revista.pucsp.br/esp | ISSN: 2318-7115

The Especialist

Volume 43 | Número 1 | Ano 2022

Does um, the medium, uh, matter? Measuring Fluency through filled pauses in Face-to-Face and Teletandem communicative groups

O, hum, meio, uh, importa? Medindo fluência por meio de pausas preenchidas em grupos de comunicação face-a-face e em teletandem.

Celia Chomón, ZAMORA (ACTFL)[1]
Abbie, FINNEGAN (GU)[2]

ABSTRACT

The frequency of hesitational phenomena, such as Filled Pauses (FP) are inversely correlated to the perceived fluency of L2 learners (e.g., DERWING; ROSSITER; MUNRO; THOMSON, 2004; LENNON, 1990; ULLAKONOJA, 2008). The current study investigated the production of these FPs by 66 naive Spanish L2 learners with an L1 of English in two types of oral tasks and compared their frequency in regard to task type and whether they participated in Face-to-Face or Telecollaboration (Teletandem) communicative dyads. Results indicate that while the medium of the tasks did not impact fluency, the task type did.

Keywords: Filled Pauses, Fluency, Hesitation, Telecollaboration, Instruction

RESUMO

A frequência de fenômenos hesitações, como Pausas preenchidas (PP) são inversamente correlacionadas à fluência percebida de alunos L2 (e.g., DERWING; ROSSITER; MUNRO; THOMSON., 2004; LENNON, 1990; ULLAKONOJA, 2008). O presente estudo investigou a produção dessas PPs por 66 aprendizes ingênuos de espanhol L2 com L1 de inglês em dois tipos de tarefas orais e comparou sua frequência em relação ao tipo de tarefa e se eles participaram de Face-a-Face ou Telecolaboração (Teletandem) díades comunicativas. Os resultados indicam que enquanto o meio das tarefas não afetou a fluência, o tipo de tarefa sim.

Palavras-Chave: Pausas preenchidas, Fluência, Hesitação, Telecolaboração, Ensino

1. Introduction

Fluency, regarded as a phonological phenomenon and automatized procedural ability, is the third dimension of the paramount SLA measure, CAF, composed also of *complexity* and *accuracy* (e.g., CARLSON; SULLIVAN; SCHNEIDER, 1989; HOUSEN; KUIKEN; VEDDER, 2012). It primarily

[1] ACTFL, Alexandria, Virginia, USA; ORCID: https://orcid.org/0000-0001-8110-3084; e-mail: czamora@actfl.org
[2] Georgetown University, Washington, D.C., USA. Department of Spanish and Portuguese; ORCID: https://orcid.org/0000-0003-2711-9098; e-mail: Aef90@georgetown.edu

refers to a speaker's "global language proficiency" (HOUSEN; KUIKEN; VEDDER, 2012, p. 4), that is, the higher command or more expertise the speaker has in the foreign language, the more fluent the speaker is expected to be (LENNON, 1990, 2000), as it is expected from native speakers. Similarly, Pawley and Syder (1983) have defined native-like fluency as comparable to the native speaker's "the ability to produce fluent stretches of discourse" (p. 191). As such, fluency is defined generally in terms of the 'native-likeness' of speech, the ease and flow of communication, and the level of eloquence in their oral or written productions (e.g., CHAMBERS, 1997; FREED, 2000; GUILLOT, 1999; HILTON, 2008; KOPONEN; RIGGENBACH, 2000; LENNON, 1990). Fluency is largely categorized in three dimensions: speed fluency (e.g., rate and/or density of linguistic units produced), breakdown fluency (e.g., number, length, and location of pauses), and repair fluency (e.g., false starts, misformulations, self-corrections, and repetitions) (TAVAKOLI; SKEHAN, 2005).

Breakdown fluency, otherwise known as disfluencies, or hesitational phenomena (HP) for the purposes of this study, is defined as a disruption in the flow of speech, which can be silent or realized in the form of a lexical or non-lexical item, and usually indicates either a disfluency or a hesitation. They are classified into three categories: unfilled pauses (UP), which are silent pauses over 400 ms (RIGGENBACH, 1991), filled pauses (FP) which are subcategorized into two groups: 1) lexical (e.g. *well, like, let's see*) and 2) non-lexical (e.g. *um, uh, er, eh*) (CLARK; FOX TREE, 2002); and asides (*God, what? I don't know*). The primary focus of the current study is the second category of filled pauses.

2. Filled Pauses

Although the implications of FPs are language-specific, they are phonetic manifestations of the internal, cognitive processes involved in speech production, making their uses and characteristics similar across languages (e.g., GOLDMAN-EISLER, 1961, 1968; LEVELT, 1983; MACLAY; OSGOOD, 1959; SWERTS, 1998). For instance, although the sounds of the production of FP may vary regularly, they share various parallel qualities, such as:

- **a)** all fillers tend to be brief, usually between one to two syllables, and correspond to the central vowels in the language (CLARK; FOX TREE, 2002),
- **b)** one or more of the fillers are contrastive, usually with a nasal (/m, n/) sound (CLARK, 1996),
- **c)** having the most common fillers be easy to prolong (e.g. *u:m, u:h, li:ke)* if a minor delay in the conversation is required (CLARK; FOX TREE, 2002),
- **d)** possessing more than one filler in order for speakers to distinguish one type of FP from another, such as a FP that differentiates a longer pause from a short one (CLARK; FOX TREE, 2002).

