Accepted, *Applied Psycholinguistics*
huensch@usf.edu

Understanding L2 fluency behavior: The effects of individual differences in L1 fluency, cross-linguistic differences, and proficiency over time

Amanda Huensch and Nicole Tracy-Ventura

University of South Florida

Correspondence concerning this article should be addressed to Amanda Huensch, Department of World Languages, University of South Florida, 4202 E. Fowler Avenue CPR 107, Tampa, FL, USA 33620-5700, Tel.: +1 813 974 2548, E-mail: huensch@usf.edu

Acknowledgments

LANGSNAP (http://langsnap.soton.ac.uk/) was funded by the ESRC (award number RES-062-23-2996). We are grateful to our colleagues Rosamond Mitchell, Kevin McManus, Laurence Richard, Patricia Romero de Mills, Karen Ruebens, the participants, and our transcribers and research assistants for their contribution to this research. We also thank those from the SPLLOC 2 project, funded by the ESRC (award number RES-062-23-1075). Finally, we wish to thank Luke Plonsky and Tom Parker for their assistance with the statistical methods and two anonymous reviewers for their comments.

Revised manuscript submitted October 3, 2016

huensch@usf.edu

Abstract

This study examined the extent to which L1 fluency behavior, cross-linguistic differences, and proficiency can predict L2 fluency behavior over time. English L1 Spanish (n = 24) and French (n = 25) majors completed a picture-based oral narrative in the L2 before and after five months residing abroad and later in the L1 after returning home. Data were coded for seven measures of speed, breakdown, and repair fluency. Results from multiple regressions indicated that L1 fluency behavior, cross-linguistic differences, and proficiency differentially contributed to explaining L2 fluency behavior prior to and during immersion. These findings suggest that when investigating L1-L2 fluency relationships considerations of mitigating factors such as cross-linguistic differences are necessary, and it is worthwhile to focus on how the contributions of these factors shift during development.

Research on second language (L2) fluency has begun to investigate the relationship between first language (L1) and L2 fluency behavior in an effort to understand whether fluency is an underlying trait of a speaker or a language specific state (De Jong, Groenhout, Schoonen, & Hulstijn, 2015; Derwing, Munro, Thomson, & Rossiter, 2009; Riazantseva, 2001; Towell, Hawkins, & Bazergui, 1996; Towell & Dewaele, 2005). Exploring this relationship is important because it not only provides evidence for the development of speech production models by helping to determine answers to questions such as how much of L2 fluency behavior can be attributed to L1 fluency behavior (Segalowitz, 2010), but it also has practical applications for language assessment in that knowing whether L1 and L2 fluency behavior are related may provide the necessary justification for correcting L2 fluency measures for L1 speaking style (De Jong et al., 2015; Ginther, Dimova, & Yang, 2010).

Thus far, the limited research from this line of inquiry has produced complicated results. Overall, however, the findings appear to suggest that the relationship between L1 and L2 fluency behavior may be mitigated by at least three main factors: (a) cross-linguistic similarities and differences between the languages under investigation (e.g., measurements of speed fluency can be confounded with phonotactics/morphology; different languages may evidence different pausing behavior patterns – Derwing et al., 2009; De Jong et al., 2015; Riazantseva, 2001), (b) changes in L2 proficiency (e.g., the relationship between L1 and L2 fluency behavior may change as learners' proficiency increases – Derwing et al., 2009; Riazantseva, 2001; Towell & Dewaele, 2005), and (c) a combination of the above (e.g., lower proficiency learners may benefit more from cross-linguistic similarities than if the L1 and the L2 are more different – Derwing et al., 2009). If cross-linguistic differences, proficiency, and L1 fluency behavior together

contribute to L2 fluency behavior, then it is likely the case that different populations studied will result in different findings if these factors are not considered in conjunction.

The current study contributes to our understanding of L2 fluency behavior by investigating the extent to which L1 fluency behavior, cross-linguistic differences, and proficiency predict L2 fluency behavior over time. Oral data were collected from two groups of temporary sojourners (English L1) who were residing abroad either in a French (n = 25) or Spanish-speaking (n = 24) country as part of a university exchange program (i.e., a form of study abroad). Data were collected twice in the L2 and once in the L1. Two additional L1 groups (French and Spanish) completed the tasks in their native language to allow for cross-linguistic comparisons of utterance fluency measures across the three languages. The relationship between the participants' L1 and L2 fluency behavior is compared both before and after 5 months' residence abroad, a period in which it is demonstrated that they also significantly improve in their L2 proficiency and L2 fluency. Finally, all of these variables are analyzed together to examine to what extent the predictive power of each variable changes over time.

Defining Fluency

Fluency as a term can have both a broad and narrow definition (Lennon, 1990). Whereas a broad definition of fluency can be equated to general proficiency, a narrow definition is concerned with the "temporal aspects of oral production that influence the degree of fluidity in speech (e.g., pauses, hesitation phenomena, speech rate)" (Derwing et al., 2009, p. 534). It is this narrow definition that we adopt in the current study. Segalowitz (2010) further differentiates between three types of fluency: (a) cognitive fluency, or the efficacy of one's underlying processing abilities; (b) utterance fluency, or temporal measures of one's oral productions; and

(c) perceived fluency, or how listeners judge one's oral productions with regard to fluency. Much of the work in the area of L2 fluency to date has explored aspects of utterance fluency while connecting these temporal measures to (a) perceived fluency and/or L2 proficiency (Baker-Smemoe, Dewey, Bown, & Martinsen, 2014; Bosker, Pinget, Quené, Sanders, & De Jong, 2012; Cucchiarini, Strik, & Boves, 2000, 2002; Derwing et al., 2009; Ginther, Dimova, & Yang, 2010; Kormos & Dénes, 2004; Riazantseva, 2001), (b) cognitive fluency (De Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2013; Kahng, 2014; Segalowitz & Freed, 2004), (c) task type (Skehan, 2003; Tavakoli & Skehan, 2005), and (d) development (Du, 2013; Freed 1995; Lennon, 1990; Mora & Valls-Ferrer, 2012; Towell, Hawkins, & Bazergui, 1996; Towell & Dewaele, 2005).

Skehan (2003) and Tavakoli and Skehan (2005) further subdivided utterance fluency into three types: speed, breakdown, and repair fluency. Speed fluency refers to the rate at which one is able to produce language. Breakdown fluency is comprised of measures of duration, location, and number of silent and filled pauses. Repair fluency is comprised of measures of repetitions, corrections, and false-starts, among others. Using these subcategories has proved beneficial in understanding utterance fluency and its connections to perceived fluency as well as L2 processing and proficiency. For example, research exploring pausing patterns has demonstrated effects of processing (Kahng, 2014; Pawley & Syder, 2000) while other work (Baker et al., 2014; Cucchiarini et al., 2002; Ginther et al., 2010; Kormos & Dénes, 2004) has found that some aspects of utterance fluency are connected to L2 proficiency (speed, breakdown), but not others (repair). Mixed/contradictory findings from this body of work, however, have highlighted the need for consistent operationalization of temporal measures of utterance fluency. Given the possibility that aspects of L2 fluency are influenced in different ways by L1 fluency behavior, cross-linguistic differences, and proficiency, then the use of less confounded, as opposed to more

global, measurements of fluency such as those in De Jong et al. (2015), is warranted in the current investigation.

Models of Speech Production

The work of Levelt (1989, 1999) and De Bot (1992) describing the internal speech production model of an L2 speaker has been influential in contributing to an understanding of what is involved in fluent speech. The model begins at conceptualization, the site of message generation (i.e., the preverbal message), then moves to formulation, where the original message is formed through grammatical and phonological encoding of lemmas from the lexicon. Finally, the last stage, articulation, is where the phonetic plan is converted to actual speech. Drawing upon this work, Segalowitz (2010) presented an adapted 'blueprint' of the L2 speaker, which summarizes the "linguistic, psycholinguistic, and cognitive issues underlying the act of speaking" (p. 8). The model also includes seven "critical points in the architecture" which can be considered "fluency vulnerability points where processing difficulties might be expected to give rise to L2 dysfluencies" (p. 17). Examples of these include grammatical encoding, lexical encoding, self-monitoring, etc. One limitation of the model, as acknowledged by Segalowitz (and Levelt and De Bot), is that it does not provide information about development, but rather a picture at a single point in time. However, in order to incorporate such developmental information into the model, it is necessary to first understand the relationship between L1 and L2 fluency, and uncover whether certain aspects of fluency are unique to L2 speech and not the result of individual differences in L1 fluency behavior (De Jong, Schoonen, & Hulstijn, 2009; Segalowitz, 2010). Additionally, it might be the case that the relationship between L1 and L2

fluency changes as L2 proficiency develops and that cross-linguistic differences between the L1 and the L2 mitigate this relationship.

