2014

Does Measuring L2 Utterance Fluency Equal Measuring Overall L2 Proficiency? Evidence From Five Languages

Wendy Baker-Smemoe
Brigham Young University, wendy_baker@byu.edu

Dan P. Dewey
Brigham Young University, ddewey@byu.edu

Jennifer Brown
Brigham Young University, jennifer_brown@byu.edu

Rob A. Martinsen
Brigham Young University, rob.martinsen@byu.edu

Original Publication Citation

Baker-Smemoe, W., Dewey, D., Bown, J., & Martinsen, R. (2014). Does measuring L2 utterance fluency equal measuring overall L2 proficiency? Evidence from five languages. Foreign Language Annals, 47, 707-728.

Foreign Language Annals
Published by the American Council on the Teaching of Foreign Languages

Does Measuring L2 Utterance Fluency Equal Measuring Overall L2 Proficiency? Evidence From Five Languages

Wendy Baker-Smemoe
Brigham Young University

Dan P. Dewey
Brigham Young University

Jennifer Bown
Brigham Young University

Rob A. Martinsen
Brigham Young University

Abstract: *The current study examined the relationship between overall second language (L2) proficiency and utterance fluency measures for several L2s in order to determine whether utterance measures can be used to predict L2 proficiency. The study measured the speech rate, number of hesitations, number and length of pauses, number and length of runs, and number of false starts using excerpts from 126 ACTFL Oral Proficiency Interviews (OPIs) spoken by 86 participants. Forty of the participants provided pre- and post-OPI speech samples, which also allowed examination of changes over time. All 86 participants were native English speakers who spoke L2 French, German, Japanese, Arabic, or Russian. They ranged in proficiency from Novice Mid to Superior. Results suggested that some L2 utterance fluency measures correlated significantly with overall L2 proficiency for all L2s, but data also revealed some differences*

Wendy Baker-Smemoe (PhD, University of Illinois) is Associate Professor in the Department of Linguistics and English Language, Brigham Young University, Provo, Utah.

Dan P. Dewey (PhD, Carnegie Mellon University) is Associate Professor in the Department of Linguistics and English Language, Brigham Young University, Provo, Utah.

Jennifer Bown (PhD, The Ohio State University) is Associate Professor of Russian, Brigham Young University, Provo, Utah.

Rob A. Martinsen (PhD, University of Texas at Austin) is Assistant Professor of Spanish Pedagogy at Brigham Young University, Provo, Utah.

Foreign Language Annals, Vol. 47, Iss. 4, pp. 707–728.
DOI: 10.1111/flan.12110

across L2s. These differences hinged partly on the L2's relative difficulty for native English speakers. Results suggested that it might be feasible to use specific fluency measures to estimate proficiency, in particular at higher levels, but that fine-grained sublevel estimates would not be recommended, in particular at the Novice and Intermediate levels.

Key words: *assessment, foreign/second language learning/acquisition, oral proficiency (OPI and OPIc), proficiency, quantitative research*

Introduction

Understanding the relationship between a second language (L2) learner's oral fluency and oral proficiency has important implications for L2 acquisition theories, pedagogical practices, high-stakes testing, and so on (de Jong, Groenhout, Schoonen, & Hulstijn, 2013; Ginther, Dimova, & Yang, 2010; Segalowitz, 2010; Skehan, 2009). For example, determining the relationship between these two may provide a better understanding of how the L2 system develops (Segalowitz, 2010; Skehan, 2009), which in turn may generate better teaching methods for improving oral L2 fluency and, as a result, help learners develop greater L2 proficiency (de Jong & Perfetti, 2011). Moreover, knowing how fluency development may vary according to proficiency from one L2 to another can allow educators to know what to expect and how to appropriately weigh various aspects of fluency as they promote the proficiency development of their learners. Finally, those interested in high-stakes testing might examine whether automatically measuring L2 fluency could replace more laborious and expensive manual rating systems for estimating L2 proficiency (Ginther et al., 2010; Ushigusa, 2009). The purpose of the current study is to better understand the relationship between oral fluency and L2 proficiency, especially as it relates to different L1-L2 combinations and at different ACTFL proficiency levels and sublevels.

Literature Review

The term *fluency*, especially in an L2 context, has several definitions, most of which refer to fluidity or ease of speech (Guillot, 1999; Kormos, 2006; Riggenbach, 2000; Schmidt, 1992; Segalowitz, 2010). In its broad sense, as defined by Lennon (1990), L2 fluency is often synonymous with overall L2 proficiency. In this sense, when laypeople use the word *fluency*, they often mean *proficiency* (Guillot, 1999). More narrowly, fluency can refer to language that is produced fluidly and smoothly as one combines words and sentences in speech. In this sense, the term is used in most L2 research to complement accuracy (i.e., error-free language) and complexity (i.e., language spoken with situation-appropriate forms) as one of the three main aspects of proficiency (Towell, 2012). However, these three factors may influence each other differently depending on the speaker's L2 proficiency (Skehan, 2001, 2009; Taguchi, 2008). Segalowitz (2010) further divided this narrow sense of fluency into three subcategories: cognitive fluency, a speaker's ability to plan and execute L2 speech smoothly and easily; perceived fluency, a native speaker's subjective impression of a nonnative speaker's ease of producing speech; and utterance fluency, often called temporal fluency, the measurable features for the ease and smoothness of L2 speech, such as the speech rate, number of hesitations, and number and length of pauses. This study focused on utterance fluency and sought to understand its relationship with overall L2 proficiency.

Tavakoli and Skehan (2005) have suggested that utterance fluency has three aspects: speed fluency, the rate at which speech is delivered; breakdown fluency, disruptions in the ongoing flow of speech; and repair fluency, how often speakers make repairs, corrections, or false starts. Each of these aspects of L2 utterance fluency is determined by measuring specific acoustic features of speech (Derwing, Munro, Thomson, & Rossiter, 2009; Fillmore, 1979; Goldman-Eisler, 1961; Lennon, 2000). While many measures of L2 utterance fluency exist,

some of the most typical are as follows: for speed fluency, syllables per second (Hilton, 2009), pruned syllables per second (i.e., the number of syllables minus hesitations or other disfluencies) (Rossiter, 2009), the number of runs or turns (Ginther et al., 2010), and mean length of run in syllables (Ginther et al., 2010) are most frequently considered. For breakdown fluency, most researchers examine the number of pauses (Trenchs-Parera, 2009) and length of pauses (Rossiter, 2009). For repair fluency, the most common measurements are the number of hesitations, false starts, and filled pauses (Borges de Almeida, 2009). Also see Kormos (2006) for other measures.

Recently, researchers have attempted to both verify and clarify the connection between utterance fluency, cognitive fluency, and overall L2 proficiency (de Jong, Groenhout, et al., 2013; de Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2013). Other such discussions can be found in Segalowitz (2010) and Larsen-Freeman (2009). Investigating such connections is both plausible and potentially beneficial to both researchers and those involved in instruction and assessment, because features of L2 utterance fluency may indicate the level of L2 automaticity, stability, and processing efficiency that a speaker has attained (Segalowitz, 2007, 2010). As Segalowitz (2007) has explained, L2 utterance fluency may be related to L2 cognitive fluency because it demonstrates access fluidity (the ability to access words and syntactic forms), attention control (the ability to focus attention during online processing), and the efficiency and control of other cognitive processes. In fact, it is possible that fluent speech requires, and likely reflects, a learner's L2 proficiency in accessing lexical forms and syntactic structures as well as a learner's phonological working memory, all of which improve with greater overall L2 proficiency (e.g., Pienemann, 1998). Thus, a strong connection may exist between accuracy and fluency, assuming that speakers cannot produce fluent speech if accuracy is not automatic (Lennon, 1990; Vercellotti, 2012). This lends credence to the hypothesis that fluid speech could indicate a speaker's proficiency. Two recent studies have found connections between cognitive fluency and overall proficiency: Segalowitz and Freed (2004), and de Jong, Steinel, et al. (2013).