In general, the majority of FP are usually produced in circumstances where the speaker faces numerous syntactic or semantic possibilities for production (e.g., SCHACHTER; CHRISTENFELD; RAVINA; BILOUS, 1991), in cases of complex noun phrases (e.g., CLARK; WASSOW, 1998) in longer utterances (e.g., SHRIBERG, 1994; OVIATT, 1995), as well as in situations where uncertainty is present, such as the content of the utterance (e.g., BRENNAN; WILLIAMS, 1995). Furthermore, FPs usually have a tendency to precede low frequency and unpredictable words (e.g., BEATTIE; BUTTERWORTH, 1979; LEVELT, 1983; SCHACHTER; CHRISTENFELD; RAVINA; BILOUS, 1991; SCHNADT; CORLEY, 2006).

3. Filled Pauses as a Measure of Fluency

Previous psycholinguistic research in oral production (e.g., BEATTIE, 1980; GOLDMAN-EISLER, 1968; GOOD; BUTTERWORTH, 1980; KOWAL; O'CONNELL, 1980; LEVELT, 1989) has created a baseline for native-speaker (NS) fluency and disfluency and found that fluent NS produce between 130 to 200 words per minute (which translates to 2-3 words per second), and approximately a third of that production time is occupied by pauses (Hilton, 2008). However, a baseline for disfluency, also referred to as "clinical disfluency" (HILTON, 2008), has been described as a speaker producing fewer than 50 words per minute. In this type of disfluent speech, "there are more pauses, which are longer and distributed differently" than those of a fluent NS (HILTON, 2008, p. 154), thereby causing shorter, less-articulate speech runs syntactically and meaning-wise.

The aforementioned, although initially referring to native speakers, also applies to foreign language (FL) and second language (L2) learners. Pausing, whether filled or unfilled, and other disfluencies are found to be just as common in nonnative speech (ULLAKONOJA, 2008). In fact, research into HP has found that the most substantial markers of fluency in the L2 is speech rate and mean length of run (MLR), or the average number of words spoken between two pauses of 0.25 seconds or longer (e.g., CUCCHIARINI; STRIK; BOVES, 2002; HILTON, 2008; EJZENBERG, 2000; FREED, 1995, 2000; LENNON, 1990; RIGGENBACH, 1991; TOWELL; HAWKINS; BAZERGUI, 1996). Fluency, as Lennon (1990) proposed, is "the highest point on a scale that measures spoken command of a FL" (p. 389). Several researchers have attempted to pinpoint what measures of fluency, if any, correlate the most with the level of perceived fluency of L2 and FL speakers. These studies have found intriguing results regarding the relationship between HP and fluency, such as: Perceived improvements over time are attributed to an increase in speech rate and a reduction in UP, FP, and repetitions (e.g. LENNON, 1990; O'BRIEN, 2014); fluency ratings being heavily influenced by speech rate and pausing (e.g. DE JONG;

PERFETTI, 2011; HANSEN; DECHENE; WANKE, 2008; RIGGENBACH, 1991); fluency judgments are predicted significantly by the frequency and duration of self-corrections and HP, such as FP, UP, and false starts (e.g. DERWING; ROSSITER; MUNRO; THOMSON, 2004; DE JONG; STEINEL; FLORIJIN; SCHOONEN; HULSTIJN, 2013; VAN GELDEREN, 1994); learners with fewer and shorter pauses are perceived to be more fluent (ULLAKONOJA, 2008); and that hesitation clusters have a negative effect on perceptions of fluency (RIGGENBACH, 1991). For instance, in Bosker et al. (2013), both pause frequency and pause length correlated negatively with fluency ratings. However, in Kormos and Deńes (2004) fluency ratings did not correlate with pause frequency but did correlate with pause length. By contrast, in Cucchiarini, Strik and Boves (2002), the opposite pattern was found; in other words, fluency ratings correlated with pause frequency but not with pause length.

4. Previous Studies on Filled Pauses as a Measure of Fluency

Freed, Segalowitz, and Dewey (2004) compared fluency ratings of learners in traditional classroom, study-abroad, and intensive immersion environments using several fluency measures, including: 1) Hesitation-Free Speech runs, reflecting the mean number of words produced without disfluencies consisting of UP of 400 ms or more; 2) Filler-Free Speech runs, reflecting the mean number of words produced without FPs (regardless of how long the pause was); 3) Fluent runs, the total number of words in the longest run of speech without UP or FP; and other measures, such as total number of words; duration of speaking time, and so forth. The purpose of the study was to see if there were any noticeable, salient differences in the levels of fluency acquired by the students in the three aforementioned learning conditions, as well as if the measured differences in fluency were associated with any time-on-task features they looked at (e.g. instructional time, time spent interacting with NS outside of class).