Relationship between L1 and L2 Fluency

A small body of research has begun to examine the relationship between L1 and L2 fluency (De Jong et al., 2015; Derwing et al., 2009; Riazantseva, 2001; Towell et al., 1996; Towell & Dewaele, 2005) and has adopted a variety of approaches including (a) comparing group differences between the L1 and L2, and learners and NSs using ANOVAs (Towell et al., 1996), (b) analyzing relationships between L1 and L2 measures using correlations (De Jong et al., 2015; Derwing et al., 2009; Towell & Dewaele, 2005), (c) investigating whether L1 measures can predict L2 measures using linear regressions (De Jong et al., 2015), and (d) attempting to partial out L1 effects either by conducting ANCOVAs (Towell & Dewaele, 2005) or using corrected measures of L2 fluency to predict L2 proficiency (De Jong et al., 2015). Although results thus far appear to support the notion that L2 fluency relates to L1 fluency (De Jong et al., 2015; Towell & Dewaele, 2005), other variables such as the structural characteristics between the L1 and L2 (De Jong et al., 2015) and participants' proficiency appear to mitigate the strength of the relationship (Derwing et al., 2009). Additionally, the fact that researchers have used different measures of utterance fluency makes comparisons across studies difficult.

Towell et al. (1996) was the first study to investigate the development of L2 fluency before and after a six-month stay abroad and then compare aspects of L2 fluency to the same learners' L1 fluency behavior. Data were collected from 12 English L2 learners of French via a film narration and three measures of speed fluency and two measures of breakdown fluency were calculated: (a) speaking rate (number of syllables divided by the total time including pauses), (b)

articulation rate (number of syllables divided by the total time excluding pauses), (c) mean length of run (average number of syllables produced between silent pauses), (d) phonation time ratio (percent proportion of time spent speaking to time spent to produce the entire sample), and (e) average length of pauses (mean duration of silent pauses). Results showed that learners improved in speed measures (speech rate, articulation rate, and mean length of run), but neither breakdown measure during study abroad. In addition, for those L2 measures that improved, they still "lagged behind" L1 measures (p. 103). Learners were, however, able to approximate L1 fluency in average length of pause. Based on these findings, Towell et al. claimed "that advanced L2 subjects reach a plateau with respect to speaking rate" that is below their L1 (p. 113).

It should be noted, however, that when comparing L2 fluency measures to L1 fluency measures it is important to consider the effect of cross-linguistic differences. For example, syllable structure and pausing can show considerable variation across languages. Roach (1998) argued that when measuring speech rate using syllable counts, the syllable structures of the languages in question will influence the results. English has a wider syllable inventory than Spanish; thus, the number of phones within the syllables produced by an English speaker will most likely be greater than those of a Spanish speaker, and thus it is predicted that the speaking rates of English speakers will be lower than those of Spanish speakers. Pellegrino, Coupé, and Marsico (2011) in fact found speech rate differences between Spanish and English when calculating the number of syllables per second. Similar results comparing French and German were found in Trouvain and Möbius (2014); the articulation rate of German speakers was slower than that of French speakers (like English, German has a more complex syllable structure than French). Thus, the different phonotactics of the languages under study in Towell et al. (English L1 and French L2) most likely influenced the plateau found in speaking rate.

Towell et al. (1996) also demonstrated that learners were also able to approximate L1 fluency in average length of pause, but without baseline data from French NSs, it is unknown whether learners were also approximating NSs and whether pausing characteristics between French and English are similar. Previous research (Grosjean & Deschamps, 1975, p. 159) comparing pausing behavior of French and English native speakers in interviews indicated cross-linguistic differences regarding the frequency and duration of pauses; English speakers paused more frequently than French speakers, but for shorter periods of time. Riazantseva (2001) included baseline data from a group of native English speakers in her cross-sectional study of pausing characteristics of Russian L1 learners of English who contributed data in both their L1 and L2. Significant cross-linguistic differences were found between native speakers of Russian and English in pause duration but not pause frequency or distribution. NSs of Russian paused longer in Russian than NSs of English paused in English, but no differences were found between these two groups in the frequency or distribution of their pauses.

To investigate whether cross-linguistic similarities and differences may influence the relationship between L1 and L2 fluency, Derwing et al. (2009) explored the L2 fluency behavior of Slavic and Mandarin L1 learners of English longitudinally. Participants included highly-educated immigrants in Canada who were beginning learners of English. Using the same narrative picture-based task, Derwing et al. collected recordings at 2 months (time 2), 10 months (time 6) and 1 year after these learners arrived in Canada (time 7). Temporal measures included (a) the number of pauses per second, (b) speech rate (i.e., the number of syllables per second which includes pausing time), and (c) pruned syllables per second (i.e., the total number of syllables without repetitions, corrections, filled pauses, etc.). Perceived fluency ratings were also collected for both groups and indicated significant gains for the Slavic group, but not the

Mandarin group from time 2 to time 7. Results also indicated significant positive correlations between L1 and L2 temporal measures of fluency for both Mandarin and Slavic speakers at time 2 but significant correlations at times 6 and 7 were found for the Slavic group only. Derwing et al. argued that the lack of correlation for the Mandarin group at later time points could have been related to the lack of improvement of the group as a whole.

Perhaps the strongest evidence for a link between L1 and L2 fluency comes from De Jong et al. (2015)'s investigation of English and Turkish L1 learners of Dutch. Their aim was to investigate whether measures of L2 fluency that were corrected for L1 fluency behavior could predict L2 Dutch proficiency better than uncorrected measures. Participants included learners with intermediate to advanced proficiency in Dutch who had been living in the Netherlands for an average of 4.5 years. Seven measures of temporal fluency were investigated following the subcategories set forth by Skehan (2003) and Tavakoli and Skehan (2005). As these are the measurements adopted in the analysis in the current study, a detailed explanation and justification for their use is presented here. Speed fluency was represented by mean syllable duration (i.e., inverse articulation rate) and was calculated as the total speaking time (excluding silent pauses) divided by the total number of syllables. This measure differs from the measure of speech rate used in previous work (e.g., Derwing et al., 2009; Towell et al., 1996; Towell & Dewaele, 2005) in that it excludes silent pauses and thus is a less confounded measure of fluency. Breakdown fluency was comprised of four measures including those that take frequency, duration, and location into account. To investigate the location of silent pauses and whether behavior changed either within or between syntactic units, the analysis of speech unit (ASU) was used. This unit is comprised of "an independent clause, or sub-clausal unit, together with any subordinate clause(s) associated with either" (Foster, Tonkyn, & Wigglesworth, 2000,

p. 365). Taking frequency, duration, and location into account, the four measures of breakdown fluency were (a) the mean length of silent pauses within ASU, (b) the mean length of silent pauses between ASU, (c) the number of silent pauses per second speaking time, and (d) the number of non-lexical filled pauses per second speaking time. In addition to using more fine-grained measures of pausing than Derwing et al., (2009), who used only average (silent) pause duration, De Jong et al. used a different silent pause threshold: 250 ms as opposed to 400 ms. A silent pause threshold of 250 ms has been shown to correlate more with measures of L2 proficiency (De Jong & Bosker, 2013; Kahng, 2014). Finally, repair fluency, which represents the online modification of utterances (Tavakoli & Skehan, 2005), was comprised of two measures and included (a) the number of repetitions per second speaking time and (b) the number of corrections per second speaking time.

Comparing L1 and L2 performance on these measurements of temporal fluency, De Jong et al. reported significant differences on all temporal measures for both groups between Dutch and their L1s. Results also indicated cross-linguistic differences between L1 Turkish and L1 English. For example, mean syllable duration differed between L1 groups which they attributed to the simpler syllable structure in Turkish. Additionally, all seven L1 and L2 fluency measures correlated significantly when both L1 groups were combined. Conducting linear regression models to predict L2 fluency from L1 fluency and language group, results indicated L1 fluency significantly improved the model for each temporal measure. Language group contributed to those models where L1 differences were found between Turkish and English. Saving the residuals from these regression analyses, they determined whether corrected measures of L2 fluency could better predict proficiency (as measured by a productive vocabulary task) than uncorrected measures. Although they found that all seven L1 measures could predict

characteristics of the L2 fluency measures, only one of the corrected measures (mean syllable duration) was a stronger predictor of L2 proficiency than the uncorrected measures. Towell & Dewaele (2005) also concluded that "fluency in the L2 reflects fluency in the L1, most directly in speaking rate" (p. 232).

