

Evaluating Measures of Pausing for Second Language Fluency Research

Christine Shea¹ and Karen Leonard²

Résumé : Dans les écrits sur la fluidité en L2, la meilleure façon de définir les pauses de manière opérationnelle et de déterminer quel est le lien entre les différentes mesures des pauses et la maîtrise de la L2 est abondamment débattue. Cette question interpelle les chercheurs qui s'intéressent à la fluidité en L2, en particulier ceux qui travaillent auprès de groupes dont le degré de maîtrise de la L2 varie. L'auteure se penche sur le sujet en étudiant les données relatives à la production orale de 48 apprenants de l'espagnol (dont la L1 est l'anglais), afin d'évaluer quelles mesures des pauses sont les meilleurs indicateurs de fluidité. Son analyse porte sur différentes mesures des pauses — par 100 syllabes ou par minute, selon la durée, la nature (remplies ou silencieuses) et l'emplacement (au milieu ou à la fin d'une proposition) — ainsi que sur leur relation avec la maîtrise de la L2. Le degré de maîtrise est opérationnalisé sous forme de score combiné en connaissances linguistiques obtenu dans des tests de grammaire et de vocabulaire. L'auteure procède à des analyses de corrélation et de régression afin de déterminer comment les différentes mesures des pauses rendent compte de la variabilité des scores quant à la maîtrise. Les résultats de l'étude indiquent que des pauses plus longues (une seconde ou plus) sont de meilleurs prédicteurs de la maîtrise de la L2 que les pauses plus courtes, que les pauses silencieuses sont un prédicteur plus fort de la maîtrise de la L2 que les pauses remplies, et que les pauses en milieu de proposition permettent d'établir plus clairement que les pauses en fin de proposition la distinction entre les niveaux de maîtrise. Les résultats de l'étude guideront les chercheurs dans le choix de la mesure des pauses qui répond le mieux à leurs besoins.

Mots clés : espagnol, fluidité, L2, mesure, pause

Abstract: In the L2 fluency literature there is considerable debate over how best to operationally define pauses and how different pause measures relate to L2 proficiency. This creates a challenge for researchers interested in L2 fluency, and particularly those who are working with groups that vary in L2 proficiency. This article addresses these issues by examining oral production data from 48 learners of Spanish (L1 English) to evaluate which pause measures best serve as markers of fluency. Different pause measures were analyzed – per 100 syllables/per minute, duration, type (filled/unfilled) and location (in the middle or at the end of a clause) – along with their relationship to L2 proficiency. Proficiency was operationalized as a composite language knowledge score from L2 grammar and vocabulary tests. Correlational and regression analyses were then carried out to determine how different pause measures account for variability in the proficiency

¹ Department of Spanish and Portuguese, University of Iowa

² Department of Modern Languages, Olivet Nazarene University

Correspondence should be addressed to Christine Shea, Department of Spanish and Portuguese, University of Iowa, 412 Phillips Hall, Iowa City, Iowa, IA 52245 USA; email: christine-shea@uiowa.edu.

scores. The findings show that longer pauses (of 1 second or longer) are better predictors of L2 proficiency than shorter pauses, unfilled pauses are a stronger predictor of L2 proficiency than filled pauses, and mid-clause pauses distinguish between proficiency levels more clearly than do end-of-clause pauses. The results will serve to guide researchers in determining which pause measure best corresponds to their needs.

Keywords: fluency, L2, measurement, pausing, Spanish

In the field of second language acquisition, fluency researchers have focused on how fluent speech sounds (the perspective of the listener); the acoustic, phonetic, and articulatory characteristics of fluency (the perspective of the speaker); and how speaker proficiency may interact with each of these factors and their respective subcomponents. Segalowitz's (2010) tripartite classification reflects this by distinguishing among three types of fluency: perceived (listener, subjective), utterance (speaker, objective) and the underlying processes responsible for the production of utterances – cognitive fluency. Strong relationships between utterance fluency and perceived fluency have been found (Derwing, Munro, Thomson, & Rossiter, 2009; Rossiter, 2009). In terms of utterance fluency and cognitive fluency, on the other hand, the relationship is more complex. As stated, cognitive fluency is assumed to underlie utterance fluency, since speakers cannot produce speech without having the lexical and syntactic knowledge to do so, independent of how listeners may perceive the actual productions themselves (Segalowitz & Freed, 2004; but see De Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2012, for an alternative account).

In this study, we present a close examination of pauses, one specific aspect of L2 fluency, and consider how pauses have been measured in learner speech, as well as how these measurements may be useful in distinguishing speakers with different proficiencies. Pauses fit into the utterance fluency component of Segalowitz's classification as part of "breakdown fluency." The other subcomponents of utterance fluency are speed (speech rate) and repair (false starts and repetitions).

Pauses – including length, type, and location – have been studied from a number of different perspectives, and it is recognized in the fluency literature that pauses play a key role in utterance and perceived fluency. For example, Bosker, Pinget, Quené, Sanders, and De Jong (2013) examined the three dimensions of utterance fluency in the speech of L2 Dutch speakers and how each dimension affected overall fluency judgments by native Dutch listeners. Their results showed that pause and speed measures significantly predicted subjective fluency ratings, but repair contributed very little.

In another study that focused more specifically on pauses and perceived fluency, De Jong and Bosker (2013) analyzed the optimal cut-off point for silent pauses in L2 Dutch speech and how pause thresholds relate to L2 vocabulary knowledge and overall fluency ratings by native-speaker judges. The authors measured pauses within Analysis of Speech Units (ASU), which can be understood as a full utterance with an independent clause and all other related dependent or subclausal units (Foster, Tonkyn, & Wigglesworth, 2000). They found low and non-significant correlations between the log duration of pauses and proficiency measures (L2 vocabulary knowledge); however, they did find significant correlations between the log duration of pauses and perceived fluency. They also found the highest correlations between vocabulary knowledge and the number of silent pauses using a threshold of 250 ms.

In a study focusing on the relationship between utterance and cognitive fluency, De Jong et al. (2012) found moderate relationships between pause frequency and measures

of cognitive fluency, and weak relationships between pause duration and measures of cognitive fluency. These findings were further examined by De Jong (2016), who showed that pauses in mid-ASU can be linked to points where the speaker is producing a less frequent word, in both the L1 *and* the L2, suggesting that lexical retrieval effects occur in the native language as well (Kirsner, Dunn, Hird, Parkin, & Clark, 2002).

Measures of pausing

Measures of pauses that have been used in the literature examining utterance fluency and, more specifically, breakdown fluency can be grouped into ratios (time of pausing in speech sample to time speaking), counts (number of pauses that occur in a given speech sample), simple duration, and location of pauses (within or across clause boundaries). Box 1 contains a summary of the different measures of pausing that have been used in previous studies; below, we present a brief discussion of how each pausing measure has been operationalized in the literature.