The participants were 28 NS of English learning French in one of the three learning environments (traditional classroom setting, study abroad, or intensive immersion), all sponsored by the same university. The data consisted of interviews ranging from 15 to 30 minutes in a manner similar to the Oral Proficiency Interview (OPI) recorded at the beginning and at the end of the semester. Although the OPI ratings were not utilized as a measure, 2 1-minute segments were extracted from each of the participant's pre- and post-test recordings, totaling 4 minutes. Participants also filled out a Language Contact Profile (LCP) to note and analyze how students spent their time outside of the learning environment.

The results showed that participants did not significantly differ in any of the fluency measures except for the Filler-Free and Fluent runs, both of which involve unfilled and filled pauses. Freed, Segalowitz and Dewey (2004) attributed this finding to the fact that FPs generate the perception that the

participant is “stumbling and stalling for time” (p. 294), and furthermore that the “extent that fluency is defined by hesitation and temporal phenomena that are both perceived by listeners and supported by empirical fact as faster, smoother uninterrupted speech” (p. 294). The context of learning that seemed to mostly facilitate this was the intensive immersion setting, followed by the study abroad context, and lastly, the traditional classroom setting. Additionally, multiple regression analyses indicated a positive correlation between the participants who gained the most in fluency measures and those who reported spending hours per week writing in French outside of the classroom setting.

In a different context, Hilton (2008) used a PAROLE corpus of university students (English, n=33; Italian, n=11; French, n=12) describing a video they had just watched in their L2. After transcribing recordings to ensure that all HP were included, Hilton found that lower-level L2 speakers who were considered disfluent, hesitated longer and more frequently than fluent speakers and that more frequent clause-internal hesitations appeared to be characteristic of L2 learners, and that of disfluent speech. Similarly, Ullakonoja (2008) also investigated fluency, particularly in hesitation phenomena, in the Russian read-aloud speech of 12 Finnish university students, particularly looking at their fluency development during a 3.5-month study period in Russia. Participants were recorded reading the same dialogue with the same partner 3 different times: before, during, and after their 3.5-month stay in Russia. However, only the most problematic section of the dialogue (lasting approximately 12 minutes) was chosen for analysis. 30 Finnish teachers of Russian were instructed to assess the fluency of these participants based on the recordings, and the samples were analyzed based on their perceptions for the frequency, duration, and placement of the pauses.

As summarized by the author, results of the study showed that, “the frequency of the fluent pauses remained the same and the frequency of the disfluent pauses decreased as the amount of experience increased” (ULLAKONOJA, 2008, p. 340). While that was the case in general, results indicated that individual differences also contributed to the frequency of the pauses. Furthermore, Ullakonoja found a statistical negative correlation between the number of pauses and the ratings of fluency received. Although the frequency of pauses varied depending on the individual, overall, the participants’ pauses were generally longer compared to the length of the pauses produced by native speakers in Volskaya’s (2004) study, which was utilized as a baseline. Lastly, although participants generally scored higher fluency ratings on the last reading, some students had the opposite results. Ullakonoja attributed this decline to the participants becoming “more conscious of their pronunciation and hence are trying to self-correct more, which causes more repairs and disfluent pauses” (p. 341-2).

Finally, O’Brien’s (2014) study investigated fluency judgments made by nonnative L2 learners of German on recorded samples of native and nonnative speakers of German, with the purpose of seeing

which fluency measures are the most influential in how fluency ratings are given. The recordings included the production of of 24 native English speakers learning German as an L2, who were residing in Germany in a study abroad program at the time of the recording; and 24 native speakers of German, who were all from Frankfurt and enrolled at the university. All of the participants (n=48) were asked to record a description of a series of pictures. The recordings were then rated by other L2 German participants (n=25), who had never spent any time in Germany. The participants were asked to rate the recordings on accentedness and when rating fluency, they were asked to focus on disfluencies (e.g., pauses). The measures of fluency included: 1) total number of FP, 2) total number of UP, 3) repetitions and self-corrections, 4) speech rate, and 5) articulation rate.

Results showed that the listeners were able to discern the recorded samples produced by the native speakers from those produced by the nonnative speakers. The ratings appointed to the recordings produced by the native speakers were not only anticipated by most of the common disfluency markers (self-corrections and repetitions, filled and unfilled pauses, and speech rate), but also by other factors, such as errors in the lexicon, syntax, and morphology. The results, O'Brien (2014) stated, were not surprising, "given that the speech samples of native and nonnative speakers differed significantly in terms of all the common disfluency markers" (p. 738). The predicted assigned fluency ratings also all accorded with previous research (e.g., DE JONG; STEINEL; FLORIJN; SCHOONEN; HULSTIJN, 2013; DERWING; ROSSITER; MUNRO; THOMSON, 2004; LENNON, 1990; RIGGENBACH, 1991; ROSSITER, 2009).

