Foster, 1996; Van Hest, 1996). Unstructured or semistructured prompts (e.g., informal interviews), associated with

greater cognitive demand on conceptual processing, have elicited more false starts than corrections (Ahmadian, Abdolrezapour, & Ketabi, 2012; Van Hest, 1996). Structured prompts requiring more precise expression, including storytelling tasks and visual networks (e.g., Ahmadian et al., 2012; Kormos, 2000a; O'Connor, 1988), have elicited more corrections than false starts. The elicitation technique therefore needs to be taken into account when interpreting the results of such fluency studies.

Pauses

Another manifestation of speakers' processing difficulties during spoken language production may be pausing. Goldman-Eisler (1968) was the first to imply an interrelation between speed and breakdown variables in L1. She claimed that fluency could be measured in the form of speech rate differences, with the length and frequency of pauses showing an inverse relation to fluency (Goldman-Eisler, 1968; Griffiths, 1991). In a comparative study of L1 and L2 Russian speakers of English, Riazantseva (2001) found that L1 pause *duration* was carried over to the L2 regardless of proficiency level, whereas pause *frequency* and *distribution* were associated with proficiency level. De Jong (2016) confirmed that only silent pauses *within* speech units correlate with L2 proficiency level: the higher the L2 proficiency level, the shorter the pause, and within speech units, the *fewer* filled pauses. Kahng (2014) reported a weak but negative correlation between L2 speaking scores and the occurrence of *silent* pauses and a weak but positive correlation between the spoken performance and the occurrence of *filled* pauses. She also reported a weak negative correlation between spoken L2 performance and mean *length* of silent pause, but Kahng found no significant correlation between L2 speaker scores and filled pause *duration*.

Although researchers generally agree that the duration and frequency of silent pauses are affected by the person's mastery of the language, the inferences about their underlying causes are often very different. Like reformulation itself, pause length is thought to reflect the monitoring process (Kormos, 2000b), an important metacognitive process thought to make learners aware of system "gaps" in their interlanguage and thus lead to L2 development (Izumi, 2003; Kormos, 1999, 2006; Swain, 1985). By investigating monitoring mechanisms and associated speech markers, researchers interested in L2 acquisition also gain information on the way learners, when producing speech, allocate their attention to particular linguistic features at the various developmental stages (Kormos, 2006). For example, increased frequency of unpredictable pauses (nonjuncture pauses) in the performance of lower proficiency speakers is thought to stem from planning and execution problems (Cenoz, 1998). In dialogues, pauses of .5 s and longer may arise when the speaker is planning or preparing to reconceptualize an utterance, or analyzing what has been said by an interlocutor (Riggenbach, 1991). As reported, pause behavior may in addition be affected by the speaker's L1 pausing patterns (Leal, 1995; Riazantseva, 2001), which further complicates the interpretation of its origins and possible functions.

Pauses within reformulations

While the evidence associating language proficiency with speakers' production of reformulations (Bosker et al., 2012; Declerck & Kormos, 2012; Green & Hecht, 1993; Kormos, 1998, 1999, 2000a; Lennon, 1990; O'Connor, 1988; Riggenbach, 1991; Van Hest, 1996) and pauses (Declerck & Kormos, 2012; de Jong, 2016; Riazantseva, 2001; Riggenbach, 1991; Tavakoli, 2011) is widely available, the reports on the interaction of these two disfluency types with L2 proficiency are only suggestive at this point. Tavakoli (2011) reports that L2 English speaker pauses are often associated with reformulations and, noting the interaction of breakdown and repair within them, suggests further research. There have been other calls for a better understanding of the relationship of reformulations (their number, location, and structure) and general speech performance through the study of pause behavior (Kormos, 1999). In her discourse analysis study, Ejzenberg (2000) associates lower proficiency L2 speakers with more pausing, and she reports that the corrections and false starts of those speakers resulted in intraclausal repetition that sounded like "debilitating hesitation" (p. 302). In her analysis of the speech of fluent and non-fluent groups of speakers, Riggenbach (1991) notes that silent pauses are mostly found within clusters of disfluencies that comprise repetitions and "restarts" (i.e., false starts) as repair: nonfluent speakers produced relatively more such clusters than fluent speakers, although it should be noted that Riggenbach's observations were based on an extremely small sample size, comparing three intermediate and three advanced speakers of English as an L2.

Finally, in a section headed "Reformulation pauses," Tavakoli (2011) refers to the importance of the compound occurrence of pauses and other forms of disfluency such as repairs, including false starts. She notes that such symptoms of disfluency are mutually interactive, and reports planning taking place during some of the pauses that occur before the start of a reformulation. However, although she gives results for the clause position of pauses, Tavakoli stops short of considering the notion of pauses *within* reformulations.

Again, there have been several studies into the duration of reformulations themselves (Plug & Carter, 2014; Van Hest, 1996). For example, Van Hest (1996) reports that false starts take significantly longer to produce than corrections, a finding confirmed by Kormos (2000b). This relationship remained the case when comparing L1 and L2 repairs (all proficiency levels): when participants spoke in their L2, the duration of their false starts was longer than when they spoke in their L1. However, although the duration of repairs has been measured, and pauses noted to occur at the start of reformulations, researchers have not analyzed the direct relationship of pauses and reformulations, nor their dependence on fluency levels. Speakers' production of pauses inside reformulations may be another possible indicator of the relationship of reformulations to proficiency level. To our knowledge, therefore, despite obvious interest, this is the first study to quantitatively examine pauses produced inside two reformulation types and to do so with reference to proficiency level. Accordingly, the following research questions were formulated:

1. Does pause behavior within reformulations depend on the kind of reformulation used?

2. Does pause behavior within reformulations depend on L2 proficiency level?
3. Does pause behavior depend on the interaction between reformulation type and L2 proficiency?