Based on this review of the literature, the research investigating the relationship between L1 and L2 fluency has generally found medium to strong correlations between L1 and L2 measures of utterance fluency. However, it has been shown through cross-linguistic comparisons that languages may differ in ways that could mediate the strength of the relationship between L1 and L2 fluency. Furthermore, longitudinal studies such as Derwing et al. (2009) show different results over time depending on L1 group and proficiency level. Therefore, to further our understanding of L2 fluency behavior it is necessary to consider the extent to which variables such as cross-linguistic differences and proficiency mediate the relationship between L1 and L2 fluency.

Research Questions

The goal of the current study is to explore the relative contribution of L1 fluency behavior, cross-linguistic differences, and proficiency in predicting L2 fluency behavior. By comparing the relative predictive power of these variables for pre-immersion L2 fluency and during-immersion L2 fluency, it is possible to examine if the predictive power of these variables shifts over time. Based on previous research, there is strong evidence that L2 fluency improves in a study abroad context (e.g., Du, 2013; Freed, 1995; Mora & Valls-Ferrer, 2012; Segalowitz & Freed, 2004; Serrano, Llanes, & Tragant, 2011). Additionally, both cross-linguistic similarities and differences, as well as individual L1 fluency behavior, appear to have an effect on the

strength of the relationship between L1 and L2 fluency. Thus, in order to understand the effects of cross-linguistic differences, it is also necessary to compare the fluency behavior of native speakers of the languages.

The research questions guiding the current research are as follows:

1. To what extent do native speakers of English, Spanish, and French differ with regard to measures of speed, breakdown, and repair fluency?
2. To what extent is there a relationship between L1 and L2 fluency behavior before and after 5-months' residence abroad?
3. To what extent can L2 fluency measures be predicted from L1 fluency behavior, target language group (Spanish or French), and proficiency before and after 5-months' residence abroad?

Based on the review of the literature, several predictions can be made. First, cross-linguistic differences between English, Spanish, and French NSs are predicted, especially regarding measures of speed fluency (i.e., mean syllable duration). It is possible that frequency and duration of pauses as well as repetitions may differ, given that these aspects of fluency have been shown to differ in previous studies. Second, correlations between L1 and L2 fluency measures are also predicted (De Jong et al., 2015). Finally, given that no previous research in L2 fluency has investigated the relative contributions of L1 fluency behavior, cross-linguistic differences, and proficiency in predicting L2 fluency behavior in a single model, it is unclear what those relative contributions will be, and whether they shift over time. Nevertheless, the findings of Derwing et al. (2009) in which significant correlations between L1 and L2 fluency measures were maintained over time for Slavic L1 learners of English but not Mandarin L1 learners, might

suggest that the contribution of cross-linguistic differences may differ as proficiency changes. It is also predicted that for those measures that differ between native speakers, the addition of the target language group variable will significantly improve models predicting L2 fluency behavior whereas for those that are similar it will not (De Jong et al., 2015).

Method

Participants

The participants come from the longitudinal Languages and Social Networks Abroad Project (LANGSNAP: Mitchell, Tracy-Ventura, & McManus, 2017) which investigates language learning before, during, and after residence abroad. They were undergraduate students from a university in the United Kingdom (UK) who were undertaking French (n = 29) or Spanish (n = 27) degrees and were required to spend their third year (of a four-year degree) abroad. Because the current study compares L1 and L2 fluency, participants whose L1 was not English were excluded (n = 5). An additional two participants were excluded because of low quality or missing recordings.

Of the remaining 24 participants in the Spanish learner group, more than half spent the academic year in Spain (n = 15) and the rest in Mexico (n = 9). Seven participants were male and 17 were female. Their mean age at the start of data collection was 21 (SD = 1.2 years) and the mean length of studying Spanish was 6 years (SD = 3.2 years). Of the remaining 25 participants in the French learner group, all spent the academic year in France. Two participants were male and 23 were female. Their mean age at the start of data collection was 20 (SD = 0.7 years) and the mean length of studying French was 11 years (SD = 2.4 years). Unlike more traditional 'study abroad' programs where students go abroad and take classes, students in the UK have a

choice of placement type while abroad: as Erasmus exchange students at a partner university, as English teaching assistants, or on other forms of work placement (Mitchell, McManus, & Tracy-Ventura, 2015; Meara, 1994). Therefore, their day-to-day experiences differed depending on the placement type they chose (see Mitchell, Tracy-Ventura, & McManus, 2017 for more details). Adult native speakers of Spanish (n =18) and French (n =10)[1] who were age matched to the learner group were also recruited to complete the narration tasks in their L1 for cross-linguistic comparison purposes.

Instruments

Speaking tasks. Two picture-based narratives were used in this study: the Cat Story, borrowed from Dominguez, Tracy-Ventura, Arche, Mitchell, and Myles (2013) and based on Langley (2000), and the Brothers Story created for the larger project based on the children's book *I Very Really Miss you* (Langley, 2006). Each story began with a prompt. The Cat Story, for example, depicts the story of a little girl and her cat and what happens one day when the cat goes missing. It begins with a prompt *Todas las mañanas eran iguales*/*Tous les matins étaient pareils* ('Every morning was the same') which signals a series of background information about what the little girl and her cat used to do every day. A few pages later the phrase *hasta que un día*/*Mais il est arrivé un jour* ('until one day') appears in order to signal that the main events of the story are about to begin. The English version of the Cat Story was also administered. The only change was that the prompts were provided in English. All participants were given approximately one minute of planning time to look through the pages and get a sense of the full

story before orally retelling the story. While narrating the story, participants were able to continue looking at the pages.

Elicited imitation test. The measure of proficiency adopted in the current study was an elicited imitation test (EIT). An EIT requires test takers to listen to test stimuli and repeat them orally as accurately as possible. The idea behind an EIT as a measure of L2 proficiency is that learners can only imitate sentences that they have both parsed and comprehended (Bley-Vroman & Chaudron, 1994). EITs, particularly those with a range of sentence lengths, have been found to be valid and reliable measures of L2 proficiency (Yan, Maeda, Lv, & Ginther, 2015). The Spanish EIT used in the current study was designed by Ortega (2000). Parallel versions of this EIT were also created in English, German, and Japanese for a cross-linguistic study of syntactic complexity measures and their relationship to L2 proficiency (Ortega, Iwashita, Rabie & Norris, 2002). Versions are also available in Chinese (Wu & Ortega, 2013) and Korean (Kim, Tracy-Ventura, & Jung, 2016). The test sentences of the Spanish EIT were constructed to include high frequency vocabulary items, a range of syntactic complexity, and typical grammatical features known to challenge instructed learners. A French version of the EIT, translated based on the English stimuli, was created for the larger study from which the data in the current study come (Tracy-Ventura, McManus, Norris, & Ortega, 2014). The Spanish and French versions of the EIT used in the current study have nearly the same stimuli; the French and Spanish versions both include 30 test sentences. The French sentences range from 7-19 syllables, and the Spanish sentences range from 7-17 syllables. This difference in the highest number of syllables is not a major concern in the comparability of the two tests, however, because item difficulty analyses demonstrated that the items with the highest number of syllables are not necessarily the most

difficult (see Tracy-Ventura et al., 2014). The Spanish EIT was also used in Bowden (2016) and is available, along with the French EIT, on IRIS (https://www.iris-database.org/iris/app/home/index).

Procedures

Data in the longitudinal investigation were collected six times over an almost two-year period (May 2011-February 2013) including three data collection points during a nine-month stay abroad when participants were visited in situ by a member of the research team; however, in the current analysis we focus on a subset of that data, from the pretest and visit 2 abroad, because these data collection rounds included the proficiency test, whereas the other two visits abroad did not. The L2 version of the Cat Story was administered at the pretest (May 2011) and the L2 version of the Brothers Story was administered at visit 2 (February 2012). The same EIT was administered at both the pretest and visit 2. Because a second version of the EIT with different test items does not exist, the decision was made to administer the EIT at visit 2 (February 2012) abroad instead of visit 1 (November 2011) to increase the time between administrations to nine months in an effort to minimize any practice effects. The L1 version of the Cat Story was administered at the last data collection round, eight months after students had returned from their stay abroad (February 2013).