The link between utterance fluency and overall proficiency, however, is still unclear and is most likely affected by several factors (Brand & Gotz, 2011; Derwing, Rossiter, Munro, & Thomson, 2004). Brand and Gotz (2011) compared the number of errors to the levels of utterance fluency in the speech of native German speakers as recorded in the Louvain International Database of Spoken English Interlanguage. They found that L2 accuracy (measured in number of errors) and L2 fluency (defined as speech rate and number of filled and unfilled pauses) were not highly correlated, suggesting that the connection between fluency and proficiency in L2 speech also may not be straightforward. Similarly, Garcia-Amaya (2009) demonstrated that the relationship between different measures of L2 utterance fluency varied greatly by L2 proficiency, although he focused mainly on examining learning context (i.e., study abroad, immersion, and at-home experiences).

However, finding a relationship between utterance fluency and overall proficiency is important for several reasons. One reason to examine this relationship is its implications for automatic proficiency scoring (Ginther et al., 2010; Higgins, Xi, Zechner, & Williamson, 2011; Ushigusa, 2009; Yoon et al., 2009). Given the expense and difficulty of rating learners' L2 proficiency, estimating overall proficiency by measuring features of utterance fluency would save both time and money. Researchers have had some success in correlating utterance fluency and overall proficiency in L2 learning (Hilton, 2009), although most automated scoring systems still measure L2 accuracy and complexity (grammar and vocabulary), as well as utterance fluency, especially when rating spontaneous speech (e.g., Higgins et al., 2011).

Another reason to examine this relationship is to better understand how the L2 system develops and, perhaps, to improve

its developmental efficiency and effectiveness (Segalowitz, 2010). This is especially important for practitioners who wish to help their students improve in overall proficiency. Understanding this relationship could lead to improvements in the teaching of L2s, with particular implications for direct instruction in classroom settings. For example, Gatbonton and Segalowitz (2005) argued that repetitive tasks in the context of a generally communication-oriented classroom can be used to help with proceduralization and automatization of formulaic expressions, which in turn improve both fluency and accuracy. Others (de Jong & Perfetti, 2011; Snellings, van Gelderen, & De Glopper, 2002) have also suggested methods of improving lexical and grammatical retrieval speed that help improve fluency. These methods may also improve overall L2 proficiency. Thus, a better understanding of the relationship between utterance fluency, cognitive fluency, and overall L2 proficiency can help practitioners select among and even develop more innovative teaching methods, strategies, and activities so as to specifically improve fluency in the classroom setting.

Thus, while some studies have revealed a correlation between cognitive fluency and overall L2 proficiency (de Jong, Steinel, et al., 2013; Segalowitz & Freed, 2004) and, to some extent, between utterance fluency and overall L2 proficiency (Ginther et al., 2010), others still question how strong this correlation may be. In short, more research is needed to understand the complex relationship between L2 utterance fluency and overall L2 proficiency. The main purpose of this study is to examine this relationship—that is, to determine whether learners' L2 utterance fluency measures could predict their overall L2 proficiency. In particular, the current study seeks to address three specific aspects of this relationship.

The first of these areas involves the way in which L2 utterance fluency develops in relationship to changes in general L2 proficiency. Most studies have examined the association between these two features using just one or two proficiency levels or have examined learners whose proficiency was not determined a priori (Kormos & Denes, 2004). Evaluating only one or two L2 proficiency levels does not clarify whether L2 utterance fluency can distinguish among very fine-grained differences in proficiency—that is, if the relationship is strong because it correlates across several proficiency levels. On the contrary, Garcia-Amaya (2009) scrutinized five proficiency levels for native English speakers learning Spanish. Results suggested that fluency differed depending on the level of proficiency and the setting in which the Spanish was learned (at home, during study abroad, or in an immersion setting). While these findings are suggestive, Garcia-Amaya (2009) assessed only 25 participants across five proficiency levels and four learning contexts. Beigi (2009), with some degree of success, did find that measures of L2 utterance fluency predicted L2 proficiency in his study, which examined 726 OPI interviews where the participants scored at either the Intermediate, Advanced, or Superior level. However, both of these studies surveyed only one L2 at a time: Spanish in Garcia-Amaya (2009) and English in Beigi (2009). No known study has examined whether variations in L2 utterance fluency might be found across a range of proficiency for several L2s. Hence, to better understand the links between L2 utterance fluency and proficiency, it is critical to examine a group of L2 speakers across a broad range of proficiency and representing many different L2s, i.e., the approach used in the present study.

The second area of research explores how L2 utterance features develop in relationship to changes in overall proficiency over time within the same speaker. Several studies have examined changes in L2 fluency over time and have found that learners can improve in overall oral fluency in general and utterance fluency in particular (e.g., DeKeyser, 2010; Derwing et al., 2009; Freed, Segalowitz, & Dewey, 2004). In multiple studies, Derwing and her colleagues (Derwing et al., 2004, 2009; Derwing, Munro, & Thomson, 2007) examined native

Slavic speakers and native Mandarin speakers who were learning English to see how their L2 fluency developed over a period of several years. Their research indicated that the native Slavic speakers did improve in oral fluency ratings over a two-year period, but the Mandarin speakers did not (Derwing et al., 2007). Such findings highlight the importance of evaluating the development of fluency across multiple L2s over time. Moreover, researchers who have studied longitudinal gains in L2 utterance fluency typically have not assessed concomitant, overall proficiency gains in the same population (e.g., Freed et al., 2004).

The third promising area of research considers how L2 fluency develops and how it is linked to proficiency in various L2s. Outcomes in this area of research are still unclear. That is, since even native speakers of different L2s use pauses and hesitations differently and speak at different rates (Derwing et al., 2004; Scanlon, 1987), the relationship between L2 proficiency and L2 utterance fluency may differ across various L2s. Certainly, a number of studies have examined the connection between fluency and oral proficiency in different L2s, such as Spanish (Garcia-Amaya, 2009), French (Freed, 1995), and Dutch (van Gelderen, 1994). However, few studies, if any, have analyzed more than one L2 in the same study. Although no existing studies examine speakers of the same L1 studying different L2s, as in the current study, Derwing et al. (2009) investigated fluency features (notably, pausing) in the English of both native Slavic speakers and native Mandarin speakers. They found that a speaker's L1 influenced how and when that speaker paused in L2 English. Similar results were found by Ginther et al. (2010). In addition, de Jong, Groenhout, et al. (2013) examined Turkish and English learners of Dutch and found, unlike the previous study, that a learner's L1 did not influence L2 fluency features. However, other than those two studies, to the current researchers' best knowledge, no study has examined more than one L1 or L2. Such cross-linguistic analyses within the same study are necessary for accurate comparisons, since methodologies, data collection methods, and the participants' ages and L2 proficiency can differ greatly across studies. Indeed, as Segalowitz (2010) argued, one of the challenges in making generalizations about L2 fluency is that researchers have not made sufficient comparisons across several L2s using the same L2 fluency measures. No known studies, then, have attempted to assess the relationship between L2 utterance fluency and overall L2 proficiency across several L2s.

Thus, the current study examined L2 utterance features in the speech of learners who differed both in L2 proficiency and in the L2s themselves. The present research also investigated which L2 utterance features changed as improvements in overall L2 proficiency were made for the same speakers over time. In particular, this study addressed the following research questions:

1. To what extent do features of L2 utterance fluency differ across proficiency regardless of the L2?
2. To what extent do specific features of L2 utterance fluency differ by L2 for the same proficiency levels and sublevels?
3. What relationship, if any, exists between increases in overall L2 proficiency from pretest to posttest and improvements in L2 utterance fluency?
4. What changes in L2 utterance fluency from pretest to posttest differ by L2 for the same L2 proficiency?

To answer these research questions, data were obtained from a short section of 126 official ACTFL Oral Proficiency Interviews (OPIs), gathered from native English speakers who were learning the following L2s: French, German, Japanese, Russian, and Arabic.

Methods

Participants

A total of 126 speech samples used in this study were obtained from 86 different native English speakers who were learning either French ($n = 36$), German ($n = 28$), Japanese

($n = 16$), Russian ($n = 10$), or Arabic ($n = 36$). These L2s were chosen because scholars have generally agreed that they vary in difficulty and thus in the amount of time required to reach a particular level of proficiency, as judged in studies from the American Council on Education's College Credit Recommendation Service (ACTFL, 2010) and from the U.S. Foreign Service Institute (see Omaggio Hadley [2001] for a description). All speech samples were from college-age students (18–26 years old) who were acquiring their respective L2s in a traditional classroom ($n = 24$), during a study abroad program ($n = 36$), or in a classroom using domestic immersion via foreign-language housing ($n = 26$). However, the setting in which each student learned his or her respective L2 was not relevant to the current research questions: This study focused on how L2 proficiency and L2 utterance fluency relate to each other across various L2s, apart from the question of the context in which the language is learned.