Although a plethora of studies provide evidence that filled and unfilled pauses are imperative markers for fluency, a few studies have reported the opposite outcome. One such study, Kormos and Dénes (2004), for instance, reported that the number of UPs, FPs, and other disfluencies were not found to substantially influence perception of fluency for these raters. They claimed that, "most of [the previous] studies … suffer from several methodological shortcomings, partly because they use very few subjects and fail to employ reliable methods of analyzing the duration of pauses" (p. 146). The participants were 16 Hungarian EFL learners in a university in Budapest, belonging to two proficiency groups: the advanced (n=8) and the intermediate (n=8). The judges were six: 3 native speakers of English, and 3 Hungarian nonnative speakers of English, who were experiences university professors of applied linguistics, and examiners of the language proficiency exams at the school. The participants were asked to perform a narrative task, where they had to make up a story relevant to a selected cartoon. All participants were given the same task and cartoon, in order to control for the content variable. The task was recorded, and the speech samples were 2-3 minutes long on average. The study measured fluency through the following: 1) number of UP per minute, 2) mean length of pauses, 3) number of FP per minute, and 4) number of disfluencies per minute (e.g., repetitions, restarts, repairs). The results showed that the NNS teachers rated

fluency based primarily on the naturalness and ease of flow of speech, the presence or lack of pauses, and self-corrections; whereas the NS teachers based their scores on the speed of delivery, the absence or presence of HP, and varied as to the importance of lexical variety and accuracy. The paramount predictors were speech rate, phonation time ration, the mean length of run, and the number of stressed words per minute. The number of UP, FP, and other disfluencies were not found to substantially influence perception of fluency for these raters. Nevertheless, according to Kormos and Dénes, the "mean length of pauses is also significantly related to the composite native and non-native fluency scores, [although] it does not correlate with all of the individual raters' scores" (p. 161).

While these studies have shed some light on the role of filled pauses in L2 production, there are some methodological limitations to these studies. For instance, it is not clear in some studies of what the rater's perceptions are regarding the assessment of fluency, as many participants and raters often associate fluency with proficiency and assessed the samples in that light instead (O'BRIEN, 2014). Additionally, the relatively low number of participants in many of the previous studies also presents a methodological limitation lowering the level of confidence in the reported findings.

The review also reveals at least three other areas less explored in the study of filled pauses in L2 oral production. The first is the need to investigate filled pauses as two distinct points in time to see if changed in gains in language development will affect fluency to complement the one-shot design employed in previous studies. The second is to address whether type of medium (e.g., video-based telecollaboration or Teletandem vs. face-to-face or FTF) may play in the production of filled pauses. While previous research has investigated the role of type of medium in comprehension ability (e.g., SAITO; AKIYAMA, 2018) and L2 development (e.g., SUH; LEOW, 2020; see ZIEGLER, 2016 for a recent meta-analysis), to our knowledge, no previous study has addressed the issue of pauses in telecollaboration, an area that clearly warrants further investigation. The third gap is type of oral task. The importance of employing multiple assessment tasks within the same research design to increase confidence in the results has been underscored in previous studies (e.g., LEOW, 1998; SHOOK, 1994). To our knowledge, only one previous study (SUH; LEOW, 2020) has addressed the role of type of assessment task (in L2 development) within the same design involving both Teletandem and FTF. Suh and Leow employed two types of oral productive assessment tasks that varied along the dimensions of openness and ecological validity: a picture description task, which they described as a relatively closed and non-ecologically valid task, and an opinion-stating task that was more open-ended and ecologically valid task that was frequently performed by participants among other activities in the classroom. Whether fluency is impacted by type of task also requires future investigation.

There is a clear need for future research to (a) probe deeper into fluency, and more specifically, filled pauses in the L2 not only from a one-shot perspective but also progressively throughout the semester, (b) compare two media of interaction in the L2 (Teletandem and FTF) given that within-speaker discrepancies regarding disfluency measures "vary according to discourse topic, situation, interlocutor, and so forth" (LENNON, 2000, p. 25), and (c) employ multiple assessment tasks to address the role of type of task. To this end this study aimed to investigate these issues by addressing the following research questions:

Research Question 1: *Does the frequency of Filled Pauses (FP) within individual speakers differ based on the medium of communication (Face-to-Face (FTF) vs. Teletandem) from the beginning to the end of the semester?*

Research Question 2: *Does the type of task affect fluency through the production of Filled Pauses?*

5. Methodology

The participants of this study were 66 students enrolled in an Advanced I Spanish course at a mid-Atlantic university. They were randomly assigned to one of two experimental study conditions: a Face-to-Face (n=34) or a Teletandem (n=32) group. 9 participants' data was excluded from the final analysis, as they did not complete either the pretest or the posttest, resulting in n=57.

Teletandem, a form of synchronous computer-mediated communication (SCMC), is a type of telecollaboration, which is "a virtual, collaborative and autonomous context for learning foreign languages in which two students help each other to learn their own languages" (TELLES, 2015, p. 604). This mode of communication affords learners the opportunity to have autonomy over their language learning, as well as the ability to work at their own pace. Learners can connect with L1 speakers of the language by connecting through a webcam or voice messaging program in order to communicate. Previous studies on SCMC have suggested that it allows for learners to be engaged in the target language, simulating a 'face-to-face' environment, regardless of the physical distance, and that it "provides students with an authentic way to learn the target language through communicative use, reduces students' anxiety of talking face-to-face, and gives students more opportunities to express their ideas than oral communication does" (LIN; HUANG; LIOU, 2013, p. 123).