Based on previous studies, we predicted that lower L2 proficiency speakers' reformulations would contain *silent* pauses that were longer and more frequent than those inserted into reformulations by higher proficiency learners, but that the frequency and duration of *filled* pauses within reformulations would be comparable across the two proficiency groups. We expected corrections generally to contain fewer and shorter silent pauses than false starts as they would be associated with repairs of surface errors, an ability that appears to develop before discourse-level skills. Assuming false starts and corrections reflect distinct cognitive mechanisms, with false starts reflecting nonlinguistic processes and corrections manifesting language processing difficulties, we expected to observe an interaction between reformulation types and L2 proficiency level, where lower and higher proficiency speakers would not differ on the duration or frequency of silent pauses within false starts, but higher proficiency speakers would produce fewer and shorter pauses within corrections.

Method

Participants and settings

The present investigation comprised 56 L2 speakers of English with a variety of different L1s (see Table 2). The data set formed part of a larger study involving 82 participants (Korko & Williams, 2017). All the participants were studying, or preparing to study, postgraduate courses at a British university and were following general English as a foreign language (EFL) classes at the same university. Their ages ranged from 19 to 45 years ($M = 26.25$). Based on assessment by two experienced EFL teachers, as described in the next section, the participants for the present study were classified as lower intermediate or advanced speakers. We excluded from the study the remaining 26 participants with middle values as inclusion of their data could mask patterns of significance. The masking may be explained by the wider variation in acquisition of different discourse features at different rates at the intermediate level (and cf. Karmiloff-Smith et al., 1993, for similar findings in relation to L1 discourse strategies). An independent samples *t* test showed that the L2 proficiency level of the two groups ($N_{low} = 25$, $N_{high} = 31$) was significantly different, $M_{low} = 5.17$, $SD_{low} = 0.17$, $M_{high} = 7.05$, $SD_{high} = 0.64$, $t(54) = 14.2$ $p < .001$.

Materials and procedure

A productive task eliciting 2 min of continuous L2 speech asked participants to describe familiar topics such as a memorable journey or a favorite TV program (an example is given in Appendix A). One of six such semistructured prompts, based on Allen, Powell, and Dolby (2007) and Hashemi and Thomas (2011), was randomly selected and presented to each participant. The speech samples were transcribed and coded for the disfluency types of interest.

Table 2. Frequencies of reported first languages

Language	Frequency	Total
German, Mandarin	13	26
Arabic	6	6
Spanish, Thai	4	8
Japanese	3	3
Bengali, French, Kurdish	2	6
Cantonese, Esan, Farsi, Gujerati, Russian, Turkish, Twi	1	7
All		56

In order to assess proficiency level, two experienced teachers of EFL rated every speech sample applying the public version of the globally recognized International English Language Testing Service (IELTS) examination Speaking Band Descriptors (http://www.ielts.org/pdf/UOBDs_SpeakingFinal.pdf), awarding a score 0–9 for each of four criteria: fluency and coherence, lexical resource, grammatical range and accuracy, and pronunciation. Interrater reliabilities reported as intraclass correlation coefficients (ICCs) and calculated using the two-way mixed, absolute agreement model were all in the acceptable range: fluency and coherence, ICC (3,2) = .796; lexical resource, ICC (3,2) = .811; grammatical range and accuracy, ICC (3,2) = .78; and pronunciation, ICC (3,2) = .842, all $ps < .001$. A composite of each of these scores, whose mean ranged from 4.75 to 9.0, was used to assign participants to the lower intermediate or advanced groups (scores 4.0–5.25 and 6.5–9.0, respectively), corresponding to Council of Europe Framework of Reference for languages levels B2 (independent user) and below, or C1 (proficient user) and above (Council of Europe, 2001). Intermediate level speakers, that is, those whose performance was rated above 5.25 and below 6.5, describing the “competent” (see <http://takeielts.britishcouncil.org/find-out-about-results/understand-your-ielts-scores>) but average speaker, were excluded from analysis in order to create groups of distinct proficiency levels (cf. Van Hest, 1996).

All reformulations in the L2 speech produced by the participants were identified from the transcripts by two independent raters, who, based on the definitions in Table 1, further divided the reformulations into false starts and corrections (Table 3). The coded variables, false starts and corrections, were assessed for interrater reliability by computing ICCs using the two-way mixed, absolute agreement model. The ICCs for false starts and corrections were in the acceptable range, with ICC (3,2) = .94, $p < .001$ and ICC (3,2) = .90, $p < .001$, respectively. The objects of the few disagreements that existed were reevaluated and resolved. One of the authors used Audacity 2.0.6 sound editor to manually calculate the length and frequency of silent pauses, mid- and end-clause, in the whole turn, adopting the minimum cutoff threshold of .25 s recommended by de Jong and Bosker (2013), whose study compared a number of pause thresholds used in monologic speaking tasks against an L2 proficiency measure. Each participant recording was uploaded

Table 3. Examples of reformulation coding with pause timings

Utterance	Reformulation instances	Reformulation type	Pause timing (seconds)
it's [0.862] I [0.310] went there with my parents	1	[False start]	0.862
and that's ... then we decided er [0.552]	3	[False start]	—
to on basing of ... [0.280] on this note		[1 st correction]	0.280
to create er maybe a small book ... [0.378] booklet		[2 nd correction]	0.378