Procedures

The OPI is a standardized set of procedures for assessing spoken language according to criteria and scale established by the ACTFL.[1] In this study, all 30-minute interviews were conducted over the phone. Two independent, ACTFL-certified raters then scored the interviews. According to standard ACTFL practices, a third reviewer scored the recorded interview if the first two reviewers disagreed on the rating.

The OPI was selected to measure each learner's proficiency because OPIs are designed to facilitate comparisons across various L2s; that is, a learner who scores Intermediate Mid (IM) in Russian, for example, is judged to have the same degree of oral L2 proficiency as a learner who scores IM in French, although the amount of learning time for each individual may differ. More detail about the OPI format can be found at http://www.languagetesting.com, and descriptors for each level and sublevel can be found at http://actflproficiencyguidelines2012.org.

Of the 126 recorded files, 80 of the sound files represented pre- and posttests that were completed by 40 participants at the beginning and end of their learning experiences in either domestic classrooms, immersion classrooms, or study abroad programs. The remaining 46 sound files were single tests; the 36 Arabic study abroad students were not tested prior to their program, and 10 students failed to take the posttest. The time difference between pre- and posttests was at least six months but no more than eight months. These pre- and posttests provided the data from which the relationship between changes in L2 proficiency and fluency over time could be explored. Scores on the OPIs ranged from Novice Mid (NM) to Superior (S). Table 1 includes a breakdown of different participants for each L2 by proficiency.[2]

Sampling

For each OPI, a segment of each speaker's interview was selected from a point approximately three minutes into the interview. Selecting the speech sample at, or immediately after, the three-minute point ensured that participants had completed opening questions, such as introductions, by which point OPI testers have typically established a floor, i.e., a level at which the participant feels comfortable speaking and that the conversation at that level is fully in progress.

Although a number of previous studies on L2 utterance fluency did not report the length of the speech segments that were included in their research (Trenchs-Parera, 2009), and although other studies have gathered data from speech segments of 30–60 seconds in length (Derwing et al., 2004; Rossiter, 2009; Yoon et al., 2009), for the present study, 20-second samples were deemed sufficient to obtain data on all of the measures of utterance fluency under consideration. The length of each speech segment was extended beyond the 20-second limit only if the speaker was in

TABLE 1

Number of L2 Speech Samples Analyzed for All Nine ACTFL OPI Proficiency Levels

Language	NM	NH	IL	IM	IH	AL	AM	AH	S	Total Speech Samples per Language
French				5	6	16	8	1		36
German			4	8	6	3	7			28
Russian							1	7	2	10
Japanese	2	3	2	3	2	4				16
Arabic			1	9	16	7	3			36

Total Speech Samples per OPI Level								
NM	NH	IL	IM	IH	AL	AM	AH	S	Total Speech Samples
2	3	7	25	30	30	19	8	2	126

Note: NM = Novice Mid, NH = Novice High, IL = Intermediate Low, IM = Intermediate Mid, IH = Intermediate High, AL = Advanced Low, AM = Advanced Mid, AH = Advanced High, S = Superior.

the middle of a syllable at the 20-second mark; in this case, the end of that single syllable was also included.

Measures

The following sources were consulted to determine how to measure each of these features: Freed et al. (2004), Rossiter (2009), and Trenchs-Parera (2009). Drawing on these sources, for each 20-second speech segment, raters recorded measurements of speed, breakdown, and repair fluency. Praat software created by Boersma and Weenink (2011) was used to measure the length of pauses, hesitations, and runs to the nearest millisecond. Specific measures in each domain are described below.

Speed Fluency

For speed fluency, the number of syllables per second, number of runs, and average length of run were coded and calculated. Runs were defined as the length of time during which the speaker produced speech without hesitations or inappropriate pauses. Inappropriate pauses are those that do not occur at a grammatical juncture where native speakers would typically pause (Lennon, 1990; Towell, Hawkins, & Bazergui, 1996). The raters also examined the number of runs and mean length of run that a speaker produced during the 20-second segment. The mean length of run equaled the average length of these runs in syllables. Finally, when the scorers had calculated the syllables per second, they counted only pruned syllables, which are defined as the number of syllables minus hesitations and false starts (Rossiter, 2009). This was done to ensure that hesitations, false starts, and so on did not factor into the speech rate (see de Jong, Groenhout, et al., 2013 for a more complete discussion). The terms *syllables per second* or *speech rate* are used throughout the current study.

Breakdown Fluency

For breakdown fluency, the number of pauses and length of pauses were recorded. Pauses were defined as any absence of speech (or vocalization) longer than 300 milliseconds. Some researchers have defined pauses as any absence of speech (or vocalization) over 200 milliseconds (Ginther

et al., 2010; Riggenbach, 2000; Towell et al., 1996), asserting that any pause shorter than this may include onsets of stop consonants and any pause longer than this may miss important shorter pauses. Others have set the minimal length at 400 milliseconds (de Jong, Groenhout, et al., 2013; Derwing et al., 2004), arguing that a higher cutoff limit better disregards micropauses that may occur normally in between words and phrases. The 300-millisecond length was chosen because it is in between the other two proposed measurements, thus avoiding some limitations of either measurement. In addition, some pausing is necessary in speech and does not necessarily represent disfluency; however, most analyses of pausing include all pauses, so the current study also did this (de Jong, Groenhout, et al., 2013; Rossiter, 2009; Trenchs-Parera, 2009).

Repair Fluency

For repair fluency, data were gathered on the number of hesitations and number of false starts (i.e., where a word or sentence is stopped mid-utterance). Hesitations were classified as either self-repetition, self-correction, or filled pauses (i.e., pauses that are filled with words such as "uh" or "oh").

Results

Research Question 1

This study's first research question investigated the extent to which features of L2 utterance fluency differed across proficiency regardless of the L2. The answer to this question was obtained by examining the seven measures of L2 utterance fluency at each of the ACTFL OPI scores for which there was a sample. This analysis included the OPI posttest speech samples from all 86 participants. Table 2 summarizes the L2 utterance fluency for all speakers in 8 of the 10 ACTFL OPI scores.

A series of one-way ANOVAs was conducted using the scores of all L2 fluency measures as the dependent variables and all L2 proficiency scores (from Novice High [NH] to S) as the independent variables. Because no participant received a posttest score of NM, this analysis examined only eight proficiency scores (minus NM since only two participants received this score and at pretest only). A Bonferroni adjustment was made to the p value in order to allow for multiple comparisons; therefore, significance was set at $p = 0.007$ (or 0.05/7 because seven analyses were run on the data). Significant differences were obtained for five of the seven L2 fluency measures across all L2 proficiency scores. Those five significant L2 fluency measures were speech rate, mean length of run, number of runs, number of pauses, and length of pauses. Posthoc Tukey tests for each of these five L2 fluency measurements were then used to determine the manner in which the speakers' scores differed across eight L2 proficiency scores: NH, Intermediate Low (IL), Intermediate Mid (IM), Intermediate High (IH), Advanced Low (AL), Advanced Mid (AM), Advanced High (AH), and S (all F's$[1,85] > 3.607.82$, all p's < 0.0002, all η_p^2's > 0.26).

Speed Fluency

The one Novice group (NH) and three Intermediate groups (IL, IM, and IH) had the lowest speed fluency; i.e., participants with these scores evidenced the slowest speech rate and shortest length of run. The three Advanced groups (AL, AM, and AH) had both a faster speech rate and longer length of run than the four lower-proficiency groups. Moreover, participants who were rated Superior had the greatest speed fluency, i.e., the fastest speech rate and longest length of run, and constituted a group that was distinctly separate from participants who received each of the other seven OPI scores.