5.1 Materials

The study consisted of two types of controlled oral production tasks (guided storytelling vs. open-ended content-based task) adapted from those employed in Suh and Leow (2020).

For the guided story-telling task, participants narrated a series of drawings and a set of questions based on the content that was currently covered in the Advanced I course. The story-telling task remained unchanged and was administered before and after experimental exposure and the open-ended questions differed before and after experimental exposure. The content-based task was a set of questions based on the content that was currently covered in the course (e.g., democracy and dictatorship in the Spanish world).

5.2 Procedure

Participants completed the two oral production tasks in a longitudinal, test 1- test 2- test 3 design. They were randomly assigned to one of 2 groups (FTF or Teletandem) and once a week for a period of 2 months, they interacted in dyads with the same assigned conversational topics that pertained to the current course content. The FTF dyads met in their regular classrooms and the Teletandem dyads connected via Skype in the language laboratory with students at a university in Mexico. The tasks were recorded with QuickTime Player. Files were saved in MP3 format and were later coded and analyzed using Fission sound editing software to measure filled pauses. The Spanish language interactions lasted for 20 minutes in both groups. Although the Teletandem dyads had a total of 40 minutes, the other 20 dedicated to English. Originally, 8 sessions were planned and only 6 sessions were completed, which amounted to a total of 2 hours of controlled interaction. The data collected from the mid-semester tasks were not analyzed for the purposes of the current study.

5.3 Coding

The recordings of the controlled oral tasks were transcribed by a team of Spanish linguistics graduate students in order to obtain the FP ratios. The transcriptions included hesitation phenomena, which were then coded by 2 researchers. Interrater agreement was 100%. For the purposes of the current study, the transcriptions were coded for both types of FP, which are lexical and non-lexical. Because solely audio, not video, recordings were collected during the oral production tasks, the researchers in the current study excluded UP, since they could not decipher if the unfilled pauses were caused by external factors such as distractions. However, when filled pauses alternated with unfilled pauses in sequence

(independent of content words), they have been analyzed and timed as a whole hesitation cluster (HILTON, 2008). The ratios were derived from the number of filled pauses over the number of words spoken. While ratios of both filled pauses being counted in the total word count and ratios with filled pauses *not* being counted in the total word count were calculated, the current data displays the FPs being considered in the total word count.

For the purposes of the current study, FPs were defined and coded as both lexical and non-lexical. Table 1 contains examples of FPs.

Table 1. Types of Filled Pauses (FP) and examples

Types of FP	Examples
Lexical	*let's see, well, in that case, so, like, entonces, bueno, pero*
Non-lexical	*uh, u:h, um, u:m, eh, em, er, erm, mm*
Hesitation Cluster	*uh...u:h, um...u:m, uh...er, mm...um*

This study implemented constructs of oral fluency based on previous literature on measures of HP (e.g., LENNON, 1990; MÖHLE, 1984; OLYNYK; D'ANGLEJAN; SANKOFF, 1990; RAUPACH, 1984, 1987; RIGGENBACH, 1991; SAJAVAARA, 1987; FREED; SEGALOWITZ; DEWEY, 2004; O'BRIEN, 2014). The analysis consisted primarily of an investigation of FP rather than a holistic view of fluency (e.g., speech rate, duration, etc.); therefore, only the number of FP were included in the data analysis. FPs were coded for each participant in 2 ways: 1) as the total sum, 2) as a ratio of the number of FPs over the number of total words (including the FPs in the total word count). While previous research on FPs did not include whether FPs were included in the analysis as words (e.g., LENNON, 1990; MÖHLE, 1984; OLYNYK; D'ANGLEJAN; SANKOFF, 1990; RAUPACH, 1984, 1987; RIGGENBACH, 1991; SAJAVAARA, 1987; FREED; SEGALOWITZ; DEWEY, 2004; O'BRIEN, 2014), the current study opted to include them. Below is an example of the coding procedure.

Participant A, September, Guided Story-Telling Task 1A; Spanish (2:44)

En *el la* **uh** primera fotografía la madre *discut*… estaba… **erm**. La madre está discutiendo con…su esposo sobre educación de la computadora para *sus hijo* su hijo. **Um** ella creo que es una bien idea pa… su hijo aprender **um**… como… **uh**... un computadora funciona y **um**… como *s*… su hijo… **um**… puede usarlo. En el segundo parte **um**… *un* una semana pasada **um**… la madre muestra… **uh** a su hijo como

usar la computadora y… **uh**… cosas diferente de **um**… la computadora. **Um**… (5 sec) a los días **uh**… **eh**…. Su hijo y… a la… **er** el amigo de su hijo **uh**… vamos a… [h]ugar con la computadora y… ella… sus padres… realiza *se*… **um** se realiza *que* que la computadora es una problema porque… **uh**… este situación.. *desar*… se desarrolla a un adicción a la computadora y ahora los hijos **uh**… nada mas jugamos… *o jueguan*… afuera de la casa. Y… sí.