Note: Pauses larger than .250 s given in brackets.

to Audacity as an MPEG Layer-3 audio file. After maximizing the waveform on the screen, and setting the counter to length, silent pauses were initially identified aurally and subsequently confirmed visually. The start and end of each pause was isolated and adjusted before recording the reading in the transcript and a separate Excel file. The pauses were then coded by T-unit (Lennon, 1990) as mid- or end-clause. For example,

ehr this June [0.546 mid] I wrote a note [0.324 mid] about erm [0.383 mid] Moscow [0.283 mid] nightlife and Moscow people [0.321 end]

A sample of 10 participants was randomly selected for a second rater to assess silent pause duration within the whole turn. This produced a total of 32 silent pauses that were reassessed. The interrater agreement on whole turn silent pauses was ICC (2,1) = .98, $p < .001$.

Following a similar procedure, the number and duration of both filled and silent pauses within reformulations, occurring after the abandonment of a problematic utterance or *reparandum* (Levelt, 1983, p. 45) and before completion of the revision, were manually quantified and analyzed by participant group. Pauses before the *reparandum* and after the repair were discounted, as were word onsets and other pauses considered too small to be timed. There was a good level of agreement between two independent raters on the occurrence of filled pauses in the L2 speech samples, with ICC (3,2) = .97, $p < .001$.

Results

Descriptive and inferential statistics for the occurrence of reformulations, silent pauses, and filled pauses per minute of speech across the two L2 proficiency groups is presented in Table 4. This is followed by analyses of the frequency and duration of silent and filled pause within the two reformulation types. Where the assumption of normality was violated and data were transformed, the type of function used for the transformation is stated in the text. Different transformation methods were applied that best served to normalize individual data sets. Where it was not possible to

Table 4. Means, standard deviations, and inferential statistics for various disfluency types per minute of speech of lower intermediate and advanced speakers of English as a second language

	Lower intermediate	Advanced	<i>F</i>	<i>p</i>	η^2
Reformulations combined (per min)	2.8 (1.6)	1.9 (1.2)	20.3	<.001	.27
False starts (per min)	1.6 (1.3)	1.56 (1.1)	0.26	.6	—
Corrections (per min)	1.2 (0.85)	0.53 (0.47)	15.4	<.001	.22
Silent pause duration (sec per min)	15 (5.9)	11.3 (4.3)	5.7	.021	.095
Silent pause frequency (per min)	24.3 (7)	18.5 (4)	15.2	<.001	.22
Filled pause frequency (per min)	17.8 (10.5)	10.8 (4.8)	9.3	.003	.15

Note: SDs are given in parentheses. See the Materials and Procedure section for details of measurement protocols.

normalize the data, untransformed scores were analyzed. In cases where the assumption of homogeneity of variance was not met, the obtained *p* value was compared to the adjusted *p* value of .01 instead of .05.

Occurrence of disfluencies per minute of speech in advanced and lower intermediate L2 speakers

Reformulation types per minute of speech

Reformulation data were transformed using the inverse function. A two-way analysis of variance (ANOVA) showed a significant difference in the number of reformulation types produced by all the L2 speakers. On average, there were more false starts ($M = 1.6$, $SD = 1.2$) than corrections ($M = 0.83$, $SD = 0.74$), $F(1, 54) = 19.2$, $p < .001$, $\eta^2 = .26$. The frequency of reformulations also depended on the L2 proficiency level, $F(1, 54) = 20.3$, $p < .001$, $\eta^2 = .27$. Lower intermediate speakers reformulated their utterances more often per minute of speech ($M = 2.8$, $SD = 1.6$) than the advanced proficiency group ($M = 1.9$, $SD = 1.20$). The interaction between L2 proficiency level and reformulation type was nonsignificant, $F(1, 54) = 2.4$, $p = .13$.

Silent pauses per minute of speech

A one-way ANOVA showed that the speech of lower intermediate speakers contained significantly more silent pauses per minute of speech ($M = 24.3$, $SD = 7$) than that of the advanced L2 learners ($M = 18.5$, $SD = 4$), $F(1, 55) = 15.2$, $p < .001$, $\eta^2 = .22$. The total duration of silent pauses per minute of speech by lower intermediate speakers was longer ($M = 15$ s, $SD = 5.9$ s) than that of silent pauses produced by the advanced proficiency group ($M = 11.3$ s, $SD = 4.3$ s) although this difference was nonsignificant after adjusting the *p* value, $F(1, 55) = 5.7$, $p = .021$, $\eta^2 = .095$.

Filled pauses per minute of speech

Filled pause data were transformed using the square root function. A one-way ANOVA showed that the lower L2 proficiency group used nearly twice as many fillers per minute of speech ($M = 17.8$, $SD = 10.5$) as the higher L2 proficiency group ($M = 10.8$, $SD = 4.8$), $F(1, 55) = 9.3$, $p = .003$, $\eta^2 = .15$.

Occurrence and duration of silent and filled pauses within reformulations

A series of 2 (L2 proficiency level: lower intermediate vs. advanced) x 2 (reformulation type: false start vs. correction) mixed-way ANOVAs was conducted to measure the effect of L2 proficiency and reformulation type as well as their interaction on the duration and frequency of silent and filled pauses inserted within reformulations.