Data Analysis

Speaking tasks. Data included 8 hours and 55 minutes of audio-recorded narrations. Participants were not given a time limit and thus the recording length varied among them. Table

1 presents the means and standard deviations of the length of the recordings in the L1 (English, Spanish, or French) and L2 (Spanish or French).

[Insert Table 1 here]

All data were first transcribed according to the Codes for Human Analysis of Transcripts (CHAT) conventions for later use with the Computerized Language Analysis (CLAN) program (MacWhinney, 2000). After the transcriptions were completed, they were checked for accuracy by another member of the research team. Transcriptions were coded for filled pauses (e.g., *um*, *uh*), repetitions (i.e., when a speaker repeated a word or phrase verbatim), and corrections (i.e., when a speaker modified a previously spoken word or phrase). Syllables for each of the language data sets (English, French, and Spanish) were counted by research assistants, and a subset of each set was also coded by the authors. Syllables were counted by hand based on the transcripts for the Spanish and English data and based on the transcripts and audio for the French data because of the possibility of differences in production of words like *finalement* (three vs. four syllables).[2] Inter-rater reliability was calculated between the two ratings for each of the data sets and the resulting Cohen's Kappa was 0.99 (English), 0.99 (Spanish), and 0.99 (French), demonstrating very good agreement.

Audio data were then automatically segmented using Praat's *Annotate to Text Grid with silences* feature (Boersma & Weenink, 2010). The minimum lower bound for silent pauses was set to 250 ms (De Jong & Bosker, 2013; Kahng, 2014). As stated previously, a lower bound of 250 ms for silent pauses (as opposed to the 400 ms used in Derwing et al., 2009) was chosen because of research (De Jong & Bosker, 2013; Kahng, 2014) demonstrating this as an optimal threshold for correlation to L2 proficiency. Next, accuracy of the segmentation was checked

manually by the first author and adjustments were made.[3] The second author manually checked 20% of the data. Inter-rater reliability was calculated between the two raters and the resulting Cohen's Kappa was 0.92, demonstrating very good agreement; thus, the first author manually checked the remaining data.

Utterance fluency in the current analysis is investigated using the subcategories proposed by Skehan (2003) and Tavakoli and Skehan (2005): speed, breakdown, and repair fluency. Following De Jong et al. (2015), seven measurements of fluency were calculated, as seen in Table 2.

[Insert Table 2 here]

Elicited imitation test. The EIT was scored using the 5-point rubric (0-4) outlined in Ortega (2000). A score of 4 was given for exact repetition, 3 for repetitions that preserve the original meaning of the stimulus but include small changes in grammar, 2 for repetitions that are meaningful but depart slightly from the original meaning of the stimulus, 1 for repetitions that are missing important content from the original stimulus and may also be incomplete sentences, and 0 for repetitions that include minimal to none of the original content of the stimulus. The total possible score on the EIT was 120 (30 items x 4 points each).

Results

Before addressing the research questions, descriptive statistics are presented for the measure of proficiency adopted in the current study, the elicited imitation test, and the measures of utterance fluency. Learners in the Spanish group (n = 24) scored a mean of 83.96 (SD = 11.73) at the pretest on the proficiency test and a mean of 98.92 (SD = 7.72) at visit 2. As the

data were normally distributed and contained no outliers, a paired-samples *t*-test was conducted and indicated significant improvement for the Spanish group from pretest to visit 2, $t(23) = 7.97$, $p < .001$, $d = 1.51$. Learners in the French group ($n = 25$) scored a mean of 59.28 (SD = 16.04) at the pretest on the proficiency test and a mean of 76.56 (SD = 14.74) at visit 2.[4] A paired-samples *t*-test indicated significant improvement for the French group from pretest to visit 2, $t(24) = 9.18$, $p < .001$, $d = 1.12$.

Table 3 provides the mean, standard deviation, and median of the speed, breakdown, and repair fluency measurements for the pretest, visit 2, and English data for both the Spanish and French groups, and the L1 data from the Spanish and French native speakers. As shown in Table 3, the English L1 fluency measures of the two learner groups appear to be comparable.[5] Table 4 provides the mean, standard deviation, and median of the changes in L2 fluency from the pretest to visit 2 for both the Spanish and French groups. Non-parametric tests were conducted to determine whether changes over time on the L2 fluency measures between the pretest and visit 2 were significant as not all data were normally distributed and several measures contained outliers. As shown in Table 4, the results of Wilcoxon signed-rank tests indicated statistically significant gains in three measures for both the Spanish and French learner groups: mean syllable duration, mean silent pause duration within ASU, and the number of silent pauses per second. Both learner groups also showed significant differences in the number of corrections per second, but the Spanish group showed an increase whereas the French group showed a decrease. For the number of filled pauses per second, the Spanish group showed a significant decrease in filled pauses, but there was no significant difference for the French group. Finally, neither group showed significant differences for the average length of silent pauses between ASU or the number of repetitions per second.

[Insert Table 3 here]

Before running inferential statistics to test whether L1 measures of fluency were comparable between English, French, and Spanish (research question one) and whether L1 and L2 measures of fluency correlated for the learners (research question two), parametric test assumptions (e.g., normality, existence of outliers) were first checked. For example, the data were inspected visually and Shapiro-Wilk normality tests indicated that of the seven sets of fluency measures (see Table 3), six contained at least one subset (pretest, visit 2, English) that failed the Shapiro-Wilk test for the French learner group data and the same was true for five sets of the Spanish learner group data. For the Spanish and French NSs, three measures failed the Shapiro-Wilk test. In addition, the data contained outliers in some of the measures: six of the 21 variables for the French learner group and 10 of the 21 variables for the Spanish learner group (many of which were in the English data), and four of the 14 variables in the Spanish and French NS data. In an attempt to obtain normally-distributed data so as to not violate assumptions of the parametric tests, a variety of transformations (e.g., Log [x+1] and SQRT[x]) were conducted where appropriate given a particular measure (Larson-Hall, 2010); however, none of the transformations resulted in normally distributed data nor did they affect outliers. Thus, the original data were retained and non-parametric tests were conducted. The alpha level was set at $p < .05$ and effect sizes are reported using the absolute value of *d*. These were interpreted following Plonsky and Oswald (2014) for within-group contrasts: $d = 0.60$ (small), $d = 1.00$ (medium), and $d = 1.40$ (large); between-group contrasts: $d = 0.40$ (small), $d = 0.70$ (medium),

and d = 1.00 (large); and correlation coefficients: r = .25 (small), r = .40 (medium), and r = .60 (large).

[Insert Table 4 here]

Research question one examined to what extent there were cross-linguistic differences between NSs of English, Spanish, and French with regard to measures of speed, breakdown, and repair fluency. Results of the Kruskal-Wallis tests indicated statistically significant differences between the NS groups on four of the seven temporal fluency measures: Mean syllable duration (speed fluency), the number of filled pauses per second (breakdown fluency), and both the number of repetitions and corrections per second (repair fluency). Table 5 lists the chi square test statistics and their associated p values for each of the seven measures. Post hoc Mann-Whitney U tests were conducted on those four measures with alpha levels adjusted to $p < .0167$. Table 5 lists the U and z test statistics and their associated p values and effect sizes for each of the four measures (Larson-Hall, 2010).

[Insert Table 5 here]

Results from the Mann-Whitney U tests indicated significant differences and large effect sizes between all language pairings for mean syllable duration with English speakers having a longer mean syllable duration than French (z = 3.78, $p < .001$, d = 1.32) and Spanish speakers (z = −6.07, $p < .001$, d = 2.47), and French speakers having a longer mean syllable duration than Spanish speakers (z = −2.97, p = .002, d = 1.33). For the measure of the number of filled pauses

per second, results indicated significant differences and large effect sizes between Spanish and both English ($z = -3.05$, $p = .002$, $d = 0.98$) and French ($z = -2.79$, $p = .004$, $d = 1.57$), such that Spanish speakers produced fewer filled pauses than French or English speakers. No significant differences were found between French and English ($p = .321$, $d = 0.22$). For the number of repetitions per second, results indicated significant differences with a medium effect size between Spanish and English ($z = 2.68$, $p = .007$, $d = 0.76$), such that Spanish speakers made fewer repetitions than English speakers. No significant differences were found between Spanish and French speakers ($p = .759$, $d = 0.14$) or English and French speakers ($p = .063$, $d = 0.79$). For the number of corrections per second, results indicated significant differences between English and both Spanish ($z = 2.03$, $p = .043$, $d = 0.50$) and French ($z = -2.51$, $p = .012$, $d = 0.88$) with a small and medium effect size, respectively, such that English speakers made more corrections than Spanish or French speakers. No significant differences were found between Spanish and French ($p = .356$, $d = 0.43$).