Ratio: 20/155 (4:31); 12.9% 164 seconds total (0.12 per second)

6. Results

The first research question addresses how the frequency of FPs is affected by the medium, face-to-face and teletandem. Table 2 below presents the average numbers of Filled Pauses across both the open-ended task (Task 1) and the spontaneous production task (Task 2) across both media of interaction (Teletandem and Face-to-Face) from the pre and posttest only.

Table 2. Average number of HP present in oral speech samples

Average number of FP	Teletandem (TTM)		Face-to-Face (FTF)	
	Guided	Open-ended	Guided	Open-ended
Pretest (September)	25.71	26.74	26.20	30.04
Posttest (December)	25.70	20.93	25.07	25.69

Based on the data, it is evident that there was a slight decrease in the total number of FPs from the pretest to the posttest in the guided, story-telling task, and a larger decrease of FPs from the pretest to the posttest in the open-ended task. It should be noted that these are the total number of FPs, and do not take into account the number of total words per task.

Table 3. Average of total number of words in per task, across interactions

Average of total words

	Guided	Open-ended
Pretest (September)	242.90	243.31
Posttest (December)	248.48	250.78

The second research question which aimed to compare FPs among different task types is discussed here. Results reveal that in the open-ended task, Mean = .2438, (24%) more FPs were produced by the participants from the pre- to the posttest in the TTM group, whereas 35% fewer FPs (M = -.3582) were present in the final iteration of the story task in the FTF group. Results for the open-ended task show that both the TTM and FTF groups saw a decrease in the presence of FPs from the pretest to the posttest (a decrease of 186% and 182%, respectively).

Table 4. Descriptive statistics for FPs based on medium

Medium	Teletandem (TTM)		Face-to-Face (FTF)	
	Guided	Open-ended	Guided	Open-ended
Mean	.2438	-1.8631	-0.3582	-1.8279
Standard Deviation	4.55818	4.48449	3.34904	7.22799
N	29		28	

Table 5. Multivariate measures

Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power c
Intercept Pillai's Trace	.099	2.966b	2.000	54.000	.060	.099	5.933	.554
Wilk's Lambda	.901	2.966b	2.000	54.000	.060	.099	5.933	.554
Hotelling's Trace	.110	2.966b	2.000	54.000	.060	.099	5.933	.554
Roy's Largest Root	.110	2.966b	2.000	54.000	.060	.099	5.933	.554
Interaction Pillai's Trace	.007	.187b	2.000	54.000	.830	.007	.375	.078
Wilk's Lambda	.993	.187b	2.000	54.000	.830	.007	.375	.078
Hotelling's Trace	.007	.187b	2.000	54.000	.830	.007	.375	.078
Roy's Largest Root	.007	.1887b	2.000	54.000	.830	.007	.75	.078

A one-way multivariate analysis of variance (MANOVA) was conducted to investigate if there would be one or more mean differences between the medium (Teletandem or Face-to-Face) and the difference of ratios between the pretest and posttest. No statistically significant MANOVA effect was obtained, Pillais' Trace = .099, F(2, 54) = 0.188, p > .005. The multivariate effect size was estimated at .099, which implies that 9.0% of the variance in the canonically derived dependent variable was accounted for by the medium of interaction

Table 6. Descriptive statistics for Tasks

Tasks	Guided	Open-ended
Mean	18.295	30.340
Standard Deviation	14.1052	23.9388
N	114	114

With regards to Task type (research question 2), open-ended, content-based tasks appear to result in more FP and breakdown of fluency (Mean= 30.34) than guided story-telling tasks (Mean= 18.295). Based on the One-Way ANOVA, there is a statistically significant difference ($p < .05$) between the means of the guided and open-ended tasks.

7. Discussion

The first research question sought to analyze fluency in two different contexts: A traditional, Face-to-Face setting and a virtual setting utilizing Teletandem. Participants performed similarly with regards to improving fluency across the tasks in the pre- and posttests, yielding insignificant results. This is similar to Ullakonoja's (2008) study, where results showed that the rate of the fluent pauses remained the same, as well as Freed et. al (2004), which found that participants in three distinct learning environments (traditional classroom setting, study abroad, or intensive immersion) did not significantly differ in any of the fluency measures conducted. However, the results indicated a small trend towards the FTF group demonstrating a slight advantage in the improvement of fluency (through the decrease of FPs).

This could be explained by the nature of the study. During the controlled experiment, participants were limited in the amount of time they were allowed to interact with their partners outside of the typical

classroom. Therefore, regardless of whether through a screen or in the traditional classroom, participants were given the same opportunity to communicate, practice, and receive additional input. For future research, it would be fruitful to allow participants autonomy with respect to the amount of contact they have with their partners, as it may be the case that a TTM virtual partner may provide more opportunities for discussion outside of the traditional classroom setting. Now that virtual learning has exponentially grown, one sees the ease and flexibility of being able to coordinate classes or calls with colleagues, which may result in an added benefit to the TTM model, in that students have additional opportunities to practice and communicate, which may not necessarily be afforded in a traditional classroom. At the same time, creating a more autonomous learning condition requires more methodological controlling of potential intervening variables such as amount of discussion and time to arrive at a true gauge of the role of medium. Future studies could include time of interaction or amount of exposure as additional variables to investigate. More research is clearly necessary.