Silent pause duration within reformulations

There was a main effect of reformulation type on silent pause duration, $F(1, 54) = 10.83, p = .002, \eta^2 = .17$. Silent pauses inserted by all the L2 speakers, regardless of their L2 proficiency, were generally much longer in false starts ($M = 0.813$ s, $SD = 1.1$ s) than in corrections ($M = 0.303$ s, $SD = 0.662$ s). The two groups differed in the total silent pause duration inserted within all the reformulations, with higher L2 proficiency speakers pausing for an average 0.4 s ($SD = 0.6$ s), and lower L2 proficiency speakers pausing for almost twice as long ($M = 0.767$ s, $SD = 1.1$ s) when reformulating. After adjusting the p value, however, this difference was nonsignificant, $F(1, 54) = 3.98, p = .051, \eta^2 = .069$. The interaction between reformulation type and L2 proficiency level was nonsignificant, $F(1, 54) = 0.61, p = .44$.

Silent pause frequency within reformulations

The silent pause frequency data within reformulations were transformed using the common logarithm function. There was a main effect of reformulation type on the frequency of silent pauses inserted by all the L2 speakers, $F(1, 54) = 20.3, p < .001, \eta^2 = .27$. False starts contained nearly twice as many silent pauses ($M = 3, SD = 2$) as corrections ($M = 1.66, SD = 1.77$). There was also a main effect of L2 proficiency on the frequency of silent pauses inserted in all the reformulations, $F(1, 54) = 7.4, p = .009, \eta^2 = .12$. Higher L2 proficiency speakers inserted fewer silent pauses within all the reformulations ($M = 1.97, SD = 1.6$) than lower L2 proficiency speakers ($M = 2.8, SD = 2$). There was a significant interaction between reformulation type and L2 proficiency, $F(1, 54) = 4.9, p = .031, \eta^2 = .08$. While silent pause frequency within false starts was comparable for the two L2 proficiency groups ($M_{low} = 3.1, SD_{low} = 2.1$ and $M_{high} = 2.9, SD_{high} = 2$), $F(1, 54) = 0.16, p = .7$, where the two groups seemed to differ for on silent pause frequency within corrections ($M_{low} = 2.5, SD_{low} = 2$ and $M_{high} = 1, SD_{high} = 1.3$), $F(1, 54) = 11.2, p = .002, \eta^2 = .17$ (Figure 1).

Filled pause duration within reformulations

Filled pauses inserted by all the L2 speakers in corrections were on average 175 ms long ($SD = 370$ ms), while those inserted in false starts were over twice as long ($M = 360$ ms, $SD = 440$ ms). This difference was not significant, however, after adjusting the p value to .01, $F(1, 54) = 6.9, p = .011, \eta^2 = .11$. There was a main effect of L2 proficiency on the duration of filled pauses inserted within both types of reformulation, $F(1, 54) = 10.8, p = .002, \eta^2 = .17$. Filled pauses inserted within both types of reformulation by lower intermediate speakers were nearly 400 ms long ($SD = 430$ ms), whereas those made by the advanced proficiency group took

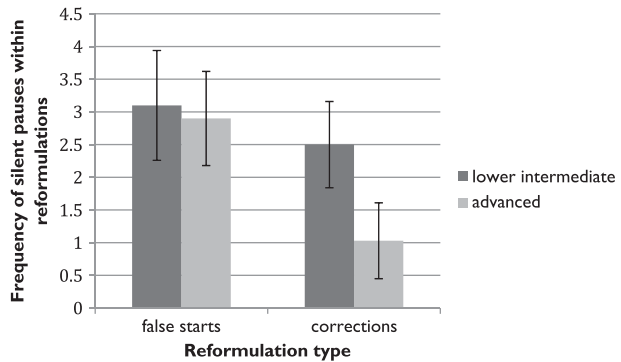


Figure 1. Frequency of silent pauses within reformulations by lower intermediate and advanced speakers of English as a second language (with 95% confidence interval bars).

on average 150 ms ($SD = 330$ ms). There was no interaction between reformulation type and L2 proficiency level, $F(1, 54) = 0.001$, $p = .97$.

Filled pause frequency within reformulations

There was a main effect of reformulation type on filled pause frequency within reformulations. False starts contained on average more filled pauses ($M = 0.75$, $SD = 0.92$) than did corrections ($M = 0.45$, $SD = 0.93$), but after adjusting the p value, this difference was not significant, $F(1, 54) = 3.6$, $p = .061$, $\eta^2 = .063$. There was a main effect of L2 proficiency on the number of filled pauses inserted within reformulations, $F(1, 54) = 14.5$, $p < .001$, $\eta^2 = .21$, with lower intermediate speakers using three times as many fillers ($M = 0.98$, $SD = 1.1$) as advanced speakers ($M = 0.31$, $SD = 0.58$). The interaction between reformulation type and L2 proficiency was nonsignificant, $F(1, 54) = 0.004$, $p = .95$.

Discussion

This section first discusses general findings concerning the speakers' pause behavior and reformulations during speakers' whole turns. It goes on to address the research questions relating to pause behavior within the reformulations, namely, whether that behavior depends on the nature of the reformulation, the proficiency level, or the interaction between the type of reformulation and L2 proficiency.

Regarding the types of reformulations, contrary to Levelt (1983), we noted more false starts than corrections per minute of speech in the speakers' whole turns. Such a discrepancy could be explained by the nature of the elicitation task and the language under investigation. In Levelt's study, participants were native speakers of Dutch who were required to describe visual networks. Consequently, much of the participants' verbal output was experimentally controlled. The present study looked at spontaneous speech produced by speakers of English as an L2. Not only did it measure the less dominant language (L2), but the speaking conditions imposed fewer production constraints. When the results of this study were compared to those reported by studies using similar elicitation tasks, better consistency was obtained. Ahmadian et al. (2012), Kormos (2000a), O'Connor (1988), and Van Hest (1996), whose elicitation tasks were less structured, comprising an oral narrative task

(“loose” plotline), role play, interview, and storytelling/interview, respectively, similarly found larger numbers of false starts than corrections.