Box 1: Measures of pausing in previous studies

Total number of pauses

- 0.4 seconds or longer (Rossiter, 2009; Tavakoli & Skehan, 2005)
- 1 second or longer (Foster & Skehan, 1996; Mehnert, 1998; Skehan & Foster, 1997)

Pause-to-time ratio

- Filled pauses per second or per minute (Bosker et al., 2013; D'Amico, 2010; Iwashita et al., 2008; Kormos & Dénes, 2004; Llanes & Muñoz, 2009; Skehan & Foster, 2005)
- Unfilled pauses per second or per minute
 - 0.2 seconds or longer (Kormos & Dénes, 2004)
 - 0.25 seconds or longer (Bosker et al., 2013)
 - 0.4 seconds or longer (D'Amico, 2010; Llanes & Muñoz, 2009; Valls Ferrer, 2008)
 - 0.5 seconds or longer (Mota, 2003)
 - 1 second or longer (Iwashita et al., 2008)
- Hesitations per minute: unfilled pauses less than 0.5 seconds + filled pauses (Mota, 2003)
- All pauses per minute
 - 0.4 seconds or longer (Derwing et al., 2009)
 - 2 seconds or longer (Nakakubo, 2011; Tajima, 2003)

Pause-to-unit ratio

- Filled pauses per turn (García-Amaya, 2009)
- Filled pauses per 100 words (Michel, Kuiken, & Vedder, 2007)
- Pauses per 1,000 words (Hilton, 2008)
- Pauses per 100 syllables (Leonard & Shea, 2017)
- Filled pauses per T-unit (Lennon, 1990)
- Unfilled pauses per T-unit (Bygate, 2001)
- Unfilled pauses per c-unit^a (Guará Tavares, 2009)

Total pausing time

- Measured as percentage of total speaking time
 - Unfilled pauses (Iwashita et al., 2008)
 - Both filled and unfilled pauses (Lennon, 1990; Leonard & Shea, 2017; Mehnert, 1998; Sangarun, 2005)
 - Unspecified (Hilton, 2008)

(Continued)

- Measured in number of seconds or unspecified (Foster & Skehan, 1996; Tavakoli & Skehan, 2005)
- Phonation-time ratio: percentage of time speaking rather than pausing (Kormos & Dénes, 2004; Préfontaine, 2010)

Other

- Mid-clause vs. end-of-clause pauses (De Jong, 2016; Skehan & Foster, 2005; Tavakoli & Foster, 2011)
- Mean length of pauses
 - All pauses or all unfilled pauses (Bosker et al., 2013; Hilton, 2008; Kormos & Dénes, 2004; Tavakoli & Skehan, 2005; Towell, 2002; Valls Ferrer, 2008)
 - Internal pauses (Valls Ferrer, 2008)
 - Pauses at T-unit boundaries (Lennon, 1990)
 - Pauses within AS-units^b (De Jong et al., 2013)
 - Pauses between AS-units (De Jong et al., 2013)
- Percentage of T-units followed by a filled or unfilled pause (Lennon, 1990)
- Percentage of total pause time at T-unit boundaries (Lennon, 1990)
- Unspecified measures of filled and unfilled pauses (Wigglesworth & Elder, 2010)

^a A c-unit is a “grammatical independent clause with any of its modifiers” (Loban, 1963, p. 7).

^b An AS-unit (Analysis of Speech Unit) is “a single speaker’s utterance consisting of an independent clause, or sub-clausal unit, together with any subordinate clause(s) associated with either” (Foster et al., 2000, p. 365).

Ratios versus counts

One of the most basic issues related to studying pausing as a factor in fluency development involves determining how to quantify the number of pauses in a speech sample. To address this, researchers have calculated pausing time as a percentage of total speaking time (Hilton, 2008; Iwashita, Brown, McNamara, & O’Hagan, 2008; Mehnert, 1998) or its inverse, the phonation-time ratio (Kormos & Dénes, 2004; Préfontaine, 2010). Other studies have used a count of the total number of pauses (Foster & Skehan, 1996, Mehnert, 1998; Rossiter, 2009; Tavakoli & Skehan, 2005). Still others have included ratios of pauses to units, such as pauses per T-unit¹ (Bygate, 2001) or pauses per 1,000 words (Hilton, 2008), or pauses to time, such as pauses per minute (Iwashita et al., 2008; Kormos & Dénes, 2004; Llanes & Muñoz, 2009; Skehan & Foster, 2005). In sum, counting the number of pauses can be relative either to a time measure or to a speech unit measure. In the present study, we compare two different commonly used measures of pause ratios: a *pause to unit ratio* (pauses per 100 syllables) to a *pause to time ratio* (pauses per minute).

Filled versus unfilled

In addition to how pauses are measured, another variation between studies is which pauses are measured. Some researchers have counted only filled² (García-Amaya, 2009) or only unfilled³ (Guará Tavares, 2012) pauses, whereas others have counted both but distinguished between them (Kormos & Dénes, 2004; Rossiter, 2009), and still others have made no distinction between them at all (Foster & Skehan, 1996; Mehnert, 1998). It has been noted that not all pauses sound equally dysfluent, with unfilled pauses perhaps creating a greater sense of dysfluency (Chambers, 1997). Others regard them as similar, arguing that both filled and unfilled pauses equally “reflect an inability to handle speech during uninterrupted time” (Skehan & Foster, 2005, p. 199).

The few studies that have measured filled and unfilled pauses separately have led to divergent conclusions. In Iwashita et al. (2008), advanced learners had significantly fewer

unfilled pauses than learners at lower levels, whereas advanced learners had slightly more filled pauses than learners at lower levels, although this difference was not significant. Cenoz (2000) also found that advanced learners used more filled pauses, while learners at lower proficiency levels used more unfilled pauses. In Kormos and Dénes's (2004) data, in contrast, low-intermediate and advanced learners had nearly identical numbers of unfilled pauses, but the low-intermediate group had almost twice as many filled pauses. Given these mixed results, we continue to investigate this question by comparing measures of filled and unfilled pauses across levels of L2 proficiency.

Duration

As mentioned above, determining the length of pauses that should be counted is also subject to a great deal of debate in the literature. Pause duration varies across speakers, languages, and even tasks, which makes it challenging to generalize findings across studies (Cucchiarini, Strik, & Boves, 2002; De Jong & Bosker, 2013). However, it is still a crucial variable to examine, because selecting a length of pause to measure is a decision that fluency researchers cannot avoid. Riggenbach (1991) compared pausing in the speech of low- and high-proficiency learners of English, finding a difference between groups in the frequency of pauses lasting 0.5 seconds or more, but no differences in the frequency of shorter pauses. In the present study, we address a similar question by examining whether pauses greater than or equal to 0.5 seconds and pauses of 0.25 to 0.49 seconds differ in their relationship to L2 proficiency.