Breakdown Fluency

Data on the number of pauses (another aspect of breakdown fluency) revealed that IL, IM, and IH speakers ($F[1,86] = 4.31$, $p = 0.0001$, $\eta_p^2 = 0.30$) had more pauses than the higher-proficiency groups. In addition, speakers rated AL and AM had more

TABLE 2

Utterance Fluency Features of L2 Speakers by ACTFL Proficiency Level

	Speed Fluency			Breakdown Fluency		Repair Fluency	
Proficiency Level	Syllables per Second	Number of Runs	Length of Run	Number of Pauses	Length of Pauses	Number of False Starts	Number of Hesitations
NH ($n = 2$)	1.58 (0.67)	9.00 (2.82)	4.46 (1.26)	5.00 (1.41)	0.95 (0.44)	2.00 (0.000)	5.50 (0.71)
IL ($n = 2$)	1.75 (0.07)	5.50 (6.36)	4.27 (1.23)	8.00 (1.41)	0.94 (0.13)	1.00 (0.000)	4.50 (0.71)
IM ($n = 17$)	2.10 (1.2)	7.17 (1.97)	4.39 (1.46)	8.70 (1.92)	0.75 (0.18)	2.23 (1.92)	7.17 (2.74)
IH ($n = 23$)	1.80 (0.74)	7.13 (2.54)	5.13 (1.71)	8.26 (1.45)	0.79 (0.29)	2.34 (1.92)	6.60 (2.87)
AL ($n = 21$)	2.78 (0.79)	8.42 (1.80)	6.92 (2.74)	6.85 (2.78)	0.65 (0.19)	1.52 (1.28)	5.47 (2.29)
AM ($n = 9$)	2.88 (0.64)	8.55 (1.74)	7.07 (2.37)	6.44 (2.18)	0.51 (0.07)	2.33 (1.58)	5.66 (1.93)
AH ($n = 4$)	3.03 (0.29)	7.25 (0.95)	8.51 (1.51)	4.44 (1.15)	0.48 (0.05)	1.25 (1.89)	5.50 (1.91)
S ($n = 2$)	3.50 (0.45)	6.50 (2.12)	11.70 (5.23)	4.16 (0.233)	0.55 (0.07)	0.50 (0.71)	4.50 (0.71)

Note: For each fluency measure, values are means with standard deviations in parentheses. NH = Novice High, IL = Intermediate Low, IM = Intermediate Mid, IH = Intermediate High, AL = Advanced Low, AM = Advanced Mid, AH = Advanced High, S = Superior.

pauses than speakers rated AH and S, who exhibited the least number of pauses of any of the groups. Interestingly, participants who were rated NH, the lowest L2 proficiency group, used a number of pauses that was similar to that of both the AH and S groups rather than more similar to the use of pauses by participants who were rated IL, IM, and IH.

NH and IL groups ($F[1,85] = 3.22$, $p = 0.005$, $\eta_p^2 = 0.24$) had longer pauses than all of the higher proficiency level groups, while, in turn, the IM, IH, and AL groups had longer pauses than the AM, AH, and S groups. No difference in length of pauses was found among these three highest scorers (AM, AH, and S).

Repair Fluency

Differences in the two repair fluency measures—the number of hesitations and number of false starts—did not reach statistical significance for any L2 proficiency (all F's < 1.09, all p's > 0.05, $\eta_p^2 < 0.08$).

In sum, measures of L2 fluency involving speed and breakdown fluency differed across L2 proficiencies, while measures of L2 fluency involving repair fluency did not. These results led to three broad conclusions. First, L2 utterance fluency measures varied across proficiency in a predictable pattern, with speakers with higher L2 proficiency demonstrating greater L2 utterance fluency than those at the lower levels. Second, it appears that the greatest differences in L2 utterance fluency occurred across the highest proficiency scores. For example, for three of the five L2 fluency measures in which a significant difference was manifested across L2 proficiency, the S group differed significantly from the groups of participants at other levels of proficiency and were set apart from those groups by their longer lengths of run, faster speech rate, and fewer pauses. Finally, although differences in L2 utterance fluency appeared across L2 proficiency, many of these differences were not statistically significant across the eight ACTFL scores under consideration (e.g., from IM to IH), with the greatest differences occurring across whole levels (e.g., from Advanced to Superior) rather than sublevels.

The above analysis determined which measures of L2 utterance fluency were related to overall L2 proficiency, but it did not reveal the overall importance of each feature when predicting L2 proficiency. To examine this matter, the current researchers ran both a correlational and a linear stepwise multiple regression analysis with L2 proficiency as the dependent variables and L2 utterance fluency measures as predictor variables (number of hesitations, number of false starts, number of pauses, length of pauses, number of runs, mean length of run, and syllables per second). Proficiency was converted to a nine-point scale with Novice Low as a one on the scale, S as a 10 on the scale, and the other proficiency scores in between.

The results of this analysis indicated that four of the seven measures of L2 utterance fluency correlated significantly with L2 proficiency as measured on the OPI: the speech rate (0.559), length of run (0.474), number of pauses (−0.386), and length of pause (−0.436). Correlations between oral proficiency and the other three variables—the number of runs (0.111), false starts (−0.128), and hesitations (−0.181)—were not statistically significant. The results of the regression analyses revealed that the speech rate (syllables per second) played the greatest role in predicting L2 proficiency, accounting for 31% of the variation. The mean length of run also played a minor role in predicting L2 proficiency. To summarize, two features of speed fluency had the greatest correlation with L2 proficiency, while the two measures of repair fluency did not predict proficiency. Table 3 provides the results of the regression analysis.

Research Question 2

The results of the initial analyses established that some L2 utterance fluency features were predictive of proficiency, regardless of L2 examined. However, it was still possible that the L2s under study could play an important role in determining the way in which L2 utterance

TABLE 3

Multiple Regression Analysis With L2 Proficiency Level as Dependent Variable and L2 Utterance Features as Predictor Variables

Predictor	Δ (Adjusted R^2)	β	F statistic	*p* value
Syllables per second	0.310	0.5	$F = 12.72$	$p = 0.0001$
Mean length of run in syllables	0.026	0.3	$F = 12.21$	$p = 0.0001$
Total	0.336			

features relate to L2 proficiency, particularly since the difficulty level of an L2, according to the U.S. Department of Defense, reflects the fact that some languages require more learning time for English speakers to progress from one proficiency level to the next (see Omaggio Hadley, 2001), as determined by proficiency testing at the Defense Language Institute. For example, for native speakers of English, French requires the least amount of allocated and engaged learning time, German requires somewhat more, and Arabic requires the most time of the three. Thus, it was hypothesized that native English–speaking learners who study L2s that are more similar to English would probably spend less time processing factors such as sentence structure or the sound system. Specifically, it was predicted that the speakers of L2 French and L2 German would attain a higher L2 fluency level (a faster speech rate, longer lengths of run, and fewer pauses) than the speakers of L2 Arabic at the same L2 proficiency.

To determine the impact of relative language difficulty on fluency, researchers evaluated OPI posttest speech samples of speakers scoring IM, IH, and AL for three of the L2s (German [$n = 14$], French [$n = 17$], and Arabic [$n = 32$]). Only these scores and languages were used because these three languages differ in terms of the amount of time that native English speakers require to reach a certain L2 proficiency and because these groups had sufficient numbers of participants. To avoid including more than one speech sample per person, the researchers examined only posttest OPI speech samples. A series of one-way ANOVAs were run with the L2 utterance fluency scores as the dependent variables and the L2s (French, German, or Arabic) as the independent variables. In each of the analyses, OPI ratings were used as a covariate because the number of participants varied for each of the three proficiency scores across each of the three L2 groups. This covariate was not significant in any of the analyses, suggesting that L2 proficiency was not a significant factor in the differences between L2 groups (all *F*'s $[1,52] < 0.500$, all *p*'s > 0.05, all η_p^2's < 0.007). As in the analyses performed for the first research question, a Bonferroni adjustment was made on the data, setting the significance at $p = 0.007$.