More importantly, the occurrence of false starts per minute of speech was comparable across the two proficiency groups; however, lower intermediate speakers produced on average twice as many corrections per minute of speech as advanced speakers. These results differ from previous work (O’Connor, 1988; Van Hest, 1996) that found advanced speakers produced more false starts. Advanced speakers might be expected to produce more false starts than corrections because they have more resources for managing discourse and require fewer for monitoring accuracy; that they did not produce more false starts in the present study might suggest that these advanced speakers were better able to plan online and preempt a certain number of them. Alternatively, false starts may depend not on proficiency level but reflect conceptual processing or other processes, such as an ability to adapt the message to the communicative situation, which occur outside the language system per se.

The higher frequency of corrections in the speech of lower intermediate learners is understandable. On the one hand, advanced speakers self-correct less as they tend to make fewer mistakes than lower proficiency speakers. We found that lower intermediate speakers produced on average nearly three times more speech errors ($M_{low} = 8.3\%$, $SD = 2.9\%$) compared to the advanced group ($M = 3.2\%$, $SD = 2.8\%$), $F(1, 54) = 42.1$, $p < .001$. In addition, there was a significant strong correlation between the total proportion of speech errors in speakers’ complete turns and their corrections ($r = .541$, $p < .001$), so this could be a likely explanation for the higher proportion of corrections made by the lower proficiency group. On the other hand, advanced L2 speakers may have sufficient cognitive resources to spot and intercept their mistakes in the form of covert repairs before actually uttering them (Levelt, 1989), which equally contributes to a lower proportion of overt corrections at this level of proficiency.

Turning to pause behavior, the advanced speakers produced fewer silent pauses per minute of speech than lower intermediate speakers. This finding is consistent with Riegenbach (1991), who found the number of silent pauses greater in lower proficiency speakers, and Kahng (2014), who found the number of silent pauses to be inversely related to L1 and L2 proficiency in Korean speakers of English. In the present study, the average silent pause *duration* per minute of speech did not differ across the two proficiency groups. This contrasts with Kahng (2014), who found in her study of L1 speakers of English and L2 Korean learners of English the *duration* as well as the *number* of silent pauses to be inversely related to L1 and L2 proficiency. However, Kahng’s inclusion of L1 English speakers in her study might explain the discrepancy. It could also be argued that the frequency of silent pauses is a better indicator of L2 proficiency than duration of silent pauses as inserting too many pauses, especially nonjuncture pauses, may give the impression of fragmented, unnaturally flowing speech. In contrast, infrequent but longer silent pauses may reflect conceptual processing.

When the two groups’ use of filled pauses per minute of speech was compared, lower intermediate speakers inserted nearly twice as many as advanced speakers. In contrast, Kahng (2014) found none of her filled pause measures correlated with the speaking scores of her L2 speakers of English. Similarly, no significant difference in measures of frequency of filled pauses were found in advanced versus intermediate

proficiency L2 speakers by Declerck and Kormos (2012). Fillers were associated only with conceptual difficulties, consistent with the proposal that they reflect a communicative signal, whereas silent pauses and repeats were also related to lexical and phonological difficulties (Fraundorf & Watson, 2013).

More pertinent to our research was the question of how pause behavior differs with regard to L2 proficiency and the type of reformulation used. The frequency and duration of silent and filled pauses within false starts and corrections may help to explain the origin of these disfluencies as well as indicate which type of disfluency is a useful means of discriminating between proficiency levels.

RQ1: Does pause behavior within reformulations depend on the kind of reformulation used?

There were significant differences in both the *occurrence* and *duration* of silent pauses within false starts and corrections. False starts contained nearly twice as many silent pauses as occurred in corrections, and the silent pauses in false starts were also longer compared to pauses in corrections. Compared to corrections, false starts also contained more filled pauses, but the duration of filled pauses in false starts and corrections was about the same. These findings are perhaps consistent with the greater total repair time of false starts over corrections reported by Kormos (2000b), whose study found corrections were significantly shorter. This finding allows the possibility of group and individual variation in pause production as false starts are more likely than corrections to be idiosyncratic in form and less dependent on the mastery of standard forms (see definitions in Table 1). A greater number of silent and filled pauses in false starts relative to corrections suggests that this type of disfluency may be associated with greater cognitive demand, which is consistent with Ahmadian et al. (2012) and Van Hest (1996). If false starts are taken to reflect conceptual encoding (e.g., the speaker plans an utterance, decides on the register, and/or tailors her message to the communicative needs of the listener), then, based on the current data, it could be argued that conceptual encoding is generally more cognitively taxing than correcting one's own speech errors.

RQ2: Does pause behavior within reformulations depend on L2 proficiency level?

Speakers' pause behavior within reformulations shows differences by proficiency level, though not for all the pause parameters used in this study. There was no difference in the duration of silent pauses within reformulations between the two groups, but lower proficiency speakers' reformulations were marked by a greater proportion of silent pauses. Advanced speakers might be expected to produce more and longer filled pauses than low-proficiency speakers, signalling their searches for less common vocabulary. Moreover, it was the lower intermediate rather than the advanced speakers who produced more frequent and longer filled pauses within reformulations (combined). Perhaps they are working harder to maintain their speaking turn to compensate for the comparative paucity of precision in their message content.