Pause location

Another important issue related to pausing and fluency concerns where the pause occurs in the utterance (De Jong, 2016; Skehan & Foster, 2005). We therefore examine whether mid-clause and end-of-clause pauses differ in their relationship to L2 proficiency to gain a better picture of these two measures and help guide decisions regarding the need to measure them separately.

L1 and L2 pausing

Along with questions of how to best operationalize pauses, it is important to take into account De Jong, Groenhout, Schoonen, and Hulstijn's (2013) proposal that pausing behaviour may be conceived of as a speaker trait and not only as an indicator of L2 proficiency. Speaker traits are characteristics particular to an individual and can be carried over to second languages. Considering the possibility of pausing as a speaker trait, we also compare patterns of pausing across L1 and L2 data in the present study.

Summary

The goal of this study is to examine how various measures of pauses relate to L2 proficiency (as measured by L2 grammar and vocabulary knowledge) to identify which pause measures best serve as markers of fluency. The research questions are as follows:

1. Pause duration: Is there a significant relationship between L2 proficiency and the rate of pauses 0.5 seconds and longer? Is there a significant relationship between L2 proficiency and the rate of shorter pauses (0.25–0.49 secs.)?
2. Pause types: Do filled and unfilled pauses differ in their relationship to L2 proficiency?
3. Pause location: Do mid-clause and end-of-clause pauses differ in their relationship to L2 proficiency?
4. L1 versus L2 pausing: How do participants' patterns of pausing in their L2 (Spanish) compare to their patterns of pausing in their L1 (English)?

Method

Participants

Forty-eight native speakers of English from the United States, Canada, and Australia participated in the study. Forty participants were studying abroad in Buenos Aires, Argentina, and completed the tasks for this study at the beginning of their semester abroad as part of a larger research project. Eight participants were studying Spanish at a large public university in the United States. All were university students between the ages of 19 and 46 ($M = 21.4$). They had a wide range of previous experience with Spanish, with the number of years of formal study varying from 1 to 12 ($M = 6.0$). Including participants who were beginning a semester abroad as well as participants in the United States served to increase the number of participants and increase statistical power; these participants were not treated as separate groups, given that the goal of this study was to determine how to operationally define pauses in L2 speech and how these different measures relate to L2 proficiency, and, therefore, context of learning is not relevant to the present discussion.

Materials, tasks, and data collection

Speaking tasks

Each participant completed three speaking tasks in English followed by three similar tasks in Spanish. Table 1 describes these tasks, all of which were monologues. The advantage of using monologues to collect the speaking data, in spite of a possible lack of ecological validity, is that monologues avoid the variability that interacting with another person introduces into the task (Segalowitz, 2010). For each task, the instructions were presented in SuperLab 5.0.⁴ Following the presentation of the instructions, participants were given 30 seconds to prepare their answer. Then they had a maximum of 2 minutes to respond to the prompt. While 30 seconds may not have been sufficient time for some participants to plan

Table 1: Speaking tasks

Task type	Prompt for L2 (Spanish) task	Prompt for L1 (English) task
Describe personal activities	Describe what you do on a typical weekday during the school year.	Describe what you do on a typical weekend during the school year.
Explain advantages and disadvantages	Explain the advantages and disadvantages of working part time while in college.	Explain the advantages and disadvantages of going to college immediately after graduating from high school.
Narrate in past time ^a	Look at the series of pictures below. Using the pictures, tell the story as a sequence of events that occurred in the past.	Look at the series of pictures below. Using the pictures, tell the story as a sequence of events that occurred in the past.

^a The L1 and L2 tasks used different picture series, both from Heaton (1966).

their responses, a short period of planning time was selected to prevent possible response rehearsal.

To verify the similarity of the tasks across English and Spanish, they were piloted with three native speakers of English and four native speakers of Spanish. Several measures of fluency were calculated, including speech rate, filled pauses (0.5 secs. and longer) per minute, and unfilled pauses (0.5 secs. and longer) per minute. No significant differences were found between the two tasks for each task type.

Grammar test

To measure grammatical knowledge, participants completed a 30-item test (see Appendix A for sample questions). Each item consisted of a sentence containing a grammatical error (verb tense, aspect or mood, adjective–noun agreement, incorrect pronoun). Participants had to detect it and write the correct form of the word in a space below the sentence.

Vocabulary test

Participants completed a vocabulary test adapted from the *Diploma de Español como Lengua Extranjera* (DELE). Portions of this test, which includes a 30-item multiple-choice vocabulary section and a 20-item multiple-choice cloze passage measuring both grammar and vocabulary, have previously been used in SLA research to assess the proficiency level of non-native speakers of Spanish (Slabakova, Rothman, & Kempchinsky, 2011; White, Valenzuela, Kozłowska-MacGregor, & Leung, 2004). As the purpose of this task in the present study was to measure vocabulary knowledge, only the 30-item vocabulary section was used.

Data analysis

Speaking tasks

Thirty-second excerpts were taken from all speaking tasks in both English and Spanish. Typically, each excerpt was taken from second 5 to second 35 of each task. Beginning at the fifth second allowed us to eliminate any fillers (e.g., *umm, so, okay, well*, etc.) that were sometimes present prior to the start of content that addressed the prompt. If a participant spoke for at least 31 but less than 35 seconds in response to a given prompt, the excerpt was taken from second 1 to second 31. In cases of short responses such as these, fillers were rarely present at the beginning of the sample. In a few cases (9 of 288 responses), participants spoke for less than 30 seconds in response to a prompt. In these cases, time was taken from the preceding or following task so that the excerpts for each participant totalled exactly 90 seconds in each language.

Additionally, following Rigggenbach (1991, 2000), unfilled pauses were limited to 3 seconds to prevent any one pause from having undue influence on the results (no filled pauses exceeded 3 seconds). Prior to taking the 30-second excerpts, pauses longer than 3 seconds were reduced to 3 seconds. There were 20 such pauses in the 288 responses.

Once the excerpts were taken, pauses greater than or equal to 0.25 seconds were marked in each excerpt using Praat 5.3.68 (Boersma & Weenink, 2014). As a first step, a script (Lennes, 2002b) was used to automatically mark silent pauses. The identified pause markers were then verified and adjusted manually as needed, and all filled pauses were marked manually. Each pause was labelled as unfilled (u) or filled (f), long (l; ≥ 0.5 secs.) or short (s; ≥ 0.25 secs. but < 0.5 secs.), and mid-clause (m) or end-of-clause (e). Pauses were considered end-of-clause pauses not only if they occurred between clauses as formally defined, but also if they occurred at a natural break, where a comma would be found in

written text (see Appendix B). Additionally, when a conjunction occurred between clauses, pauses on either side of the conjunction were marked as end-of-clause pauses, as in the sentence *I pack my bag after getting dressed and putting on my makeup [pause] and [pause] then I go to a class*. After the pauses were marked, another script (Lennes, 2002a) was used to produce a list of all pauses and their durations. Finally, for each excerpt, the number of syllables was counted for the purpose of calculating rates of pausing per 100 syllables.