The results of these analyses revealed that three of the seven L2 utterance fluency measures differed significantly across the three L2s: the length of run ($F[1,52] = 12.54$, $p = 0.0001$, $\eta_p^2 = 0.30$), number of runs ($F[1,52] = 8.65$, $p = 0.001$, $\eta_p^2 = 0.22$), and number of hesitations ($F[1,52] = 5.39$, $p = 0.0007$, $\eta_p^2 = 0.15$). None of the other fluency measures differed significantly across the three L2 groups (all *F*'s < 4.54, all *p*'s > 0.007, all η_p^2's < 0.13). Second, according to posthoc Tukey tests, the number of hesitations differed notably between the German L2 learners and the other two L2 groups, with the German L2 group having fewer hesitations than the French L2 and Arabic L2 groups. Data for the L2 learners of French and German were similar with respect to the number of runs and length of run, which were longer than those of the L2 learners of Arabic. As data showed in response to the first research question, speed

fluency features were the most decisive measures in determining differences in fluency—this time across L2s. Table 4 presents the results for this research question.

Research Question 3

As stated above, 40 of the 86 participants completed both an OPI pretest and posttest, so researchers could examine whether these participants made significant gains in L2 utterance fluency and in overall L2 proficiency over time, i.e., from the OPI pretest to the posttest (see Table 5 for details on the number of participants for each language). These participants' speech samples were used to determine the extent to which learners whose overall L2 proficiency increased from pretest to posttest also improved in L2 utterance fluency.

Of the 40 participants, 22 were non-gainers, meaning either they did not gain at least one ACTFL sublevel from pretest to posttest, or they received a lower score on the posttest than they received on the pretest. Interestingly, of these 22 non-gainers, 5 learned only in classrooms, and 17 learned in immersive experiences coupled with classroom experiences. On the other hand, the other 18 of the 40 participants were gainers, meaning that they moved up at least one ACTFL sublevel from pretest to posttest (e.g., from IM to IH). Nine of these gainers learned their L2 in an immersive, foreign housing experience coupled with classroom experience; the other nine learned an L2 only in a classroom setting. Pre- and posttest data were available for classroom and immersion learners of French, German, Russian, and Japanese. Pretest data were not available for any of the students who had participated in study abroad programs, including learners of Arabic. However, the context in which participants learned the L2 was not the focus of this study. Instead, this study sought to determine if L2 speakers who made significant gains in overall L2 proficiency also made concomitant gains in L2 utterance fluency.

TABLE 4

L2 Utterance Fluency Measures for Three L2s: French, German, and Arabic

Language	Speed Fluency			Breakdown Fluency		Repair Fluency	
	Syllables per Second	Number of Runs	Length of Run	Number of Pauses	Length of Pauses	Number of False Starts	Number of Hesitations
French ($n = 17$)	2.09 (0.638)	5.84 (1.75)	1.45 (0.498)	7.62 (2.11)	0.71 (0.242)	1.63 (1.31)	6.42 (2.11)
German ($n = 14$)	2.28 (0.842)	6.25 (2.00)	1.75 (0.661)	7.75 (2.23)	0.68 (0.174)	1.19 (0.949)	4.19 (2.53)
Arabic ($n = 32$)	1.75 (0.692)	4.49 (1.19)	1.11 (0.291)	8.71 (1.87)	0.75 (0.233)	2.72 (1.94)	7.22 (2.21)

Note: For each fluency measure, values are means with standard deviations in parentheses.

TABLE 5

Changes in All Measures of L2 Utterance Fluency for Gainers and Nongainers in Overall ACTFL OPI Proficiency from Pretest to Posttest

	Speed Fluency			Breakdown Fluency		Repair Fluency	
	Syllables per Second	Number of Runs	Length of Run	Number of Pauses	Length of Pauses	Number of Hesitations	Number of False Starts
Gainers ($n = 18$)	0.322 (1.82)	−0.44 (2.48)	0.17 (0.358)	−1.08 (2.32)	0.0001 (0.221)	0.61 (2.92)	0.33 (1.45)
Nongainers ($n = 22$)	−0.318 (14.01)	−0.22 (2.55)	0.03 (0.697)	0.05 (2.98)	0.00084 (0.257)	0.31 (14.01)	−0.04 (1.66)

A series of repeated-measures ANOVAS was run with L2 utterance fluency measures as the dependent variables on the pretests and posttests (with the p value Bonferroni adjustment at 0.003, since 14 separate analyses were performed on the data). The findings revealed that the gainers improved from pretest to posttest for two of the seven features of L2 utterance fluency: gainers produced more syllables per second and longer run lengths from pretest to posttest (all F's > 8.42, $p = 0.0001$, $\eta_p^2 = 0.33$) than nongainers. By contrast, the nongainers did not improve significantly on any of the L2 features of utterance fluency and, in some instances, actually performed slightly worse on the posttest than on the pretest (all F's < 0.455, all p's > 0.05, all η_p^{2}'s < 0.02). Table 5 provides a representation of these findings. Again, the biggest differences across proficiency levels were found for speed fluency features.

Research Question 4

Having established that speakers who improved in overall L2 proficiency also improved on some measures of L2 utterance fluency, the way in which a particular L2 proficiency level and a particular L2 were associated was examined, using measures of gain from pretest to posttest in overall L2 proficiency and L2 utterance fluency measures. In this manner, the changes that the gainers in each L2 group demonstrated from pretest to posttest in each of the seven measures of L2 utterance fluency could be analyzed. Similar analyses were then performed for the nongainers in each L2 group. Table 6 provides the results of these analyses.

When divided by L2, in some cases the number of speakers in each group was too low to perform reliable statistical analyses, in particular the Russian and Japanese L2 speakers. However, we were still able to make some comparisons across the Russian gainers and nongainers since both groups scored at the Advanced level on their pretests. Simple visual comparisons of L2 gains across the Russian gainers and nongainers in

TABLE 6

Changes in All Seven Measures of L2 Fluency, by L2, for Gainers and Nongainers From Pretest to Posttest

Language (Number of Learners/ Proficiency Level)	Speed Fluency			Breakdown Fluency		Repair Fluency	
	Syllables per Second	Number of Runs	Length of Run	Number of Pauses	Length of Pauses	Number of Hesitations	Number of False Starts
Russian gainers (3 A)	0.317 (0.931)	0 (1.41)	0.275 (0.555)	−2.52 (1.06)	−0.003 (0.149)	1 (1.63)	−0.333 (1.24)
Russian nongainers (2 A)	−0.05 (0)	1.00 (1.00)	−0.46 (0.125)	3.62 (1.99)	0.029 (0.0935)	0.5 (0.50)	1.5 (1.5)
French gainers (3 I, 3 A)	0.425 (0.855)	−1.16 (2.91)	0.347 (0.243)	−2.33 (1.79)	0.053 (0.257)	1.16 (3.07)	−0.166 (1.34)
French nongainers (3 I, 6 A)	−0.067 (0.714)	0.777 (2.77)	−0.146 (0.589)	−0.889 (2.80)	0.004 (0.289)	1 (1.76)	−0.55 (1.77)
German gainers (5 I, 1 A)	0.225 (0.665)	1 (1.73)	−0.08 (0.254)	1.16 (2.11)	−0.067 (0.227)	−0.166 (3.67)	0.666 (1.37)
German nongainers (3 I, 4 A)	0.193 (0.532)	−2.14 (1.95)	0.59 (0.691)	−0.714 (1.48)	0.044 (0.206)	−0.57 (2.61)	−0.42 (1.39)
Japanese gainers (2 N, 1 I)	0.35 (0.353)	−2.33 (1.69)	0.266 (0.258)	−1.66 (0.471)	0.03 (0.148)	0.667 (0.942)	1.33 (1.25)
Japanese nongainers (2 I, 2 A)	−0.25 (0.957)	0.25 (1.29)	−0.29 (0.454)	1.75 (3.63)	−0.056 (0.301)	0.25 (2.04)	1.00 (0.707)

Notes: N = Novice Proficiency Level, I = Intermediate Proficiency Level, A = Advanced Proficiency Level. For each fluency measure, values are means with standard deviations in parentheses.

the syllables per second, number of pauses, and length of run suggested that the gainers did indeed improve more on these fluency measures than the nongainers did. Although a similar pattern was present across the Japanese L2 groups, the L2 proficiency levels were highly dissimilar for gainers vs. nongainers. The Japanese L2 gainers were mostly Novice and Intermediate speakers, while the Japanese L2 nongainers were mostly Intermediate and Advanced speakers. Such a discrepancy in proficiency levels made comparisons across the two Japanese L2 groups difficult.