RQ3: Does pause behavior depend on the interaction between reformulation type and L2 proficiency?

There was a significant interaction between L2 proficiency level and reformulation type in the frequency of silent pauses. Specifically, the two proficiency groups did not differ in the frequency of silent pauses within false starts (the means are almost the same), but they did within corrections, with lower intermediate speakers producing more frequent silent pauses within that reformulation type. From this, it is possible to infer that false starts are not dependent on proficiency level because both advanced and lower intermediate speakers insert about the same number of pauses within them. This may provide further evidence that false starts are discourse related, having regard to the speaker's concern with the quality of the message or interaction. For example, Giles, Taylor, and Bourhis (1973) found that speakers who adapted their speech to match their interlocutors, including "many filled and unfilled pauses [and] speech disturbances" (p. 181), were thought of more favorably by the latter. This would be an example of context adaptation, that is, an ability to take the perspective of the listener into account to establish common ground. Speakers may thus qualify their message to satisfy the communicative needs of their interlocutor (see also Garrod & Anderson, 1987). In contrast, as noted earlier, advanced learners in particular are reported to become more attentive to discourse-level problems than surface errors such as lexis, grammar, and phonology (e.g., Kormos, 1998; Kovac & Milatovic, 2012; O'Connor, 1988; Van Hest, 1996). In addition, apart from the suggestion that they might have recourse to discourse strategies in order to avoid correction, it is possible that the discourse strategies themselves take the form of false starts. It might be supposed that low-proficiency false starts would be more fluent than low-proficiency corrections, and associated pauses *shorter*, as speakers have more flexibility as to the conceptual content of false starts and are less driven to search for particular standard forms, but their pauses would be *more frequent* as the speakers reconceptualize and select from a variety of syntagmatic and paradigmatic options. Yet, the low-proficiency speakers' silent pauses within false starts were neither significantly more frequent nor shorter than in their corrections.

Corrections, however, may depend on proficiency level because we know that lower intermediate speakers pause more frequently when self-correcting, possibly because they are not sure whether or how to correct an erroneous structure. Lower intermediate speakers inserted on average nearly twice as many silent pauses within corrections as the advanced group. Pause behavior within corrections is therefore clearly distinguished by L2 proficiency level, while the silent pause duration is not. Silent pause frequency can thus be argued to be a better marker of reformulation types across different L2 proficiency groups.

In both kinds of reformulation, advanced speakers' pauses would be predicted to be shorter and less frequent as a result of greater automaticity in standard forms, a wider repertoire of "formulaic sequences" (Tavakoli, 2011), and more working memory capacity. However, advanced speakers' pauses were shorter and less frequent only in corrections and not in both types of reformulation. The greater number and duration of lower intermediate speaker breakdowns in corrections most likely reflects their relative lack of linguistic and cognitive resources to adopt alternative strategies

at these points. Finally, lower intermediate speakers produced more filled pauses than advanced speakers within both reformulation types combined.

Consistent with the literature, analysis of reformulations as false starts and corrections in these data has shown differences by proficiency level. This suggests that false starts may not depend so much on proficiency level or how much exposure speakers have had to the L2, and by inference do not necessarily reflect temporary difficulties in finding words, constructing an appropriate grammatical structure, or selecting a relevant phonological segment (which is related to language proficiency), but are possibly an outcome of extralinguistic abilities, such as preverbal message planning, an ability to adapt to the communicative context, or perhaps proneness to intrusions (false starts could be interpreted as temporary inattentiveness). This trend cannot be observed with filled pauses, but their production may be confounded by cultural or individual differences (Igras-Cybulska, Ziolk, Zelasko, & Witkowski, 2016).

Conclusion


It may be self-evident that speakers, in both L1 and L2, “monitor what they are saying and how they are saying it” (Levelt, 1989, p. 458), resulting in a self-initiated and self-completed revision when the speaker notices that his or her output includes erroneous or infelicitous language content, interrupts the speech flow, often by means of a pause, and finally accomplishes a repair (Kormos, 2006; Levelt, 1989). However, the results indicate that the two types of reformulation, false start and correction, and their pause patterns, are probably separate categories, underpinned by different processes. This explanation would contradict Goldman-Eisler’s (1968) claim that speakers conceptualize during silent pauses and not during articulation. The fewer silent pauses the advanced speakers produced within reformulations provides supporting evidence for this claim, consistent with Gilabert (2007).

The reformulation and pause characteristics of advanced speakers (i.e., that they reformulate and pause less) indicate that, probably as a result of automaticity, they have greater attentional capacity to monitor their production. That they produce fewer silent pauses and fewer filled pauses per minute of speech, and fewer silent pauses within reformulations, suggests that they have more efficient access to lemmas and syntax coding, and are better able to (re-)conceptualize their thoughts prior to speaking. The fact that this pause behavior is true of their complete speech sample, as well as their reformulations, bears this out. The difference in the reformulation and pause behavior of the two proficiency levels may be conceptualized as the tension between psycholinguistic (the need to accurately express one’s thoughts) and sociolinguistic (the need to produce linguistic forms that are understood) imperatives (see Ejzenberg, 2000). Both groups prioritize self-expression, but the lower intermediate group appears to expend more cognitive resources on producing standard forms, and signals this with a higher proportion of filled and silent pauses and greater duration of silent pauses.