Grammar and vocabulary tests

On the grammar and vocabulary tests, each correct answer received one point. The highest possible score on each test was 30 points.

Data transformations

Prior to carrying out other statistical analyses, all data were tested for normality using the Shapiro-Wilk test. For data found not to be normally distributed, the following transformations were applied and the data were then checked again for normality. No variables included in the analyses violated normality assumptions.

- Log transformations
 - L1 data: mid-clause pauses per 100 syllables, unfilled pauses per 100 syllables, pauses 0.25–0.49 seconds per 100 syllables
 - L2 data: all pauses per 100 syllables, mid-clause pauses per 100 syllables, unfilled pauses per 100 syllables, filled pauses per 100 syllables, pauses 0.25–0.49 seconds per 100 syllables, pauses ≥ 0.5 seconds per 100 syllables
- Square root transformations
 - L1 data: filled pauses per 100 syllables, filled pauses per minute
 - L2 data: pauses 0.25–0.49 seconds per minute

Addressing the research questions

A composite variable for proficiency was created by calculating the z-scores for the vocabulary and grammar tests and averaging them together. The averaged z-scores were the dependent variable for the correlation analysis and regression models that addressed the research questions above. Correlation tables and figures are presented for each set of variables with their corresponding research question.

Given the highly interrelated nature of the variables in this study, multi-collinearity was an issue (see De Jong et al., 2012). Many of the variables included were very highly correlated among themselves, rendering the interpretation of multiple regression models difficult. To address this, we conducted an exploratory factor analysis to reduce the data to a smaller set of summary variables, or latent factors.

Scree plots (*stats* package in R, R Core Team) were used to determine the number of latent factors and then factor scores were created for each participant. A factor score is a composite variable providing information on the participant's placement on a factor (DiStefano, Zhu, & Mindrila, 2009; Tabachnick & Fidell, 2007). The *sum scores by factor* method was used, by which scores on each factor are summed and then averaged, creating a factor score for each individual. For example, the factor analysis from the pause location data (Question 3) revealed that mid-clause pauses and end-of-clause pauses were highly correlated across both the per-minute and per-100-syllable measures, and indeed, scree-plot examination showed that two factors were appropriate for the exploratory factor analysis, which loaded on mid- and end-clause pauses. This factor score was included in the multiple regression analysis.

Results

Table 2 shows descriptive statistics for all pause measures. The results for each research question are given below.

Table 2: Descriptive statistics for all participants, Spanish and English

Variable	<i>M (SD)</i>	Min–Max
Proficiency score (Spanish)	37.17 (10.38)	8.00–57.00
Articulation rate, Spanish	3.83 (0.74)	2.14–5.71
Articulation rate, English	4.34 (0.54)	3.30–5.54
Pauses 0.25–0.49 secs. per 100 syllables, Spanish	4.38 (2.31)	1.32–12.35
Pauses 0.25–0.49 secs. per 100 syllables, English	3.31 (1.25)	1.14–7.32
Pauses 0.25–0.49 secs. per minute, Spanish	5.69 (2.76)	1.33–15.33
Pauses 0.25–0.49 secs. per minute, English	6.18 (2.67)	2.00–15.33
Pauses ≥ 0.50 secs. per 100 syllables, Spanish	14.15 (7.73)	2.23–37.21
Pauses ≥ 0.50 secs. per 100 syllables, English	7.64 (2.93)	2.23–15.43
Pauses ≥ 0.50 secs. per minute, Spanish	16.49 (3.63)	6.67–24.00
Pauses ≥ 0.50 secs. per minute, English	13.46 (3.38)	4.67–19.33
Pauses 0.50–0.99 secs. per minute, Spanish	8.11 (2.73)	2.67–14.67
Pauses 0.50–0.99 secs. per minute, English	7.83 (2.20)	3.34–12.67
Pauses 1.00–1.49 secs. per minute, Spanish	3.51 (1.65)	0.00–8.00
Pauses 1.00–1.49 secs. per minute, English	2.93 (1.78)	0.00–9.33
Pauses ≥ 1.50 secs. per minute, Spanish	4.85 (2.83)	0.00–11.33
Pauses ≥ 1.50 secs. per minute, English	2.76 (2.06)	0.00–8.00
Filled ^a pauses per 100 syllables, Spanish	7.63 (7.23)	0.00–36.05
Filled pauses per 100 syllables, English	3.13 (2.36)	0.00–12.25

(Continued)

Variable	<i>M (SD)</i>	Min–Max
Filled pauses per minute, Spanish	8.42 (4.97)	0.00–20.67
Filled pauses per minute, English	5.51 (3.60)	0.00–17.33
Unfilled pauses per 100 syllables, Spanish	20.77 (10.82)	5.13–53.85
Unfilled pauses per 100 syllables, English	11.52 (3.75)	4.83–22.84
Unfilled pauses per minute, Spanish	24.33 (4.59)	14.00–34.67
Unfilled pauses per minute, English	20.50 (3.92)	13.33–30.00
Mid-clause pauses per 100 syllables, Spanish	10.40 (7.27)	0.45–34.88
Mid-clause pauses per 100 syllables, English	3.59 (1.95)	0.58–9.88
Mid-clause pauses per minute, Spanish	11.65 (4.25)	1.33–22.67
Mid-clause pauses per minute, English	6.21 (2.47)	1.33–12.67
End-of-clause pauses per 100 syllables, Spanish	8.12 (2.20)	3.74–12.82
End-of-clause pauses per 100 syllables, English	7.37 (1.56)	3.91–10.33
End-of-clause pauses per minute, Spanish	10.49 (2.47)	5.33–15.33
End-of-clause pauses per minute, English	13.43 (2.31)	9.33–18.00

^a The rate of unfilled pauses and the rate of filled pauses add up to more than the rate of all pauses because of the way in which the pauses were counted. For the rate of all pauses, mid-clause pauses, end-of-clause pauses, long pauses, and short pauses, instances of a filled pause immediately followed by an unfilled pause (or vice versa) were counted as a single pause. For the rate of unfilled pauses and filled pauses, filled and unfilled pauses were all counted as separate pauses.