By contrast, both the gainers and nongainers in the French L2 and German L2 groups included Intermediate and Advanced speakers, which again facilitated comparisons. Interesting patterns appeared across these two L2s. First, little difference seemed to exist regarding gains in speech rate between the German L2 gainers (0.225) and nongainers (0.193). In contrast, the difference in L2 gains in speech rate for the French L2 gainers (0.425) vs. the French L2 nongainers (−0.067) seemed much greater. Such findings suggest that the relationship between changes in overall L2 proficiency and changes in L2 utterance fluency may differ according to the L2 learned. However, statistical analyses of both nongainers and gainers across the French L2 and German L2 groups did not reveal significant differences for any of the chosen L2 utterance features (all F's < 2.31, all p's > 0.05).

Discussion

The main purpose of this study was to determine how L2 utterance fluency relates to overall L2 proficiency across several different L2s and proficiency levels. This relationship was evaluated in two ways. First, analyses were conducted to examine the differences in measures of L2 utterance fluency for several L2s across several overall L2 proficiency levels. Changes in L2 utterance fluency for several L2s from pretest to posttest for overall OPI L2 proficiency were addressed. Both of these analyses investigated whether or not the specific L2 affected the relationship between overall L2 proficiency and L2 utterance fluency. What follows is a discussion of these findings.

The Relationship Between L2 Utterance Fluency and L2 Proficiency

This study's findings revealed that certain features of L2 utterance fluency do, in fact, differ according to L2 proficiency level. These results mirror the results of Garcia-Amaya (2009) and Beigi (2009), which suggested that L2 fluency is related, at least partly, to L2 proficiency. The findings also expand on these studies to indicate that utterance fluency can, and does, distinguish among multiple different proficiency levels rather than just across one or two different levels, as well as across several L2s.

In addition, the current results provide several additional insights into the relationship between L2 oral proficiency and L2 utterance fluency. First, these results demonstrate that particular features of L2 utterance fluency, including speed fluency (speech rate and mean length of run) and one measure of breakdown fluency (number of pauses) vary across L2 proficiency levels, especially in more advanced levels. These findings are consistent with earlier studies (de Jong, Groenhout, et al., 2013; Derwing et al., 2009) that established that these fluency measures are often indicators of L2 proficiency level (at least when only comparing two vastly different proficiency levels without considering increments in between). The current study is the first to demonstrate this variation across several L2s in the same study and across a range of proficiency levels and sublevels.

Although speech rate in the present study seemed to relate to overall L2 proficiency more than all other six features of L2 utterance fluency, Segalowitz (2010) has cautioned against approaches focusing mainly on speech rate as an L2 fluency measure because it is unclear whether speech rate correlates with efficient L2 processing. However, data from this study indicate that

speech rate seems to be the strongest fluency indicator of L2 proficiency. In addition, clear differences in L2 fluency emerged across broad L2 proficiency levels (i.e., Novice, Intermediate, Advanced, and Superior). However, clear distinctions in L2 proficiency sublevels arose only at the more advanced sublevels (AL, AM, AH, and S). In other words, L2 utterance fluency measures did not help distinguish among groups at lower L2 proficiency levels as accurately as they did at higher levels. This is the first study to demonstrate the link between L2 measures of utterance fluency and overall L2 proficiency at several proficiency levels for several L2s.

In terms of direct application of these findings, the results suggest that using L2 measures of utterance fluency alone for automatic scoring and estimation of L2 proficiency would not allow for sufficiently precise scoring, in particular at lower proficiency levels, but such L2 fluency measures may be useful at higher proficiency levels. Rather, it appears that rate of speech, length of run, and number of pauses would be the best combination of tools to use for such estimates.

It is interesting to consider why L2 utterance fluency features would differ more at the higher L2 proficiency levels than for participants at lower L2 proficiency levels. Most previous studies have argued that L2 measures of utterance fluency are also indications of L2 cognitive fluency; the current study's results suggest that L2 speakers at higher L2 proficiency levels will exhibit greater L2 cognitive fluency than speakers at lower L2 proficiency levels. Hence, based on the results of this study, one can assume that faster speech rates, longer runs, and fewer pauses suggest greater automaticity, greater attention control, and a more stable and efficient system (see above discussion of Segalowitz [2007,2010]).

If this is true, then the current results indicate that speakers at lower L2 proficiency levels may still be developing a stable and efficient L2 system and therefore do not truly differ on these measures of L2 utterance fluency, even though participants across different L2 proficiency levels may vary in the accuracy and complexity of their speech. On the other hand, it could be that changes in L2 cognitive fluency at these L2 proficiency levels are not noticeable in terms of L2 utterance fluency. For example, it could be that a learner's L2 system becomes more stable at an IM L2 proficiency level in comparison to an IL L2 proficiency level; but as an L2 learner gains more complex grammar or vocabulary, he or she must pause more or speak slower to access these new features. Thus, learners at less advanced L2 proficiency levels may increase in the number of pauses and speech rate only very slowly as the L2 system continues to develop. However, at the more advanced L2 proficiency levels, learners have typically acquired most of the L2 grammar and enough of the L2 vocabulary that they can finally advance in L2 cognitive fluency. This hypothesis needs to be verified by further research since the higher number of speakers at Advanced and Intermediate levels than at Novice levels in this study may have influenced these results.

In addition, certain L2 measures of utterance fluency (the number of hesitations and false starts)—did not vary across L2 proficiency levels. Interestingly, the L2 features of utterance fluency that did not vary were all repair fluency measures, which corroborates Tavakoli and Skehan's (2005) findings. This lack of correlation between repair fluency measures and L2 proficiency levels may indicate that these features differ from speaker to speaker—in other words, they are more indicative of a speaker-specific "trait" than indicative of an average speaker's current state of L2 development (Derwing et al., 2009). Clark and Wasow (1998), for example, found that some native speakers of English regularly demonstrated disfluencies—particularly false starts and repetitions—even when speaking English. In other words, these L2 fluency features may reflect differences in an L2 speaker's individual speech characteristics instead of differences in an L2 system's development.

For similar arguments, see de Jong, Groenhout, et al. (2013) and Derwing et al. (2009). More research is needed that explores the relationship between measures for L1 and L2 utterance fluency.

Another possible explanation for a lack of difference across L2 proficiency levels for some L2 utterance fluency measures is that accuracy and complexity may interact differently with L2 utterance fluency at various L2 proficiency levels. For example, large numbers of hesitations at lower L2 proficiency levels could be caused by a learner's attempts to retrieve simple lexical items, whereas large numbers of hesitations at higher L2 proficiency levels could be caused by a learner's attempt to produce more complicated syntactic structures in the L2 (Hilton, 2009). In both cases, the number of hesitations could be the same, but this similarity masks an increase in overall L2 proficiency. Research has suggested that the location of the pauses might be as important an indicator of L2 proficiency as the number and length of pauses (Hilton, 2009).

As far as pedagogical implications are concerned, the results of this study seem to encourage using features of L2 utterance fluency to locate more fine-grained differences in L2 proficiency—especially at higher proficiency levels. Beyond that, they suggest that instructors need not be concerned by aspects of fluency such as false starts (stopping mid-word or mid-utterance) or hesitations (self-correction, self-repetition, using natural fillers such as "um" and "uh," etc.), since these occur regularly in native speech and do not appear to relate closely to L2 proficiency.

L2 Utterance Fluency, L2 Proficiency Level, and the L2 Itself

In addition to examining the relationship between L2 utterance fluency and overall L2 proficiency, this study also sought to determine whether this relationship differed depending on the speaker's L2. To address this, interviews with L2 learners of German, French, and Arabic at Intermediate L2 proficiency levels were analyzed. The hypothesis was that native English speakers studying L2s that are less closely related to English (e.g., Arabic) would be less fluent at the same L2 proficiency level than would native English speakers studying L2s that are more closely related to English (e.g., French and German). The current findings appear to support, at least partly, this hypothesis. Native English speakers studying Arabic as an L2 (a language that is typically considered to be more difficult for native English speakers than French or German) demonstrated less fluency on certain measures than native English speakers studying French and German as L2s. In particular, the German L2 speakers demonstrated fewer hesitations than Arabic L2 speakers; moreover, French and German L2 learners produced more and longer runs than the Arabic L2 speakers did at the same L2 proficiency level. However, it was somewhat surprising that the German L2 speakers demonstrated greater fluency than the French L2 speakers, at least in terms of hesitations. Presumably, the more complex grammar of German would make it more challenging for L2 learners than French. In fact, researchers at the Foreign Service Institute have reported this very conclusion: Reaching an Intermediate L2 proficiency level in German takes longer than reaching the same L2 proficiency level in French for native English speakers (Liskin-Gasparro, 1982; Thompson, 2013).