Although no differences were found between learning conditions, both groups improved in fluency over time. The number of FPs from the pretest to posttest decreased in both communicative groups, as was also seen in Ullakonoja (2008), where the frequency of the pauses decreased as the amount of language experience grew. It should also be noted, however, that some participants produced more FPs in their posttest in comparison to their pretest. However, this may be due to these participants also producing more language in the posttest compared to the pretest, which could be interpreted as an increase in confidence with the topic and language. More research is required to be able to identify the potential causes of these results.

The second research question addressed the effect of Task Type on the usage of FP. The results demonstrated a higher number of FPs in the open-ended, content-based activities than they did during the guided story-telling tasks. The difficulty posed by the task type, implied by the increased presence of FPs, is an interesting observation. It is possible that students may be more successful and feel more confident in completing guided oral tasks that afford them the opportunity to use rehearsed knowledge that is gradually built upon versus activities that requite deeper retrieval, thinking, creativity, and some level of spontaneity. While the open-ended prompts were based on information learned in the course, they require more cognitive effort on behalf of the students to create with the language on their own with less assistance than that provided by the guided tasks. Additionally, it could be the case that participants repeating the guided task in both the pre- and posttest faced a potential test effect, which resulted in a decrease of the FPs, although the time between the tests was over a month. To remedy this, future studies should ensure that the guided tasks change, even slightly, between the pre- and posttest. Future research may also want to address the role of individual differences, which could be completed via debriefing interviews and

surveys. It may be the case that language anxiety or confidence were the leading causes in participants producing more FPs, especially since they were aware that they were being recorded. Furthermore, it may be the case that language learners may feel more confident answering open-ended questions due to the potential of being able to perform circumlocution if a word is forgotten or unknown, whereas they may feel more constricted and less confident in answering a question with a clear right or wrong answer, or where they are expected to produce specific words. Finally, it is imperative that future studies also focus on providing an opportunity for participants to record the task in their L1 as a baseline to omit the possibility that participants naturally produce FPs in their daily speech in their first language. Such important data will be able to tease apart participants whose FPs may not be reflective of a breakdown in fluency in order to avoid miscoding of participants' fluency.

Conclusion

This study has contributed to the body of L2 research that investigates fluency by comparing indicators of fluency, in this case, Filled Pauses, across two media of communication (FTF and Teletandem) and task types (a guided storytelling vs. open-ended content-based task). There was no significance difference between the traditional Face-to-Face and virtual Teletandem media in terms of the number of FPs. Furthermore, the open-ended, content-based tasks appeared to elicit more production of FPs as compared to the guided story-telling tasks, which may imply that open-ended oral production tasks may pose a greater cognitive challenge.

While the current results are inconclusive, it is apparent that further research on this topic will yield important pedagogical implications for the L2 classroom. For instance, instructors should take note on how task selection will affect the results of their students' performance. Perhaps instructors should ensure to use various types of oral tasks in order to accurately gauge their students' fluency.

Whether students interact with other language speakers in the target language via a virtual setting such as Teletandem, or in a traditional Face-to-Face setting, what is crucial is that students consistently utilize the opportunity to speak and practice with both their classmates and other speakers of the target language, and that students continuously stay motivated to continue. It appears that regardless of communication occurring via video-based teletandem or a traditional face-to-face setting, the goal is to be able to make communication with other language learners accessible, beneficial, and entertaining with the result of providing more opportunities for conversation and practice. To conclude, well, um, the medium doesn't, uh, matter.