Research question two examined to what extent there is a relationship between temporal measures of fluency in the L1 and L2 before and after 5-months' residence abroad. To address this question, Spearman's correlations were conducted for the seven temporal measures between the L1 and the L2 at the pretest and at visit 2 for both the French and Spanish groups (see Table 6).

[Insert Table 6 here]

As can be seen in Table 6, the results for both groups demonstrate positive correlations between the L1 and the L2 fluency measures (except the number of corrections per second for

the French group at the pretest, $r = -0.07$) that range from weak to strong, $r = 0.08$ to $r = 0.72$, yet only a subset of these correlations are significant. For the Spanish group, the same three fluency measures are significant at both the pretest and visit 2: mean syllable duration (pretest: $r = 0.55$, visit 2: $r = 0.66$), mean silent pause duration between ASU (pretest: $r = 0.48$, visit 2: $r = 0.60$), and number of silent pauses per second (pretest: $r = 0.42$, visit 2: $r = 0.46$). In contrast, for the French group, whereas only two of the correlations are significant at the pretest (MSD, $r = 0.49$ and number of silent pauses per second, $r = 0.72$), five of the seven measures correlate significantly at visit 2 (all but mean silent pause duration between ASU and number of corrections per second).

Research question three examined to what extent L2 measures of fluency can be predicted from L1 measures of fluency, target language group, and proficiency before and after 5-months' residence abroad. To address this question, multiple regressions were conducted (variables entered simultaneously) with the L2 fluency measure as the dependent variable and the L1 fluency measure, target language group (Spanish or French, coded as a dummy variable), and proficiency (EIT score) as independent variables. Assumptions for multiple regression (e.g., normal distribution of residuals, independence of observations, linearity between dependent and independent variables, homoscedasticity) were checked and met. Results from the multiple regressions at the pretest are presented in Table 7 and those from visit 2 are presented in Table 8.

[Insert Table 7 here]

At the pretest, models predicting L2 fluency measures from L1 fluency measures, target language group, and proficiency were significant for mean syllable duration, $F(3, 45) = 8.76$, $p <$

.001 (speed fluency), and two measures of breakdown fluency: mean silent pause duration within ASU, $F(3, 45) = 4.65$, $p = .007$, and the number of silent pauses per second, $F(3, 45) = 9.16$, $p < .001$. The amount of variance explained for mean syllable duration was 33%, for mean silent pause duration within ASU was 19%, and for the number of silent pauses per second was 34%. For mean syllable duration, the L1 fluency measure and proficiency significantly contributed to the model, $p < .001$ and $p = .013$, respectively, and target language group approached significance, $p = .059$. For mean silent pause duration within ASU, target language group and proficiency significantly contributed to the model, $p = .023$ and $p = .002$, respectively, whereas the L1 fluency measure did not, $p = .325$. For the number of silent pauses per second the L1 fluency measure significantly contributed to the model, $p < .001$, whereas target language group and proficiency did not, $p = .899$ and $p = .209$, respectively.

[Insert Table 8 here]

At visit 2, models predicting L2 fluency measures from L1 fluency measures, target language group, and proficiency were significant for all seven temporal measures of fluency, with the amount of variance explained ranging from 13% − 47%. In contrast to results from the pretest, the L1 fluency measure contributed significantly to each model, whereas proficiency did not significantly contribute to any model. Target language group significantly contributed to mean syllable duration, the number of filled pauses per second, the number of repetitions per

second and the number of repairs per second. Note that these were the same measures that differed among the native speakers.

Discussion

The purpose of this study was to investigate the effects of individual differences in L1 fluency behavior, cross-linguistic differences, and proficiency on L2 fluency behavior over time. Results indicated that both groups' L2 proficiency and several measures of L2 fluency significantly improved after 5-months' residence abroad and the effect sizes were large, which supports previous research demonstrating development as a result of residence/study abroad (e.g., Du, 2013; Freed, 1995; Mora & Valls-Ferrer, 2012). It was also the case that English, Spanish, and French NSs differed significantly on some measures of fluency: mean syllable duration, number of filled pauses per second, number of repetitions per second, and number of corrections per second, but not average silent pause duration within or between ASU or the number of silent pauses per second. In the current study, we found that English NSs had a longer mean syllable duration than Spanish and French NSs, and French NSs, in turn, had a longer mean syllable duration than Spanish NSs. These results are in line with the syllable structures of these given languages and previous work indicating cross-linguistic differences regarding speech rate with other languages (e.g., De Jong et al., 2015; Roach, 1998). Additionally, results of the current study demonstrated that Spanish NSs produced fewer filled pauses and repetitions than English NSs, and that English NSs used more corrections than Spanish or French NSs. Unlike the results from Grosjean and Deschamps (1975), it was not the case that differences were found between English and French NSs with regard to pause frequency. One possible explanation for this may be a result of different tasks: the current study examined data from picture-based

narratives whereas Grosjean and Deschamps examined interview data. Task type has been shown to influence temporal measures of fluency (Préfontaine & Kormos, 2015), including articulation rate and pause duration. The fact that cross-linguistic differences were found between the three languages supports the idea that the fluency characteristics of the languages under investigation might influence the strength of the relationship between L1 and L2 fluency. This finding is addressed in research question three.

The second research question examined to what extent there is a relationship between L1 and L2 temporal measures of fluency before and after 5-months' residence abroad. Only two of the seven temporal measures of fluency, mean syllable duration and the number of silent pauses per second, correlated at the pretest (and visit 2) for both groups. For the Spanish group, the average pause duration between ASU was also significant at the pretest and visit 2 but no other measures were significant at visit 2. For the French group, all but two measures, average pause duration between ASU and the number of corrections per second, were significant at visit 2. These results are in contrast to those presented in De Jong et al. (2015) who found significant correlations between all measures in the L1 and the L2 and Derwing et al. (2009), who found significant L1-L2 correlations at time 2 for all three of their measures. One explanation for the current findings could be related to the exposure that participants in each of the studies had at the time of testing. Participants in De Jong et al. had already resided abroad for an average of 4.5 years compared to participants in Derwing et al. who had been residing abroad for an average of 4.8 months. The participants in the current study had never resided abroad before the pretest for any significant amount of time. It is important to note that significant correlations between L1 and L2 fluency were found for measures of mean syllable duration and pausing frequency from early on, and this result parallels previous findings like those in Derwing et al. and Towell and

Dewaele (2005). Derwing et al. did not find significant L1-L2 correlations for the Mandarin group at times 6 and 7 (although significant correlations were found for learners in the Slavic group); however, this result might be explained by the fact that learners in the Mandarin group did not show improvement between the two time points whereas learners in the Slavic group and learners in the current study did improve over time.

The third research question examined to what extent L2 measures of fluency can be predicted from L1 measures of fluency, target language group, and proficiency before and after 5-months' residence abroad. Results indicated that at the pretest, only models predicting L2 mean syllable duration, mean silent pause duration within ASU, and the number of silent pauses per second were significant; however, L1 fluency only contributed to models of mean syllable duration and the number of silent pauses per second. In contrast, at visit 2 L1 fluency significantly contributed to all models, target language group contributed to those models where cross-linguistic differences were found between native speakers, and proficiency contributed to none. These findings suggest that ultimately L2 fluency behavior and its relationship to L1 fluency behavior are mitigated by both cross-linguistic differences and proficiency and that the relative explanatory power of these factors shifts with language experience. The contribution of cross-linguistic differences changes between pretest and visit 2, with target language group at visit 2 contributing to those models only in which differences were found between NS groups. In other words, when predicting L2 fluency behavior at the pretest, not much information is gained by knowing whether these English speakers are learning French or Spanish. At visit 2, however, knowing the language an English speaker is learning significantly contributes to explaining variation in those measures that differ between English, Spanish, and French native speakers. Although not highlighted in their paper, these results parallel those from De Jong et al. (2015)

who found that language group contributed to a best fit model of the three fluency measurements that differed between Turkish and English native speakers (mean syllable duration, number of silent pauses per second, and number of repetitions per second). Based on these results, it seems possible that cross-linguistic differences in fluency characteristics between a learner's L1 and L2 could affect the amount of gains that can be made. For example, L1 English speakers who are learning Spanish have more room to develop in the area of mean syllable duration than if they were learning French because the difference between English and Spanish is greater than between English and French. Future research is needed to investigate this claim.