Q1: Pause duration and proficiency

The relationship between the L2 proficiency score (the composite L2 grammar and vocabulary score) and short pauses (0.25–0.49 secs.), measured per minute (unit measurement) and per 100 syllables (ratio measurement), was examined. The correlation between the number of pauses per 100 syllables and the proficiency score was negative and significant ($r = -0.48$, $p = 0.003$), whereas the correlation with pauses per minute was not ($r = 0.05$, $p = 0.66$). The unit and ratio measures were highly correlated ($r = 0.69$, $p < 0.001$).

For long pauses (≥ 0.5 secs.), the 100-syllable measure was negatively correlated with proficiency ($r = -0.60$, $p < 0.001$). The per-minute measure had a moderate negative correlation ($r = -0.30$, $p = 0.04$). The unit and ratio measures were also highly correlated ($r = 0.80$, $p < 0.001$).

The examination of scree plots showed that two latent factors were required (Factor 1: pauses 0.25–0.49 secs. per minute and per 100 syllables; Factor 2: pauses ≥ 0.5 secs., collapsed across pauses per minute and unit measures). Factor scores (see above) were then created for each participant, which served as the independent variables for the regression model. The overall model was significant ($R^2 = 0.271$, $F(2, 45) = 7.3$, $p = 0.014$). The 0.25–0.49-second pause duration variable did not reach significance ($\beta = -0.093$, $t(-1.6) = 0.118$), nor did the ≥ 0.5 -second pause measure ($\beta = 0.066$, $t(1.33) = 0.19$).

Since the overall model was significant but the individual variables were not, a multiple regression model was then run with the independent variables of pauses of 0.25–0.49, 0.5–0.99, 1–1.49, and ≥ 1.5 seconds. Multi-collinearity was not a problem; all relevant correlations were below 0.22 and non-significant. Overall, the model was significant ($R^2 = 0.28$, $F(4, 43) = 3.65$, $p = 0.012$). All coefficients were negative, but for pauses ≤ 0.99 seconds, neither variable reached significance (0.25–0.49 secs.: $\beta = -0.005$, $t(-0.103) = 0.411$; 0.5–0.99 secs.: $\beta = -0.008$, $t(-0.031) = 0.897$). The two independent variables that corresponded to longer pauses did reach significance: 1–1.49 seconds ($\beta = -0.126$, $t(-2.3) = 0.041$) and pauses greater than 1.5 seconds ($\beta = -0.126$, $t(-2.3) = 0.041$).

Q2: Pause type

The second research question concerns how the type of pause (filled or unfilled) per ratio (per minute) or unit measure (per 100 syllables) relates to the proficiency score. The correlation results are presented in Table 3.

All correlations between the proficiency score and filled/unfilled pauses were significant ($p < 0.05$) and negative. Among the different pause measures themselves, the highest correlation was observed between filled pauses per 100 syllables and filled pauses per minute ($r = 0.90$, $p < 0.001$), while the lowest was between unfilled pauses per minute and filled pauses per minute ($r = 0.52$, $p = 0.005$).

Scree plots showed that two latent factors were required (Factor 1: unfilled pauses per 100 syllables/per minute; Factor 2: filled pauses per 100 syllables/per minute). Factor scores were created from these variables for each participant, which served as the independent variables

Table 3: Correlations between filled and unfilled pauses

	Unfilled pauses/ 100 syllables	Unfilled pauses/ minute	Filled pauses/ 100 syllables	Filled pauses/ minute
Proficiency	-0.67**	-0.41**	-0.40**	-0.26*
Unfilled pauses/ 100 syllables		0.75**	0.76**	0.52**
Unfilled pauses/ minute			0.60**	0.54**
Filled pauses/ 100 syllables				0.90**

* $p < 0.05$.

** $p < 0.01$.

for the regression model. The overall model was significant ($R^2 = 0.398$, $F(2, 45) = 14.42$, $p = 0.014$). The filled pause variable was not significant ($\beta = 0.014$, $t(0.518) = 0.61$). The unfilled pause variable did reach significance, however ($\beta = -0.09$, $t(-4.15) < 0.001$).

Q3: Pause location

Next, the relationship between the location of pauses (mid vs. end-of-clause, per minute and per 100 syllables) and proficiency scores was investigated. The correlation results are presented in Table 4.

Proficiency was negatively correlated with the number of pauses across both mid-clause measures (all $ps < 0.001$). For end-of-clause measures per minute, the correlation with proficiency was positive, but for pauses per 100 syllables, the correlation was negative.

Scree plots showed that two latent factors were required (Factor 1: end-of-clause pauses per 100 syllables/per minute; Factor 2: mid-clause pauses per 100 syllables/per minute). Factor scores from these variables were created for each participant, which served as the independent variables for the regression model. The overall model was significant ($R^2 = 0.433$, $F(2, 45) = 17.19$, $p < 0.001$). Only the mid-clause pauses reached significance ($\beta = -0.113$, $t(-5.75) < 0.001$); the end-of-clause variable did not ($\beta = 0.009$, $t(0.146) = 0.884$).

Q4: Pause differences across L1 and L2

To address Question 4, focused on pause differences across L1 English and L2 Spanish, two sets of correlational analyses were carried out. The first included the number of pauses with durations of 0.25–0.49 and 0.5–0.99 seconds per minute and per 100 syllables for English and Spanish. The second correlational analysis expanded the length of pauses to include those between 1 and 1.49 seconds and those greater than 1.5 seconds per minute and per 100 syllables.

The correlation between the number of pauses per minute across English and Spanish was not significant for pauses of 0.25–0.49 seconds or for pauses of 0.5–0.99

Table 4: Correlations, pause location

	Mid-clause pauses/ 100 syllables	Mid-clause pauses/minute	End-of-clause pauses/ 100 syllables	End-of-clause pauses/ minute
Proficiency	-0.67**	-0.54**	-0.42**	0.54**
Mid-clause pauses/ 100 syllables		0.77**	0.57**	-0.68**
Mid-clause pauses/ minute			0.28*	-0.56**
End-of-clause pauses/ 100 syllables				-0.02

* $p < 0.05$.

** $p < 0.01$.

seconds. For pauses between 1 and 1.49 seconds, the correlation was positive, as was the correlation between pauses greater than 1.5 seconds in length. For pauses per 100 syllables, the correlation was not significant for pauses of 0.25–0.49 seconds, but there was a significant positive correlation for pauses of 0.5–0.99 seconds. Tables 5–7 show these correlations.

Table 5: L1 and L2 per-minute pause correlations

	Pauses 0.25–0.49 secs., English	Pauses 0.25–0.49 secs., Spanish	Pauses 0.5–0.99 secs., English	Pauses 0.5–0.99 secs., Spanish
Proficiency	–0.05	0.04	0.10	0.18
Pauses 0.25–0.49 secs., English		0.22	–0.38**	0.14
Pauses 0.25–0.49 secs., Spanish			–0.22	0.22
Pauses 0.5–0.99 secs., English				–0.06

* $p < 0.05$.