Particular features of the task set added to the cognitive pressure on the speakers, regardless of their proficiency level. These features included the monologic nature of the task, the time constraint, the implied requirement for extended planning, and

possibly some narrative content. However, because advanced speakers are able to access lemmas and self-monitor more efficiently, thanks to greater processing capacity, they have less need to reformulate their message, and when they do, they produce fewer pauses as they order their thoughts, access lemmas, and parse syntax.

Limitations of the present study include the range of first languages, which reflects the choice of a convenience sample but could be seen as a major weakness. On the other hand, it was felt that if statistically significant associations were found in this data, then we could be more confident that they would apply across a broad range of language learners. We acknowledge that filled pauses at least may be idiosyncratic, and possibly culture related, and the study avoids making generalizations on their association with proficiency, although it does reprise some widely recognized observations regarding their function. It would be interesting to investigate whether silent pauses are used for covert editing, or whether covert editing and planning in advanced or lower intermediate learners takes place during articulation, while filled pauses in contrast are used for word searches. Further investigation might also address the area of individual differences in terms of how various disfluency types vary with the L2 speaker's working memory, attention, or perspective taking.

Author ORCID.  Simon A. Williams, 0000-0002-4807-9243

Acknowledgments. The authors would like to thank the editors and two anonymous reviewers for their helpful comments.

Note

1. These and other examples are taken from the data unless indicated otherwise.

References

- Ahmadian, M. J., Abdolrezapour, P., & Ketabi, S. (2012). Task difficulty and self-repair behavior in second language oral production. *International Journal of Applied Linguistics*, *22*, 310–330.
- Allen, M., Powell, D., & Dolby, D. (2007). *IELTS graduation*. Oxford, UK: Macmillan.
- Baker-Smemoe, W., Dewey, D. P., Bown, J., & Martinsen, R. A. (2014). Does measuring L2 utterance fluency equal measuring overall L2 proficiency? Evidence from five languages. *Foreign Language Annals*, *47*, 707–728.
- Beattie, G. (1977). The dynamics of interruption and the filled pause. *British Journal of Social and Clinical Psychology*, *16*, 283–284.
- Blackmer, E. R., & Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, *39*, 173–194.
- Bosker, H., Pinget, A., Quené, H., Sanders, T., & de Jong, N. (2012). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, *30*, 159–175.
- Brédart, S. (1991). Word interruption in self-repairing. *Journal of Psycholinguistic Research*, *20*, 123–137.
- Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language*, *44*, 274–296.
- Cenoz, J. (1998). *Pauses and communication strategies in second language speech* (ERIC Document ED 426630). Rockville, MD: Educational Resources Information Center.
- Chambers, F. (1997). What do we mean by fluency? *System*, *25*, 535–544.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge, UK: Cambridge University Press.

- Cucchiari, C., Strik, H., & Boves, L.** (2012). Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *Journal of the Acoustic Society of America*, **111**, 2862–2873.
- de Bot, K.** (1992). A bilingual production model: Levelt's "speaking" model adapted. *Applied Linguistics*, **13**, 1–24.
- Declerck, M., & Kormos, J.** (2012). The effect of dual task demands and proficiency on second language speech production. *Bilingualism: Language and Cognition*, **15**, 782–796.
- de Jong, N. H.** (2016). Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching*, **54**, 113–132.
- de Jong, N. H., & Bosker, H. R.** (2013). Choosing a threshold for silent pauses to measure second language fluency. In R. Eklund (Ed.), *The 6th Workshop on Disfluency in Spontaneous Speech (DiSS)* (pp. 17–20). Stockholm, Sweden: KTH Royal Institute of Technology.
- de Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, R.** (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, **34**, 893–916.
- Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J.** (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, **31**, 533–557.
- Ejzenberg, R.** (2000). The juggling act of oral fluency: A psycho-sociolinguistic metaphor. In H. Riggenbach (Ed.), *Perspectives on fluency*. Ann Arbor, MI: University of Michigan Press.
- Foster, P., & Skehan, P.** (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, **8**, 299–323.
- Fox Tree, J. E.** (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language*, **34**, 709–738.
- Fraundorf, S. H., & Watson, D. G.** (2013). Alice's adventures in um-derland: Psycholinguistic sources of variation in disfluency production. *Language Cognition Process*, **29**, 1083–1096.
- Garrod, S., & Anderson, A.** (1987). Saying what you mean in dialog: A study in conceptual and semantic co-ordination. *Cognition*, **27**, 181–218.
- Gilbert, R.** (2007). Effects of manipulating task complexity on self-repairs during L2 production. *IRAL*, **45**, 215–240.
- Giles, H., Taylor, D. M., & Bourhis, R.** (1973). Towards a theory of interpersonal accommodation through language: Some Canadian data. *Language in Society*, **2**, 177–192.
- Ginther, A., Dimova, S., & Yang, R.** (2010). Conceptual and empirical relationships between temporal measures of fluency and oral English proficiency with implications for automated scoring. *Language Testing*, **27**, 379–399.
- Goldman-Eisler, F.** (1968). *Psycholinguistics: Experiments in spontaneous speech*. London, UK: Academic Press.
- Green, P. S., & Hecht, K.** (1993). Pupil self-correction in oral communication in English as a foreign language. *System*, **21**, 151–163.
- Griffiths, R.** (1991). Pausological research in an L2 context: A rationale and review of selected studies. *Applied Linguistics*, **12**, 345–364.
- Hartsuiker, R. J., & Notebaert, L.** (2011). Lexical access problems lead to disfluencies in speech. *Experimental Psychology*, **57**, 169–177.
- Hashemi, L., & Thomas, B.** (2011). *IELTS Trainer*. Cambridge, UK: Cambridge University Press.
- Housen, A., & Kuiken, F.** (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, **30**, 461–473.
- Igras-Cybulska, M., Ziolko, B., Zelasko, P., & Witkowski, M.** (2016). Structure of pauses in speech in the context of speaker verification and classification of speech type. *EURASIP Journal on Audio, Speech and Music Processing*, **18**, 1–16.
- Izumi, S.** (2003). Comprehension and production processes in second language learning: In search of the psycholinguistic rationale for the output hypothesis. *Applied Linguistics*, **24**, 168–196.
- Kahng, J.** (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning*, **64**, 809–854.
- Karmiloff-Smith, A., Johnson, H., Grant, J., Jones, M.-C., Karmiloff, Y.-N., Bartrip, J., & Cuckle, P.** (1993). From sentential to discourse functions: Detection and explanation of speech repairs by children and adults. *Discourse Processes*, **16**, 565–589.