The finding that L1 fluency behavior and cross-linguistic differences contribute to explaining more variation in L2 fluency behavior as L2 fluency improves is perhaps explained by processing difficulties at lower proficiency levels. It might be the case that at the pretest, learners' weaker L2 cognitive fluency contributed more to explaining variation in L2 utterance fluency than the factors investigated in the current study. Although we did not include any tests of cognitive fluency to examine, for example, whether participants became more efficient at making word-meaning links, given that at visit 2 they had been residing abroad for five months and practicing the L2 regularly, their L2 cognitive processing skills likely became more automatized as proficiency improved. As the difference between their L1 and L2 cognitive fluency decreased, it became possible to observe the influence of other variables on the participants' L2 utterance fluency, including L1 fluency behavior and cross-linguistic

differences. To empirically test this claim, future research is needed which includes tests of cognitive fluency as well.

Conclusion

Some limitations of the current study should be acknowledged. First, the same measure of proficiency, the elicited imitation test, was administered at both the pretest and visit 2 abroad. Although there were nine months between administrations of the test, it is possible that there may have been some practice effects. Additionally, findings from the current study are based on comparisons of English, Spanish, and French only. While evidence was found for cross-linguistic differences with regard to temporal measures of fluency, more work investigating cross-language differences of these features is needed with a wider variety of combinations of L1s and L2s and larger sample sizes to determine whether these results hold across language pairs. Furthermore, while it was demonstrated that learners' proficiency and several measures of L2 fluency improved between the pretest and visit 2 abroad, the current study did not aim to explain what might account for these gains and whether L1 fluency behavior, target language group, and proficiency at pre-departure might play a role. Thus, it may be beneficial to include these variables in future research investigating L2 fluency development. A final direction for future work is to investigate the relationship between L1 and L2 fluency in a variety of tasks (Derwing, Rossiter, Munro, & Thompson, 2004; Foster & Skehan, 1996; Tavakoli & Skehan, 2005). The current study used a picture-based oral narration. Future research should include additional tasks such as those that elicit spontaneous speech (e.g., an interview or conversation).

Ultimately, incorporating results from a variety of task types will provide a more complete picture of the relationship between L1 and L2 fluency and how task type might mitigate results.

These limitations notwithstanding, results of this study have demonstrated that variation in L2 fluency behavior is not only explained by individual differences in L1 fluency behavior, but is also related to other factors such as proficiency and cross-linguistic differences. Perhaps more importantly, the findings indicated that these factors differentially contribute to explaining variance in L2 utterance fluency between pre-immersion and after 5-months' residence abroad. Thus, it may be worthwhile for future research to shift the focus beyond investigating *if* there is a relationship between L1 and L2 fluency, to rather how that relationship shifts as development occurs. In sum, these results suggest that factors influencing L2 fluency behavior are dynamic, and as such, L2 models of speech production should be equipped to explain these shifts.

Notes

1 It may be noted that the number of the participants for the French native speaker group is smaller than that of the other two native speaker groups; however, the standard deviations of the fluency measures reported in Table 3 indicate that the French native speaker group is relatively homogenous and shows similar standard deviations as the Spanish native speaker group. Control groups of native speakers are often smaller given their homogeneity.

2 To test whether using the audio and transcript for the French data rather than just the transcript alone resulted in significantly different syllable counts, a subset of that data (~10%) was later counted by the first author on the basis of the transcript alone. Inter-rater reliability was calculated between the two rating types and the resulting Cohen's Kappa was 0.99, demonstrating very good agreement.

3 Because data collection occurred on site abroad, finding a quiet place to record was sometimes difficult. Therefore, the sound quality of the recordings varied which affected how well the automated script was able to detect silences. Manually checking the script's output was thus necessary.

4 An anonymous reviewer questioned whether differences in proficiency between the two learner groups as measured by the EIT were true differences in proficiency or differences due to the task. Based on Ortega et al. (2002)'s cross-linguistic study of four different language EITs, it is unlikely that the raw EIT scores are directly comparable across languages. Additionally, even though the participants come from the same institution and are at the same institutional level, it

does not necessarily mean they would be at exactly the same proficiency level (see Callies, Díez-Bedmar, & Zaytseva, 2014). For the current analysis, what is most important is that each group's proficiency improves over time.

5 Results of non-parametric Mann-Whitney U tests indicated no statistically significant differences between learner groups for any of the English L1 temporal fluency measures.

References

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.

Bley-Vroman, R., & Chaudron, C. (1994). Elicited imitation as a measure of second-language competence. *Research methodology in second-language acquisition*, 245-261.

Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer (Version 5.2.04) [Computer program]. Retrieved from http://www.praat.org/

Bosker, H. R., Pinget, A. F., Quené, H., Sanders, T., & De Jong, N. H. (2012). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, *30*, 159-172.

Bowden, H. W. (2016). Assessing second language oral proficiency for research. *Studies in Second Language Acquisition*, 1-29. doi: http://dx.doi.org/10.1017/S0272263115000443

Callies, M., Díez-Bedmar, M. B., & Zaytseva, E. (2014). Using learner corpora for testing and assessing L2 proficiency. In P. Leclercq, A. Edmonds, & H. Hilton (Eds.), *Measuring L2 proficiency: Perspectives from SLA* (pp. 71–90). Bristol, UK: Multilingual Matters.

Cucchiarini, C., Strik, H., & Boves, L. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *The Journal of the Acoustical Society of America, 107*, 989-999.

Cucchiarini, 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.

De Bot, K. (1992). A bilingual production model: Levelt's "Speaking" model adapted. *Applied Linguistics, 13,* 1-24.

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

De Jong, N. H., Schoonen, R., & Hulstijn, J. H. (2009). Fluency in L2 is related to fluency in L1. Paper presented at the Seventh International Symposium on Bilingualism (ISB7), Utrecht, The Netherlands.

De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. N. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics, 34,* 893-916.

De Jong, N. H., & Bosker, H. R. (2013). Choosing a threshold for silent pauses to measure second language fluency. *Proceedings of DiSS*, 17-20.

Derwing, T., Munro, M., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition, 31,* 533-557.

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language Learning, 54,* 655-679.

Dominguez, L., Tracy-Ventura, N., Arche, M., Mitchell, R., & Myles, F. (2013). The role of dynamic contrasts in the L2 acquisition of Spanish past tense morphology. *Bilingualism: Language and Cognition, 16*, 558-577.

Du, H. (2013). The development of Chinese fluency during study abroad in China. *The Modern Language Journal, 97*, 131-143.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition, 18,* 299-323.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21,* 354-375.

Freed, B. F. (1995). What makes us think that students who study abroad become fluent. *Second language acquisition in a study abroad context, 9,* 123-148.

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.

Grosjean, F., & Deschamps, A. (1975). Analyse contrastive des variables temporelles de l'anglais et du français: vitesse de parole et variables composantes, phénomènes d'hésitation. *Phonetica, 31,* 144-184.

Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning, 64,* 809-854.

Kim, Y., Tracy-Ventura, N., & Jung, Y. (2016). A measure of proficiency or short-term memory? Validation of an elicited imitation test for SLA research. *Modern Language Journal, 100,* 655-673.

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System, 32,* 145-164.

Langley, J. (2000). *Missing*. London: Francis Lincoln.

Langley, J. (2006). *I Very Really Miss You.* (illustrations). London: Francis Lincoln.

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York, NY: Routledge.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40*, 387-417.

Levelt, W. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Levelt, W. (1999). Producing spoken language: A blueprint of the speaker. In C. Brown and P. Hagoort (Eds.), *The neurocognition of language* (pp. 83-122). Oxford, UK: Oxford University Press.

MacWhinney, B. (2000). The CHILDES Project: Tools for Analyzing Talk. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Meara, P. (1994). The year abroad and its effects. *Language Learning Journal*, *10*, 32-38.