** $p < 0.01$.

Table 6: L1 and L2 per-100-syllable pause correlations

	Pauses 0.25–0.49 secs., English	Pauses 0.25–0.49 secs., Spanish	Pauses 0.5–0.99 secs., English	Pauses 0.5–0.99 secs., Spanish
Proficiency	–0.14	–0.50**	–0.16	–0.59**
Pauses 0.25–0.49 secs., English		0.17	–0.15	0.05
Pauses 0.25–0.49 secs., Spanish			–0.23	0.31*
Pauses 0.5–0.99 secs., English				0.61**

* $p < 0.05$.

** $p < 0.01$.

Table 7: L1 and L2 per-minute pause correlations, long pauses

	Pauses 1–1.49 secs., English	Pauses 1–1.49 secs., Spanish	Pauses ≥ 1.5 secs., English	Pauses ≥ 1.5 secs., Spanish
Proficiency	–0.35*	–0.14	–0.02	–0.49**
Pauses 1–1.49 secs., English		0.37**	0.29*	0.41**
Pauses 1–1.49 secs., Spanish			0.30*	0.12
Pauses ≥ 1.5 secs., English				0.54**

* $p < 0.05$.** $p < 0.01$.

Discussion

This study examined oral production data from L1 English/L2 Spanish speakers with the goal of operationalizing the L2 fluency construct of pauses and determining how different pause measurements relate to L2 proficiency. Proficiency was operationalized by calculating a normalized composite grammatical and lexical knowledge score for each participant. Four research questions were set out to address these issues, related to pause duration (long vs. short pauses), pause types (filled vs. unfilled), pause location (mid- vs. end-of-clause), and comparisons in pauses across the L1 and L2.

Pause length and ratios per minute versus per 100 syllables

For pauses per 100 syllables, there was a significant negative correlation between proficiency (operationalized as the composite score from grammatical and vocabulary tests) and both short (0.25–0.49 secs.) and longer (≥ 0.5 secs.) pauses. For pauses per minute, on the other hand, the negative correlation between proficiency and short pauses was not significant, and the negative correlation with longer pauses just reached significance. Importantly, both short-pause measurements (unit and ratio measures) were significantly correlated with one another, as were both long-pause measurements. The correlations between unit and ratio measures indicate a strong relationship between the two types of measure, so separating them may not always be necessary for researchers interested in examining the pause–proficiency relationship. However, the fact that the unit measure correlated with proficiency more strongly than the ratio measure suggests that they are not identical, and selecting one over the other could potentially influence the results of a study. The difference between these measures is discussed further in relation to the third research question.

Looking at the multiple regression results, neither pauses of 0.25–0.49 seconds nor pauses greater than or equal to 0.5 seconds significantly predicted changes in the proficiency score. Given this, we can conclude that pauses below 0.5 seconds do not account for a significant portion of the variability in the language proficiency score. For the longer

pauses, however, the lack of significance seems to be due to the fact that the general category of “pauses over 0.5 seconds” fails to capture the differences between “shorter” long pauses (i.e., under 1 sec.) and those that are longer (i.e., over 1 sec. in length). As noted above, when this category was further divided into pauses of 0.5–0.99, 1–1.49, and ≥ 1.5 seconds, the latter two subcategories did reach significance.

Taking into account pause threshold is an important issue for researchers to keep in mind. De Jong and Bosker (2013) analyzed the specific pause threshold that best captured the relationship between L2 proficiency and number of pauses and determined that 250 ms was optimal. Pauses shorter than this were irrelevant for calculating L2 fluency measures. Our results support their conclusion that longer pauses are more strongly related to L2 proficiency, although the threshold that we found for pauses to be a significant predictor of proficiency was considerably higher (1 sec.). Our findings suggest that if researchers select a lower threshold (such as 250 ms), it may be valuable to count shorter and longer pauses in separate categories to avoid missing possible differences in fluency revealed more clearly by longer pauses.

Filled and unfilled pauses

Research question 2 examined the relationship between pause type (whether filled or unfilled per minute and per 100 syllables) and proficiency. The filled and unfilled pause measures were all positively correlated with each other and negatively correlated with the proficiency score.

The regression results showed that the filled pause variable did not reach significance, but the unfilled pause variable did. The negative regression coefficient indicates that a higher number of unfilled pauses predicts a lower speaker proficiency level. This may be because speakers are treating filled pauses as a more natural part of their speech patterns and may in fact reflect personal speaking style, rather than L2 proficiency. Previous studies (Cenoz, 2000; Iwashita et al., 2008; Kormos & Dénes, 2004) found varying trends in the percentage of filled pauses for low- and high-proficiency learners. In De Jong et al. (2012), the authors suggest that both filled and unfilled pauses may in fact be the result of the same speech-planning process and will depend upon the speaker as to which is used. The results of the present study indicate a difference between filled and unfilled pauses across proficiency levels, but we concur with De Jong et al.'s (2012) suggestion that at least filled pauses may reflect individual speaker tendencies that could be consistent across the L1 and L2. Unfilled pauses, according to the results here, may not be. The type of pause that speakers use does depend upon proficiency; specifically, more proficient learners have fewer unfilled pauses.

Mid-clause and end-of-clause pauses

Research question 3 examined how pause location related to proficiency. Mid-clause pauses per 100 syllables and also per minute were negatively correlated with proficiency (and positively correlated between each other). The end-of-clause pauses patterned differently, however. The per-minute measure was positively correlated with proficiency and the per-100-syllable measure was negatively correlated. The multiple regression model showed that only the mid-clause pauses were significant: As the number of mid-clause pauses went up, proficiency scores went down. These results parallel those in a recent study by De Jong (2016), who examined the distribution of silent and filled pauses in L2 Dutch speakers across different pause locations. She found that within-utterance pauses were more frequent in L2 speech but there was an interaction between pause location and lexical frequency

of the word following the mid-clause pause: Pauses were more likely before less frequent words for both L1 and L2 speakers. These results indicate that mid-clause pauses for both native and non-native speakers are a consequence of lexical-retrieval difficulties, related to the cognitive fluency aspect of overall fluency (Segalowitz, 2010). Thus, the findings here align with those of De Jong (2016): End-of-clause pauses do not seem to be affected by L2 proficiency while mid-clause pauses are, and this may be a consequence of lexical-retrieval processes. End-of-clause pauses may fulfil a different fluency function.

L1 versus L2 pausing

The lack of correlation across English and Spanish for short pauses (0.25–0.49 secs. per minute and per 100 syllables) suggests that short pause patterning is not consistent across a speaker's two languages. For longer pauses, however, there were significant positive correlations.