The reasons for this discrepancy remain unclear. One possible explanation is that the length of runs and number of hesitations vary in the speech of native speakers of both German and French, and, therefore, the L2 learners imitated these variations. Relatively few studies besides Alcala (2009) and Scanlon (1987) have examined native French and/or native German fluency measures. Since no known studies have compared fluency across native speakers of different languages studying the same L2, understanding these differences is challenging. Indeed, although more researchers are comparing the L2 fluency abilities of L1 speakers and L2 speakers (e.g., Osborne,

2007; Yoon et al., 2009), little research exists from which to base solid comparisons across different L2s. Future research in this area would be helpful to determine why learners of different L2s might achieve different levels of L2 utterance fluency at the same L2 proficiency level.

The current study's findings also raise the question of whether definitions of fluency vary from language to language. For example, speech rate for native Japanese speakers may be quite different from that of native French speakers. Furthermore, such differences demonstrate that improvements in speech rate may be less important for one language than another. Derwing et al. (2009) found some indication that variations may occur across learner groups when contrasting the differences in English L2 fluency between native Slavic speakers and native Mandarin speakers. These scholars discovered that the Slavic speakers attained a greater level of L2 utterance fluency from pretest to posttest than the Mandarin speakers. Derwing et al. (2009) hypothesized that this could be the result of different L1s, but they also determined that the native Slavic speakers spoke L2 English more often and had stronger ties to the native English-speaking community than the native Mandarin speakers did. Moreover, even various dialects of the same language can differ in terms of speech rate (see Jacewicz, Fox, O'Neill, & Salmons, 2009). Future research should examine the measures of L2 utterance fluency for native speakers of various languages and dialects to establish a baseline against which to judge the L2 speech fluency of nonnative speakers. Research in natural language processing, which indicates differences across languages such as the location of naturally occurring pauses, the length of acceptable pauses, etc., can inform these efforts (Chou, Tseng, & Lee, 1996; Kaiki & Sagisaka, 1992; Li & Yang, 2009, Quimbo, Kawahara, & Doshita, 1998).

Given differences across languages in terms of connections between fluency and proficiency, it appears that language educators may benefit from a heightened consciousness of the norms of the L2 they are teaching, with particular attention to fluency features such as frequency and length of pauses and rate of speech production. If native speakers of the target language routinely incorporate more frequent and longer pauses, then perhaps learners can use this to their advantage, learning to pause at natural points in their speech as they prepare both conceptually and linguistically to produce their subsequent speech. On the other hand, if speech is expected to be more rapid and include fewer unfilled pauses, then encouraging more practice at rapid flowing speech would seem appropriate. Although there is a vast literature in second language acquisition suggesting methods for increasing reading fluency (see Grabe & Stoller, 2013, for an overview), very little has been written about promoting oral fluency (some exceptions include Brown, 2003; Gatbonton & Segalowitz, 1988, 2005). Indeed, Gatbonton and Segalowitz (2005) argued that the kind of repetitive practice that is necessary to develop automatic fluency, defined as "the smooth and rapid production of utterances, without undue hesitations and pauses" (p. 326), is often seen as counterproductive under current Communicative Language Teaching methodology. They have suggested creative ways to resolve the apparent conflict. The findings reported here suggest that more attention be given to the development of fluency in the language classroom and call for more research-based practice.

L2 Utterance Fluency, L2 Proficiency Level, and Changes in L2 Utterance Fluency Over Time

The current study also examined how L2 utterance fluency evolved over time for speakers as they improved from one L2 proficiency level to another. In particular, 40 participants' pretest and posttest OPIs were examined to address this evolution in L2 fluency. As expected, some of these participants gained in L2 proficiency from pretest to posttest, and some did not. Incremental improvements in L2 proficiency did occur

with concomitant changes in L2 utterance fluency for two of the measures (faster speech rate and longer run length). These two measures also predicted L2 proficiency levels in general, as indicated in the findings regarding the first research question.

Changes in L2 utterance fluency were also compared across those who improved from pretest to posttest and those who did not. It is apparent, given the large standard deviations and limited improvement in certain L2 utterance features (e.g., number of hesitations and false starts), that changes in these areas of L2 utterance fluency were unrelated to improvements in overall oral L2 proficiency. By contrast, comparing speech rate improvements across the gainers and nongainers certainly demonstrated a vast difference across these two groups—and low standard deviations in the gainers' changes in speech rate also suggests consistency in improvement in this area. Indeed, future research could investigate if learners with similar levels of L2 accuracy and complexity would be rated higher in L2 proficiency simply because they have a faster speech rate or fewer pauses or if these differences are simply a result of greater oral L2 proficiency. Pedagogically speaking, it might be possible to formally or informally estimate whether learners have gained in terms of proficiency over time by precisely measuring or even roughly estimating how fast they speak and how much they are able to speak between pauses.

Changes in L2 Utterance Fluency Across L2s

Finally, the study examined possible differences in how utterance measures of fluency changed from pretest to posttest across different L2s. Although it was not possible to perform statistical analyses due to the low number of learners in each group, variations did seem to occur across L2 learners of German and French, at the very least. In particular, French L2 gainers seemed to exhibit much faster speech rates and longer run lengths than French L2 nongainers. By contrast, the German L2 gainers did not appear to differ very much from the German L2 nongainers in these two areas of utterance fluency. Continuing research comparing the L1 and L2 of both native and nonnative speakers is necessary to explore these differences in more detail. Knowledge of the target language norms and of learners' typical development (e.g., the fact that German learners may exhibit less development in utterance fluency as they reach higher levels than their French-learning peers) can certainly help instructors as they seek to encourage both fluency and proficiency. They can set realistic expectations that take into consideration both the target language and learner tendencies. It may also be that there are tendencies in practices related to pedagogy and assessment that are influencing these cross-linguistic differences and that may merit reconsideration.

Conclusion

This study represents the first attempt to examine measures of L2 utterance fluency across several different L2s and several different L2 proficiency levels. Results indicated that measures of L2 utterance fluency can predict broad differences in L2 proficiency but do not correlate precisely with OPI proficiency ratings, especially at lower sublevels. Thus, for instructors and institutions interested in finding broad measures of L2 proficiency to determine, e.g., whether L2 learners have reached Intermediate or Advanced levels, automatic measures of L2 utterance fluency may provide sufficient data. However, for a more refined analysis of learners' L2 abilities, a more comprehensive measure such as the OPI may yet be necessary. Moreover, this study also found that L2 utterance fluency differs depending on the L2 studied, suggesting that L2 fluency measures may be language specific rather than universal. Language educators certainly need to be sensitive to these differences as they prepare their learners, making them aware of native norms and helping them to develop greater automaticity, allowing

them to pause and subsequently produce more or less speech when appropriate. Finally, improvements in fluency were more apparent for those who made gains on the OPI than for those who did not and, even for gainers, fluency changes were more apparent in some languages than others. Awareness of these trends can help language educators as they seek to identify and understand learner progression and as they reflect on the qualities they emphasize in their teaching and assessment. Finally, as a next step, researchers should measure the utterance features of native speakers to find those features of L2 utterance fluency that differ from language to language. Doing so will greatly inform language instructors as they help learners understand and move toward native fluency.

Notes

1. The ACTFL scale consists of four levels, the first three of which are also divided into the sublevels Low, Mid, and High. For this study, participants fell into nine L2 proficiency levels: Novice Mid, Novice High, Intermediate Low, Intermediate Mid, Intermediate High, Advanced Low, Advanced Mid, Advanced High, and Superior.
2. As shown in Table 1, there are fewer participants at the novice level than at the other levels, which is a limitation of this study.