Mitchell, R., McManus, K., & Tracy-Ventura, N. (2015). Comparison of language development during different residence abroad programmes. In R. Mitchell, N. Tracy-Ventura, & K. McManus (Eds), *Social interaction, identity and language learning during residence abroad* (pp.115-138). EUROSLA Monograph Series 4.

Mitchell, R., Tracy-Ventura, N., & McManus, K. (2017). *The Anglophone student abroad: Identity, social relationships and language learning*. Routledge.

Mora, J. C., & Valls-Ferrer, M. (2012). Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly*, *46*, 610-641.

Ortega, L. (2000). *Understanding syntactic complexity: The measurement of change in the syntax of instructed L2 Spanish learners*. Unpublished Doctoral dissertation, University of Hawaii, Honolulu, HI.

Ortega, L., Iwashita, N., Norris, J. M., & Rabie, S. (2002, October). An investigation of elicited imitation tasks in crosslinguistic SLA research. In *Second Language Research Forum, Toronto*.

Pawley, A., & Syder, F. (2000). The one clause at a time hypothesis. In H. Riggenbach (Ed.), *Perspectives on Fluency* (pp. 163–191). Ann Arbor: University of Michigan Press.

Pellegrino, F., Coupé, C., & Marsico, E. (2011). A cross-language perspective on speech information rate. *Language, 87,* 539-558.

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878-912.

Préfontaine, Y., & Kormos, J. (2015). The relationship between task difficulty and second language fluency in French: A mixed methods approach. *The Modern Language Journal, 99, 1,* 96-112.

Riazentseva, A. (2001). Second language proficiency and pausing: A study of Russian speakers of English. *Studies in Second Language Acquisition, 23,* 497-526.

Roach, P. (1998). Some languages are spoken more quickly than others. In L. Bauer & P. Trudgill (Eds.) *Language Myths*. London, UK: Penguin.

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,* 173-199.

Serrano, R., Llanes, À., & Tragant, E. (2011). Analyzing the effect of context of second language learning: domestic intensive and semi-intensive courses vs. study abroad in Europe. *System, 39,* 133-143.

Skehan, P. (2003). Task-based instruction. *Language Teaching, 36,* 1-14.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239-276). Amsterdam: John Benjamins.

Towell, R., & Dewaele, J. M. (2005). The role of psycholinguistic factors in the development of fluency amongst advanced learners of French. In J. M. Dewaele (Ed.), *Focus on French as a foreign language* (pp. 210-239). Clevedon, UK: Multilingual Matters.

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics, 17,* 84-119.

Tracy-Ventura, N., McManus, K., Norris, J., & Ortega, L. (2014). "Repeat as much as you can": Elicited imitation as a measure of oral proficiency in L2 French. In P. Leclercq, A. Edmonds & H. Hilton (Eds.), *Measuring L2 Proficiency: Perspectives from SLA* (pp 143-166). Bristol: Multilingual Matters.

Trouvain, J., & Möbius, B. (2014). Sources of variation of articulation rate in native and non-native speech: comparisons of French and German. *Proceedings of Speech Prosody (SP7)*, 275-279.

Wu, S.-L., & Ortega, L. (2013). Measuring global oral proficiency in SLA research: A new elicited imitation test of L2 Chinese. *Foreign Language Annals, 46,* 680-704.

Yan, X., Maeda, Y., Lv, J., & Ginther, A. (2015). Elicited imitation as a measure of second language proficiency: A narrative review and meta-analysis. *Language Testing*. Epub ahead of print 17 August 2015. DOI: 10.1177/0265532215594643.

Table 1

Means (and standard deviations) of recording length separated by learner group and language

	Pretest (L2)	Visit 2 (L2)	Native Language
Spanish Learners (n = 24)	4 min 23 sec (SD: 1 min 47 sec)	3 min 05 sec (SD: 1 min 27 sec)	1 min 48 sec (SD: 53 sec)
French Learners (n = 25)	4 min 58 sec (SD: 1 min 47 sec)	2 min 56 sec (SD: 1 min 11 sec)	1 min 55 sec (SD: 36 sec)
Spanish Native Speakers (n = 18)	NA	NA	2 min 25 sec (SD: 47 sec)
French Native Speakers (n = 10)	NA	NA	2 min 25 sec (SD: 1 min 07 sec)

Table 2

Fluency measure calculations

	Calculation
Speed Fluency	
Mean syllable duration	speaking time (excluding pauses) / number of syllables
Breakdown Fluency	
Mean silent pause duration	
Within ASU	duration of silent pauses within ASU / number of silent pauses within ASU
Between ASU	duration of silent pauses between ASU / number of silent pauses between ASU
Number of	
Silent pauses	number of silent pauses / speaking time (excluding pauses)
Filled pauses	number of filled pauses / speaking time (excluding pauses)
Repair Fluency	
Repetitions/second	number of repetitions / speaking time (excluding pauses)
Corrections/second	number of corrections / speaking time (excluding pauses)

huensch@usf.edu

Table 3

Means (standard deviations) and medians of temporal fluency measures from the pretest, visit 2, and English separated by learner group and L1 Spanish and L1 French from the native speaker groups

	Spanish Learners			French Learners			Native Speakers	
	L1	L2		L1	L2		L1	L1
	English	Pretest	Visit 2	English	Pretest	Visit 2	Spanish	French
Speed Fluency								
Mean syllable duration (ms)	238	350	246	243	350	310	163	195
	(41)	(85)	(59)	(40)	(56)	(58)	(19)	(28)
Median	227	340	234	228	364	306	159	189
Breakdown Fluency								
Mean silent pause duration								
Within ASU (ms)	527	803	632	570	804	664	551	600
	(99)	(158)	(117)	(131)	(179)	(172)	(71)	(167)
Median	516	770	609	568	750	610	554	582
Between ASU (ms)	958	1171	1199	1039	1217	1167	1010	1039
	(235)	(295)	(345)	(201)	(333)	(289)	(239)	(230)
Median	962	1123	1201	1074	1121	1125	959	998
Number of								
Silent pauses/second	0.53	0.76	0.63	0.50	0.80	0.64	0.50	0.49
	(0.18)	(0.18)	(0.17)	(0.13)	(0.17)	(0.19)	(0.11)	(0.05)
Median	0.52	0.76	0.60	0.46	0.81	0.59	0.51	0.48
Filled pauses/second	0.08	0.29	0.13	0.09	0.28	0.24	0.03	0.09
	(0.07)	(0.17)	(0.11)	(0.08)	(0.13)	(0.13)	(0.02)	(0.06)
Median	0.06	0.35	0.10	0.05	0.27	0.26	0.02	0.10
Repair Fluency								
Repetitions/second	0.02	0.08	0.08	0.01	0.09	0.10	0.04	0.04

	(0.03)	(0.05)	(0.06)	(0.02)	(0.06)	(0.05)	(0.03)	(0.04)
Median	0.02	0.07	0.07	0.01	0.09	0.10	0.03	0.04
Corrections/second	0.02	0.07	0.09	0.02	0.07	0.05	0.03	0.04
	(0.02)	(0.03)	(0.03)	(0.02)	(0.04)	(0.04)	(0.02)	(0.02)
Median	0.02	0.06	0.08	0.02	0.06	0.04	0.03	0.03

huensch@usf.edu

Table 4

Changes in L2 fluency for the Spanish and French groups

	Spanish Learners		French Learners	
	L2 Fluency Change	Pairwise Comparison	L2 Fluency Change	Pairwise Comparison
Speed Fluency				
Mean syllable duration (ms)	−104	$z = -4.29$, $p < .001$*,	−41	$z = -3.59$, $p < .001$*,
	(52)	$d = 1.45$	(40)	$d = 0.71$
Median	−107		−45	
Breakdown Fluency				
Mean silent pause duration				
Within ASU (ms)	−172	$z = -4.03$, $p < .001$*,	−140	$z = -3.57$, $p < .001$*,
	(133)	$d = 1.24$	(146)	$d = 0.80$
Median	−170		−131	
Between ASU (ms)	29	$z = 0.40$, $p = .689$,	−50	$z = -0.71$, $p = .476$,
	(320)	$d = 0.09$	(278)	$d = 0.16$
Median	9		−71	
Number of				
Silent pauses/second	−0.13	$z = -3.74$, $p < .001$*,	−0.16	$z = -3.75$, $p < .001$*,
	(0.15)	$d = 0.77$	(0.14)	$d = 0.92$
Median	−0.11		−0.18	
Filled pauses/second	−0.16	$z = -4.06$, $p < .001$*,	−0.05	$z = -1.09$, $p = .276$,
	(0.13)	$d = 1.10$	(0.14)	$d = 0.36$
Median	−0.15		−0.01	
Repair Fluency				
Repetitions/second	−0.00	$z = -0.09$, $p = .932$,	0.01	$z = 0.42$, $p = .677$,
	(0.06)	$d = 0.02$	(0.05)	$d = 0.17$
Median	−0.01		0.00	
Corrections/second	0.02	$z = 2.63$, $p = .009$*,	−0.02	$z = -2.14$, $p = .032$*,

huensch@usf.edu

	(0.03)	$d = 0.74$	(0.03)	$d = 0.47$
Median	0.03		−0.02	

Table 5

Chi square test statistics and their associated p *values as a result of the Kruskal-Wallis tests and* U *and* z *test statistics and their associated* p *values and effect sizes as a result of the Mann-Whitney U tests*