As noted above, De Jong et al. (2013) suggest that pausing behaviour is a speaker trait and not just an indicator of L2 proficiency; pausing seems to vary across individual speakers and is consistent across L1 and L2. They found that various measures of L1 and L2 fluency related to pauses (pause duration and number of pauses, among others) were positively correlated in the speech of L2 Dutch learners, and a subsequent regression analysis showed that L2 fluency measures could be predicted on the basis of the same L1 fluency measure. They recommend using L1-corrected measures of L2 fluency in order to account for these differences. In the present study, the finding of significant correlations between several measures of L1 and L2 pausing lends support to the idea of pausing as (in part) a speaker trait. This is an important issue for researchers to consider when investigating pause behaviour in an L2 and particularly when the stakes are high for those being evaluated.

As has been well documented in the L2 fluency literature (Derwing, 2017), articulation-rate (the number of syllables per second after the removal of pauses) differences between the L1 and L2 can also greatly influence pause-measure comparisons across learners' L1 and L2. For example, articulation-rate differences may influence comparisons in the number of pauses per minute versus the number of pauses per 100 syllables by enlarging the difference between the two measures, given that speakers may produce fewer syllables in their L2 than in their L1.⁵ The present study does not explicitly consider articulation rate, but it is important for future researchers to keep in mind.

Conclusion

The overall goal of this study was to examine how different pause measures relate to L2 proficiency (as measured by a composite normalized z-score of grammar and vocabulary knowledge) to identify which pause measures best serve as markers of fluency. The research questions focused on how pause durations, pause location, and pause type differ across per-minute and per-100-syllable measures of pauses in the L2 and how pauses of different durations compare across the L1 and L2.

The present study has certain limitations that should be addressed in future research. To more precisely describe participants' levels of L2 ability, it would be ideal to conduct more comprehensive proficiency testing, rather than rely solely on written grammar and vocabulary tests, and also to include processing measures as part of cognitive fluency. Indeed, work by De Jong and colleagues shows that proficiency may not be as clearly connected to pause durations as researchers initially believed. Related to this is the need to bring in a wider variety of L1s and L2s, to determine how much of pausing behaviour may be language-particular (see De Jong & Bosker, 2013, for a study that did this).

When examining a speech feature such as pauses, broad generalizations may be difficult to make. However, the findings of this study suggest that different methods of measuring pauses can lead to different conclusions about L2 fluency. To summarize, it was found that longer pauses (≥ 1 sec.) are better predictors of L2 proficiency than shorter pauses, unfilled pauses are a stronger predictor of L2 proficiency than filled pauses, and mid-clause pauses distinguish between proficiency levels more clearly than do end-of-clause pauses. Additionally, significant correlations between L1 and L2 pauses support the idea (De Jong et al., 2013) that pausing patterns are partially a speaker trait; therefore, it is valuable to consider L1 pausing behaviour when measuring L2 pauses.

Christine Shea is an associate professor of Spanish linguistics at the University of Iowa. Her research focuses on second language acquisition of phonetics, phonology, and speech perception.

Karen Leonard is an assistant professor at Olivet Nazarene University. Her research focuses on Spanish applied linguistics.

Notes

1. A T-unit is “one main clause plus whatever subordinate clauses happen to be attached to or embedded within it” (Hunt, 1965, p. 305).
2. Common examples of filled pauses in English include “uh” and “umm.”
3. Unfilled pauses refer to periods of silence.
4. See <https://www.cedrus.com/superlab/>
5. We thank an anonymous reviewer for pointing this out.

References

- Boersma, P., & Weenink, D. (2014). Praat 5.3.68 [Computer software]. Retrieved from <http://www.praat.org>
- Bosker, H.R., Pinget, A.-F., Quené, H., Sanders, T., & De Jong, N.H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, 30(2), 159–175. <https://doi.org/10.1177/0265532212455394>
- Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching, and testing* (pp. 23–48). Harlow, England: Longman.
- Cenoz, J. (2000). Pauses and hesitation phenomena in second language production. *International Journal of Applied Linguistics*, 127(1), 53–69. <https://doi.org/10.1075/itl.127-128.03cen>
- Chambers, F. (1997). What do we mean by fluency? *System*, 25(4), 535–544. [https://doi.org/10.1016/S0346-251X\(97\)00046-8](https://doi.org/10.1016/S0346-251X(97)00046-8)
- Cucchiari, C., Strik, H., & Boves, L. (2002). Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *Journal of the Acoustical Society of America*, 111, 2862–2873. <https://doi.org/10.1121/1.1471894>
- D'Amico, M.L. (2010). *The effects of intensive study abroad and at home language programs on second language acquisition of Spanish* (Unpublished doctoral dissertation). University of Florida, Gainesville, FL.
- 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. <https://doi.org/10.1515/iral-2016-9993>
- De Jong, N.H., & Bosker, H.R. (2013). Choosing a threshold for silent pauses to measure second language fluency. In R. Eklund (Ed.), *Proceedings of the 6th workshop on disfluency in spontaneous speech* (pp. 17–20). Stockholm, Sweden: KTH Royal Institute of Technology.

- De Jong, N.H., Groenhout, R., Schoonen, R., & Hulstijn, J.H. (2013). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, 36(2), 223–243. <https://doi.org/10.1017/S0142716413000210>
- De Jong, N.H., Steinel, M.P., Florijn, A., Schoonen, R., & Hulstijn, J.H. (2012). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, 34(5), 893–916. <https://doi.org/10.1017/S0142716412000069>
- Derwing, T.M. (2017). L2 fluency development. In S. Leowen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 246–259). New York, NY: Routledge.
- Derwing, T.M., Munro, M.J., Thomson, R.I., & Rossiter, M.J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, 31(4), 533–557. <https://doi.org/10.1017/S0272263109990015>
- DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation*, 14, 1–11. Retrieved from <https://pareonline.net/getvn.asp?v=14&n=20>
- Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, 18(3), 299–323. <https://doi.org/10.1017/S0272263100015047>
- Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21(3), 354–375. <https://doi.org/10.1093/applin/21.3.354>
- García-Amaya, L. (2009). New findings on fluency measures across three different learning contexts. In J. Collentine (Ed.), *Selected proceedings of the 11th Hispanic linguistics symposium* (pp. 68–80). Somerville, MA: Cascadilla Proceedings Project.
- Guará Tavares, M.G. (2009). The relationship among pre-task planning, working memory, capacity, and L2 speech performance: A pilot study. *Linguagem e Ensino*, 12, 165–194.
- Guará Tavares, M.G. (2012). *Pre-task planning, working memory capacity, and L2 speech performance* (Unpublished doctoral dissertation). Universidade Federal de Santa Catarina, Florianópolis, Brazil. Retrieved from <https://repositorio.ufsc.br/bitstream/handle/123456789/91831/254908.pdf?sequence=1>
- Heaton, J.B. (1966). *Composition through pictures*. London, England: Longman.
- Hilton, H. (2008). The link between vocabulary knowledge and spoken L2 fluency. *Language Learning Journal*, 36(2), 153–166. <https://doi.org/10.1080/09571730802389983>
- Hunt, K.H. (1965). A synopsis of clause-to-sentence length factors. *English Journal*, 54(4), 300+305–309. <https://doi.org/10.2307/811114>
- Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics*, 29(1), 24–49. <https://doi.org/10.1093/applin/amm017>
- Kirsner, K., Dunn, J., Hird, K., Parkin, T., & Clark, C. (2002). Time for a pause. In C. Bow (Ed.), *Proceedings of the 9th Australian international conference on speech science & technology* (pp. 52–57). Canberra City, Australia: Australian Speech Science and Technology Association.
- Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32(2), 145–164. <https://doi.org/10.1016/j.system.2004.01.001>
- Lennes, M. (2002a). Calculate segment durations [Praat script]. Retrieved from <https://lennes.github.io/spect/>
- Lennes, M. (2002b). Mark pauses [Praat script]. Retrieved from <https://lennes.github.io/spect/>
- Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40(3), 387–417. <https://doi.org/10.1111/j.1467-1770.1990.tb00669.x>
- Leonard, K.R., & Shea, C.E. (2017). L2 speaking development during study abroad: Fluency, accuracy, complexity, and underlying cognitive factors. *Modern Language Journal*, 101(1), 179–193. <https://doi.org/10.1111/modl.12382>
- Llanes, À., & Muñoz, C. (2009). A short stay abroad: Does it make a difference? *System*, 37(3), 353–365. <https://doi.org/10.1016/j.system.2009.03.001>