References

Alcala, Y. (2009). Discontinuous discourse construction in German native speakers: Aspects of the relationship between verbalization and reference procedures. *Revista de Filologia Alemana*, *17*, 159–181.

ACTFL. (2010, June 10). *ACTFL announces ACE college credit recommendation for official ACTFL OPI rating*. Retrieved July 11, 2014, from http://www.languagetesting.com/american-council-on-education

Beigi, H. (2009). *Computer rating of oral test responses using verbosity*. Report No. RTI-20091211-01. Yorktown Heights, NY: Recognition Technologies.

Boersma, P., & Weenink, D. (2011). Praat (Version 5.3.03) [Software]. Retrieved April 23, 2014, from http://www.fon.hum.uva.nl/praat/download_win.html

Borges de Almeida, V. (2009). Filled-in pauses and prosodic domains: Evidence for the validation of the fluency descriptor of an oral proficiency test in a foreign language. *Revista de Linguistica*, *53*, 167–193.

Brand, C., & Gotz, S. (2011). Fluency versus accuracy in advanced spoken learner language: A multi-method approach. *International Journal of Corpus Linguistics*, *16*, 255–275.

Brown, J. D (2003, May 10–11). Promoting fluency in EFL classrooms. In *Proceedings of the second annual JALT Pan-SIG Conference*. Retrieved August 13, 2014, from https://jalt.org/pansig/2003/HTML/Brown.htm

Chou, F.-C., Tseng, C.-Y., & Lee, L.-S (1996, August). *Automatic generation of prosodic structure for high quality Mandarin speech synthesis*. Paper presented at the 1996 International Conference on Spoken Language Processing, Philadelphia, PA. Retrieved August 7, 2014, from http://www.asel.udel.edu/icslp/cdrom/vol3/317/a317.pdf

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, *37*, 201–242.

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*, 1–21. doi:10.1017/S0142716413000210

de Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, *34*, 893–916. doi:10.1017/S0142716412000069

de Jong, N., & Perfetti, C. A. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning*, *61*, 533–568.

DeKeyser, R. (2010). Monitoring processes in Spanish as a second language during a study abroad program. *Foreign Language Annals*, *43*, 80–92.

Derwing, T. M., Munro, M. J., & Thomson, R. I. (2007). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, *29*, 359–380.

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

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.

Fillmore, C. (1979). On fluency. In C. Fillmore, D. Kempler, & W. Wang (Eds.), *Individual differences in language ability and language behavior* (pp. 85–101). New York: Academic Press.

Freed, B. F. (1995). What makes us think that students who study abroad become fluent? In B. F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 123–148). Philadelphia, PA: John Benjamins.

Freed, B. F., Segalowitz, N., & Dewey, D. P. (2004). Context of learning and second language fluency in French. *Second Language Acquisition*, *26*, 275–301.

Garcia-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.

Gatbonton, E., & Segalowitz, N. (1988). Creative automatization: Principles for promoting fluency in a communicative framework. *TESOL Quarterly*, *22*, 473–492.

Gatbonton, E., & Segalowitz, N. (2005). Rethinking communicative language teaching: A focus on access to fluency. *Canadian Modern Language Review*, *61*, 325–353.

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*, *3*, 379–399.

Goldman-Eisler, F. (1961). The distribution of pause duration in speech. *Language and Speech*, *4*, 232–237.

Grabe, W., & Stoller, F. L. (2013). *Teaching and researching reading*. New York: Longman.

Guillot, M. N. (1999). *Fluency and its teaching*. Clevedon, UK: Multilingual Matters.

Higgins, D., Xi, X., Zechner, K., & Williamson, D. (2011). A three-stage approach to the automated scoring of spontaneous spoken responses. *Computer Speech and Language*, *25*, 282–306.

Hilton, H. (2009). Annotation and analyses of temporal aspects of spoken fluency. *CALICO Journal*, *26*, 644–651.

Jacewicz, E., Fox, R. A., O'Neill, C., & Salmons, J. (2009). Articulation rate across dialect, age, and gender. *Language Variation and Change*, *21*, 233–256.

Kaiki, N., & Sagisaka, Y. (1992, October). *Pause characteristics and local phrase-dependency structure in Japanese*. Paper presented at the International Conference on Spoken Language Processing, Banff, Alberta, Canada. Retrieved April 17, 2013, from http://www.isca-speech.org/archive/icslp_1992/i92_0357.html

Kormos, J. (2006). *Speech production and second language acquisition*. New York: Routledge.

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

Larsen-Freeman, D. (2009). Adjusting expectations: The study of complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, *30*, 579–589.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, *40*, 387–412.

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

Li, W., & Yang, Y. (2009). Perception of prosodic hierarchical boundaries in Mandarin Chinese sentences. *Neuroscience*, *158*, 1416–1425. doi:10.1016/j.neuroscience.2008.10.065

Liskin-Gasparro, J. E. (1982). *ETS oral proficiency testing manual*. Princeton, NJ: Educational Testing Service.

Omaggio Hadley, A. (2001). *Teaching language in context*. (3rd ed.). Boston: Heinle & Heinle.

Osborne, J. (2007). Investigating L2 fluency through oral learner corpora. *Linguistic Insights—Studies in Language and Communication*, *51*, 181–197.

Pienemann, M. (1998). *Language processing and second language development: Processability theory*. Amsterdam: John Benjamins.

Quimbo, F. C. M., Kawahara, T., & Doshita, S. (1998, November-December). *Prosodic analysis of fillers and self-repair in Japanese speech*. Paper presented at the 1998 International Conference on Spoken Language Processing, Philadelphia, PA. Retrieved March 22, 2014, from http://www.mirlab.org/conference_papers/International_Conference/ICSLP%201998/PDF/AUTHOR/SL980762.PDF

Riggenbach, H. (Ed.). (2000). *Perspectives on fluency*. Ann Arbor: University of Michigan.

Rossiter, M. J. (2009). Perceptions of L2 fluency by native and non-native speakers of English. *Canadian Modern Language Review*, *65*, 395–412.

Scanlon, T. (1987). Improving fluency in spoken French through a study of native pause behavior. *Foreign Language Annals*, *20*, 345–352.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, *14*, 357–385.

Segalowitz, N. (2007). Access fluidity, attention control, and acquisition of fluency in a second language. *TESOL Quarterly*, *41*, 181–186.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: 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.

Skehan, P. (2001). Tasks and language performance. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 167–185). New York: Longman.

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

Snellings, P., van Gelderen, A., & De Glopper, K. (2002). Lexical retrieval: An aspect of fluency second language production that can be enhanced. *Language Learning*, *52*, 723–754.

Taguchi, N. (2008). The role of learning environment in the development of pragmatic comprehension. *Studies in Second Language Acquisition*, *30*, 423–452.

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.

Thompson, I. (2013). *Language learning difficulty* [Electronic article]. Retrieved April 24, 2013, from http://aboutworldlanguages.com/language-difficulty

Towell, R. (2012). Complexity, accuracy and fluency from the perspective of psycholinguistic second language acquisition research. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 47–70). Amsterdam: University of Brussels, University of Amsterdam.

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

Trenchs-Parera, M. (2009). Effects of formal instruction and a stay abroad on the acquisition of native-like oral fluency. *Canadian Modern Language Review*, *65*, 365–393.

Ushigusa, S. (2009). *The relationship between oral fluency, multiword units, and proficiency scores* (Unpublished doctoral dissertation). Purdue University, West Lafayette, IN.

van Gelderen, A. (1994). Prediction of global ratings of fluency and delivery in narrative discourse by linguistic and phonetic measures: Oral performances of students aged 11–12 years. *Language Testing*, *11*, 291–319.

Vercellotti, M. L. (2012). *Complexity, accuracy, and fluency as properties of language performance: The development of multiple subsystems over time and in relation to each other* (Unpublished doctoral dissertation). University of Pittsburgh, PA.

Yoon, S. Y., Pierce, L., Huensch, A., Juul, E., Perkins, S., Sproat, R., & Hasegawa-Johnson, M. (2009). Construction of a rated speech corpus of L2 learners' spontaneous speech. *CALICO Journal*, *26*, 662–673.

Submitted July 9, 2014

Accepted August 19, 2014