- Korko, M., & Williams, S. A.** (2017). Inhibitory control and the speech patterns of second language users. *British Journal of Psychology*, *108*, 43–72.
- Kormos, J.** (1998). A new psycholinguistic taxonomy of self-repairs in L2: A qualitative analysis with retrospection. *Even Yearbook, ELTE SEAS Working Papers in Linguistics*, *3*, 43–68.
- Kormos, J.** (1999). Monitoring and self-repair in L2. *Language Learning*, *49*, 303–342.
- Kormos, J.** (2000a). The role of attention in monitoring second language speech production. *Language Learning*, *50*, 343–384.
- Kormos, J.** (2000b). The timing of self-repairs in second language speech production. *Studies in Second Language Acquisition*, *22*, 145–169.
- Kormos, J.** (2006). *Speech Production and Second Language Acquisition*. London, UK: Erlbaum.
- Kovac, M. M., & Milatovic, B.** (2012). Analysis of repair distribution, error correction rates, and repair successfulness in L2. *Studia Linguistica*, *67*, 225–255.
- Leal, C. F.** (1995). *A pause as a cause of change*. ERIC document ED 379935. Rockville, MD: Educational Resources Information Center.
- Lennon, P.** (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, *40*, 387–417.
- Levelt, W. J. M.** (1983). Monitoring and self-repair in speech. *Cognition*, *14*, 41–104.
- Levelt, W. J. M.** (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M.** (1999). Models of word production. *Trends in Cognitive Sciences*, *3*, 223–232.
- Martin, N., Weisberg, R. W., & Saffran, E. M.** (1989). Variables influencing the occurrence of naming errors: Implications for models of lexical retrieval. *Journal of Memory and Language*, *28*, 462–485.
- O'Connor, N.** (1988). Repairs as indicative of interlanguage variation and change. In T. J. Walsh (Ed.), *Georgetown University Round Table in Languages and Linguistics 1988: Synchronic and diachronic approaches to linguistic variation and change*. Washington, DC: Georgetown University Press.
- Oomen, C. C. E., & Postma, A.** (2001). Effects of time pressure on mechanisms of speech production and self-monitoring. *Journal of Psycholinguistic Research*, *30*, 163–184.
- Oomen, C. C. E., & Postma, A.** (2002). Limitations in processing resources and speech monitoring. *Language and Cognitive Processes*, *17*, 163–184.
- Plug, L., & Carter, P.** (2014). Timing and tempo in spontaneous phonological error repair. *Journal of Phonetics*, *45*, 52–63.
- Riazantseva, A.** (2001). Second language proficiency and pausing: A study of Russian speakers of English. *Studies in Second Language Acquisition*, *23*, 497–526.
- Riggenbach, H.** (1991). Toward an understanding of fluency: A microanalysis of non-native speaker conversations. *Discourse Processes*, *14*, 423–441.
- Rossiter, M. J.** (2009). Perceptions of L2 fluency by native and non-native speakers of English. *Canadian Modern Language Review*, *65*, 395–412.
- Schmidt, R.** (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, *14*, 357–385.
- Schnadt, M. J., & Corley, M.** (2006). The influence of lexical, conceptual and planning based factors on disfluency production. *Proceedings of the twenty-eighth meeting of the Cognitive Science Society [CD-ROM]*. Mahwah, NJ: Erlbaum.
- Schuller, B., Steidl, S., Batliner, A., Burkhardt, F., Devillers, L., Müller, C., & Narayanan, S.** (2013). Paralinguistics in speech and language—State-of-the-art and the challenge. *Computer Speech & Language*, *27*, 4–39.
- Shehan, P., & Foster, P.** (1996). *Task type and task processing conditions as influences on foreign language performance. Paper presented at the annual conference of the American Association of Applied Linguistics*, Chicago.
- Swain, M.** (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition*. Rowley: Newbury House.
- Tavakoli, P.** (2011). Pausing patterns: Differences between L2 learners and native speakers. *ELT Journal*, *65*, 71–79.
- Van Hest, E.** (1996). *Self-repair in L1 and L2 production*. Tilburg, the Netherlands: Tilburg University Press.

APPENDIX A. Example of oral production task

Describe a TV program that made you laugh.
You should say:

- What the program was about
- When it was broadcast
- How often you watched it

And explain why you found the program funny.

Cite this article: Williams S.A. and Korko M. (2019). Pause behavior within reformulations and the proficiency level of second language learners of English. *Applied Psycholinguistics* **40**, 723–742. <https://doi.org/10.1017/S0142716418000802>