- Loban, W. (1963). *Language of elementary school children*. Champaign, IL: National Council of Teachers.
- Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, 20(1), 83–108. <https://doi.org/10.1017/S0272263198001041>
- Michel, M.C., Kuiken, F., & Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *International Review of Applied Linguistics in Language Teaching*, 45(3), 241–259. <https://doi.org/10.1515/iral.2007.011>
- Mota, M.B. (2003). Working memory capacity and fluency, accuracy, complexity, and lexical density in L2 speech production. *Fragmentos*, 24, 69–104.
- Nakakubo, T. (2011). *The effects of planning on second language oral performance in Japanese: Processes and production* (Unpublished doctoral dissertation). University of Iowa, Iowa City, IA. Retrieved from <http://ir.uiowa.edu/etd/1038/>
- Préfontaine, Y. (2010). Differences in perceived fluency and utterance fluency across speech elicitation tasks: A pilot study. In G. Bota, H. Hargreaves, L. Chia-Chun, & R. Rong (Eds.), *Papers from the Lancaster University postgraduate conference in linguistics & language teaching* (pp. 134–154). Lancaster, England: Lancaster University.
- Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes*, 14(4), 423–441. <https://doi.org/10.1080/01638539109544795>
- Riggenbach, H. (2000). *Perspectives on fluency*. Ann Arbor, MI: University of Michigan Press.
- Rossiter, M.J. (2009). Perceptions of L2 fluency by native and non-native speakers of English. *The Canadian Modern Language Review*, 65(3), 395–412. <https://doi.org/10.3138/cmlr.65.3.395>
- Sangarun, J. (2005). The effects of focusing on meaning and form in strategic planning. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 111–141). Amsterdam, Netherlands: John Benjamins.
- Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York, NY: Routledge.
- Segalowitz, N., & Freed, B.F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition*, 26(2), 173–199. <https://doi.org/10.1017/S0272263104262027>
- Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, 1(3), 185–211. <https://doi.org/10.1177/136216889700100302>
- Skehan, P., & Foster, P. (2005). The effects of focusing on meaning and form in strategic planning. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 193–216). Amsterdam, Netherlands: John Benjamins.
- Slabakova, R., Rothman, J., & Kempchinsky, P. (2011). Gradient competence at the syntax-discourse interface. *EUROSLA Yearbook*, 11(1), 218–243. <https://doi.org/10.1075/eurosla.11.12rot>
- Tabachnick B.G., & Fidell, L.S. (2007). *Using multivariate statistics*. (5th ed). Boston, MA: Pearson Education.
- Tajima, M. (2003). *The effects of planning on oral performance of Japanese as a foreign language* (Unpublished doctoral dissertation). University of Iowa, Iowa City, IA. Retrieved from <https://ir.uiowa.edu/cgi/viewcontent.cgi?article=2423&context=etd>
- Tavakoli, P., & Foster, P. (2011). Task design and second language performance: The effect of narrative type on learner output. *Language Learning*, 61(s1), 37–72. <https://doi.org/10.1111/j.1467-9922.2011.00642.x>
- Tavakoli, P., & Skehan, P. (2005). The effects of focusing on meaning and form in strategic planning. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–273). Amsterdam, Netherlands: John Benjamins.
- Towell, R. (2002). Relative degrees of fluency: A comparative case study of advanced learners of French. *International Review of Applied Linguistics in Language Teaching*, 40(2), 117–150. <https://doi.org/10.1515/iral.2002.005>

- Valls Ferrer, M. (2008). Language acquisition during a stay abroad period following formal instruction: Temporal effects on oral fluency development (Unpublished doctoral dissertation). Pompeu Fabra University, Barcelona, Spain. Retrieved from http://www.academia.edu/535791/Language_acquisition_during_a_stay_abroad_period_following_formal_instruction_temporal_effects_on_oral_fluency_development
- White, L., Valenzuela, E., Kozłowska-MacGregor, M., & Leung, Y.-K.I. (2004). Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics*, 25(1), 105–133. <https://doi.org/10.1017/S0142716404001067>
- Wigglesworth, G., & Elder, C. (2010). An investigation of the effectiveness and validity of planning time in speaking test tasks. *Language Assessment Quarterly*, 7(1), 1–24. <https://doi.org/10.1080/15434300903031779>

Appendix A: Sample questions from grammar test

Each sentence below contains a grammatical error. Your task is to find the error and write the corrected form in the box below the sentence.

Model: Adriana están feliz.

Está

3. Se le mandé a las diez de la mañana por correo electrónico.

6. Se prohíben fumar aquí.

Appendix B: Examples of mid-clause and end-of-clause pauses

Mid-clause pauses, English:

And you'll also finish your undergraduate [pause] sooner.

Mid-clause pauses, Spanish:

Muchas veces vamos [pause] vamos al [pause] cine. = Many times we go [pause] we go to the [pause] movies.

End-of-clause pauses, English:

Taking time off helps them prepare for that [pause] but [pause] I think the advantage is that . . .

End-of-clause pauses, Spanish:

Los sábados a veces trabajo otra vez [pause] y si no [pause] hago mi tarea. = On Saturdays sometimes I work again [pause] and if not [pause] I do my homework.