	Comparison of L1	Post hoc Comparisons		
	Kruskal-Wallis	Mann-Whitney U test		
Speed Fluency		Spanish-English	Spanish-French	French-English
Mean Syllable Duration (ms)	$\chi^2(2) = 45.35$, $p < .001$	$U = 12.00$, $p < .001$ $z = -6.07$, $d = 2.47$	$U = 28.00$, $p = .002$ $z = -2.97$, $d = 1.33$	$U = 432.00$, $p < .001$ z = 3.78, $d = 1.32$
Breakdown Fluency				
Mean Silent Pause duration Within ASU (ms)	$\chi^2(2) = 1.12$, $p = .572$			
Between ASU (ms)	$\chi^2(2) = 5.82$, $p = .055$			
Number of Silent Pauses/second	$\chi^2(2) = 0.03$, $p = .984$			
Filled Pauses/second	$\chi^2(2) = 11.74$, $p = .003$	$U = 226.00$, $p = .002$ $z = -3.05$, $d = 0.98$	$U = 32.00$, $p = .004$ $z = -2.79$, $d = 1.57$	$U = 196.00$, $p = .321$ $z = -0.99$, $d = 0.22$
Repair Fluency				
Repetitions/second	$\chi^2(2) = 8.84$, $p = .012$	$U = 628.00$, $p = .007$ $z = 2.68$, $d = 0.76$	$U = 83.00$, $p = .759$ $z = -0.34$, $d = 0.14$	$U = 155.50$, $p = .063$ $z = -1.86$, $d = 0.79$
Corrections/second	$\chi^2(2) = 8.71$, $p = .013$	$U = 583.00$, $p = .043$ $z = 2.03$, $d = 0.50$	$U = 70.00$, $p = .356$ $z = -0.96$, $d = 0.43$	$U = 122.00$, $p = .012$ $z = -2.51$, $d = 0.88$

Table 6

Correlations between L1 fluency and L2 fluency measures for the Spanish group and the French group at the pretest and at visit 2

	Spanish		French	
	Pretest	Visit 2	Pretest	Visit 2
Speed Fluency				
Mean syllable duration (ms)	0.55*	0.66*	0.49*	0.58*
Breakdown Fluency				
Mean silent pause duration				
Within ASU (ms)	0.08	0.35	0.35	0.55*
Between ASU (ms)	0.48*	0.60*	0.32	0.37
Number of				
Silent pauses/second	0.42*	0.46*	0.72*	0.65*
Filled pauses/second	0.18	0.22	0.26	0.66*
Repair Fluency				
Repetitions/second	0.16	0.36	0.12	0.55*
Corrections/second	0.11	0.34	-0.07	0.33

Table 7

Results of the multiple regression models predicting L2 fluency at the pretest from L1 fluency, target language (TL) group, and proficiency

		Test Statistics					
		Unstandardized regression coefficient *B*	SE_b	Standardized coefficient β	*F*	*p*	*Adjusted* R^2
Speed Fluency							
Mean syllable duration (ms)	Intercept	292.135	81.552		$F(3, 45) = 8.76$	.000	.33
	L1 Measure	0.805	0.216	0.455*		.001	
	TL Group	−43.298	22.385	−0.309		.059	
	Proficiency	−1.591	0.617	−0.420*		.013	
Breakdown Fluency							
Mean silent pause duration							
Within ASU (ms)	Intercept	1138.961	192.310		$F(3, 45) = 4.65$	.007	.19
	L1 Measure	0.194	0.196	0.136		.325	
	TL Group	−136.734	57.902	−0.413*		.023	
	Proficiency	−5.219	1.615	−0.583*		.002	
Between ASU (ms)	Intercept	1186.005	334.280		$F(3, 45) = 1.96$	.134	.06
	L1 Measure	0.379	0.203	0.267		.068	
	TL Group	−95.254	117.406	−0.154		.421	
	Proficiency	−4.502	3.141	−0.270		.159	
Number of Silent pauses/second	Intercept	0.632	0.186		$F(3, 45) = 9.16$	.000	.34
	L1 Measure	0.576	0.147	0.514*		.000	
	TL Group	0.007	0.058	0.002		.899	
	Proficiency	−0.002	0.002	−0.224		.209	
Filled pauses/second	Intercept	0.424	0.136		$F(3, 45) = 1.45$	.241	.03
	L1 Measure	0.438	0.283	0.221		.129	

	TL Group	−0.061	0.057	−0.205		.290	
	Proficiency	−0.002	0.002	−0.244		.210	
Repair Fluency							
Repetitions/second	Intercept	0.043	0.049		$F(3, 45) = 0.85$	.475	−.01
	L1 Measure	0.472	0.341	0.211		.173	
	TL Group	0.024	0.021	0.231		.265	
	Proficiency	0.000	0.001	0.110		.583	
Corrections/second	Intercept	0.102	0.030		$F(3, 45) = 0.54$	.660	−.03
	L1 Measure	0.013	0.242	0.008		.957	
	TL Group	−0.008	0.012	−0.136		.501	
	Proficiency	0.000	0.000	−0.245		.221	

Table 8

Results of the multiple regression models predicting L2 fluency at visit 2 from L1 fluency, target language (TL) group, and proficiency

		Test Statistics					
		Unstandardized regression coefficient *B*	SE_b	Standardized coefficient β	*F*	*p*	*Adjusted* R^2
Speed Fluency							
Mean syllable duration (ms)	Intercept	148.688	71.193		$F(3, 45) = 8.76$	.000	.47
	L1 measure	0.782	0.171	0.483*		.000	
	TL Group	39.333	18.763	0.307*		.042	
	Proficiency	−0.891	0.582	−0.224		.133	
Breakdown Fluency							
Mean silent pause duration							
Within ASU (ms)	Intercept	535.705	194.062		$F(3, 45) = 3.45$	.024	.13
	L1 measure	0.499	0.171	0.399*		.006	
	TL Group	−27.181	54.609	−0.094		.621	
	Proficiency	−1.686	1.685	−0.187		.322	
Between ASU (ms)	Intercept	580.590	374.135		$F(3, 45) = 4.51$	.008	.18
	L1 measure	0.700	0.193	0.489*		.001	
	TL Group	−100.323	116.788	−0.161		.395	
	Proficiency	−0.523	3.562	−0.027		.884	
Number of Silent pauses/second	Intercept	0.311	0.204		$F(3, 45) = 10.32$	.000	.37
	L1 measure	0.734	0.136	0.630*		.000	
	TL Group	0.018	0.058	0.052		.751	
	Proficiency	−0.001	0.002	−0.064		.693	
Filled pauses/second	Intercept	0.068	0.141		$F(3, 45) = 7.68$	.000	.29
	L1 measure	0.743	0.214	0.423*		.001	
	TL Group	0.101	0.045	0.383*		.028	
	Proficiency	0.000	0.001	0.007		.969	

Repair Fluency							
Repetitions/second	Intercept	−0.001	0.063		$F(3, 45) = 4.19$	.011	.17
	L1 measure	1.021	0.328	0.422*		.003	
	TL Group	0.045	0.020	0.408*		.034	
	Proficiency	0.001	0.001	0.168		.364	
Corrections/second	Intercept	0.112	0.039		$F(3, 45) = 7.15$	.001	.28
	L1 measure	0.467	0.232	0.252*		.050	
	TL Group	−0.040	0.012	−0.560*		.002	
	Proficiency	0.000	0.000	−0.162		.351	