References

BEATTIE, G. W.; BUTTERWORTH, B. L. 1979. Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. *Language and Speech*, 22.3: 201-211.
BRENNAN, S. E.; WILLIAMS, M. 1995. The feeling of Another′ s Knowing: prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of memory and language*, 34.3: 383-398.
CARLSON, R. A.; SULLIVAN, M. A.; SCHNEIDER, W. 1989. Practice and working memory effects in building procedural skill. *Journal of Experimental Psychology:* Learning, Memory, and Cognition, 15.3: 517–526. https://doi.org/10.1037/0278-7393.15.3.517
CHAMBERS, F. 1997. What do we mean by fluency? *System*, 25.4: 535-544.
CLARK, H. H. 1996. *Using language.* Cambridge University Press.
CLARK, H. H.; TREE, J. E. F. 2002. Using uh and um in spontaneous speaking. *Cognition*, 84.1: 73-111.
CLARK, H. H.; WASOW, T. 1998. Repeating Words in Spontaneous Speech. *Cognitive Psychology,* 37: 201-242.
CUCCHIARINI, C.; STRIK, H.; BOVES, L. 2002. Quantitative assessment of second language learners' fluency: comparisons between read and spontaneous speech. *The Journal of the Acoustical Society of America*, 111:6: 2862–2873. https://doi.org/10.1121/1.1471894
DE JONG, N.; PERFETTI, C. A. 2011. Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning*, 61.2: 533-568.
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*, 34.05: 893-916.
DERWING, T. M.; ROSSITER, M. J.; MUNRO, M. J.; THOMSON, R. I. 2004. Second language fluency: Judgments on different tasks. *Language learning*, 54.4: 655-679.
EJZENBERG, R. 2000. The juggling act of oral fluency: A psycho-sociolinguistic metaphor. In: RIGGENBACH, H. (Ed.). *Perspectives on fluency:* 287–314. Ann Arbor, MI: The University of Michigan Press.
FREED, B. 1995. What makes us think that students who study abroad become fluent? In: FREED, B. (Ed.). *Second Language Acquisition in a Study Abroad Context*: 37-66. Philidelphia, PA: John Benjamin.
FREED, B. 2000. Is fluency, like beauty, the eyes (and ears) of the beholder? In: RIGGENBACH, H. (Ed.). *Perspectives on fluency:* 243-265. University of Michigan Press.
FREED, B. F.; SEGALOWITZ, N.; DEWEY, D. P. 2004. Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition,* 26.2: 275-301.
GOLDMAN-EISLER, F. 1961. The distribution of pause durations in speech. *Language and Speech*, 4.4: 232-237.
GOLDMAN-EISLER, F. 1968. *Psycholinguistics:* Experiments in Spontaneous Speech. Academic Press.
GOOD, D. A.; BUTTERWORTH, B. L. 1980. Hesitancy as a conversational resource: Some methodological implications. In: DECHERT, H. W.; RAUPACH, M. (Eds.). *Temporal Variables in Speech*: 145-52. Mouton.
GUILLOT, M. N. 1999. *Fluency and its teaching* (Vol. 11). Multilingual Matters.
HANSEN, J.; DECHENE, A.; WÄNKE, M. 2008. Discrepant fluency increases subjective truth. *Journal of Experimental Social Psychology*, 44.3: 687-691.
HILTON, H. 2008. The link between vocabulary knowledge and spoken L2 fluency. *Language Learning Journal*, 36.2: 153-166.
HOUSEN, A.; KUIKEN, F.; VEDDER, I. 2012. (Eds.) *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*. Amsterdam and Philadelphia: John Benjamins.

KOPONEN, M.; RIGGENBACH, H. 2000. Overview: Varying perspectives on fluency. *Perspectives on Fluency*, 5-24.
KORMOS, J.; DENES, M. 2004. Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32.2: 145-164.
KOWAL, S.; O'CONNELL, D. C. 1980. Pausological research at Saint Louis University. In: DECHERT, H. W.; RAUPACH, M. (Eds.). *Temporal Variables in Speech:* 61-6. Mouton.
LENNON, P. 1990. The advanced learner at large in the L2 community: Developments in spoken performance. *IRAL: International Review of Applied Linguistics in Language Teaching*, 28.4: 309.
LENNON, P. 2000. The lexical element in spoken second language fluency. In: RIGGENBACH, H. (Ed.). *Perspectives on Fluency*: 25-42. University of Michigan Press.
LEOW, R. P. 1998. The effects of amount and type of exposure on adult learners' L2 development in SLA. *Modern Language Journal,* 82: 49-68.
LEVELT, W. J. 1983. Monitoring and self-repair in speech. *Cognition*, 14.1: 41-104.
LEVELT, W. J. 1989. *Speaking: From Intention to Articulation.* MIT Press.
LIN, W.; HUANG, H.; LIOU, H. 2013. The Effects of text-based SCMC on SLA: A meta analysis. *Language Learning & Technology,* 17: 123-142.
MÖHLE, D. 1984. A comparison of the second language speech production of different native speakers. *Second language productions*, 26-49.
O'BRIEN, M. G. 2014. L2 Learners' Assessments of accentedness, fluency, and comprehensibility of native and nonnative German speech. *Language Learning*, 64.4: 715-748.
OLYNYK, M.; D'ANGLEJAN, A.; SANKOFF, D. 1990. A quantitative and qualitative analysis of speech markers in the native and second language speech of bilinguals. In: SCARCELLA, R.; ANDERSEN, R.; KRASHEN, S. (Eds.). *Developing communicative competence in a second language:* 139-155. Newbury House.
OVIATT, S. 1995. Predicting spoken disfluencies during human–computer interaction. *Computer Speech & Language*, 9.1: 19-35.
PAWLEY, A.; SYDER, F. H. 1983. Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *Language and communication*: 191-225.
RAUPACH, M. 1984. Formulae in second language speech production. *Second language productions:* 114-137.
RAUPACH, M. 1987. Procedural knowledge in advanced learners of a foreign language. In: COLEMAN, J.; TOWELL, R. (Eds.). *The advanced language learner:* 123-155. CILT.
RIGGENBACH, H. 1991. Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse processes*, 14.4: 423-441.
ROSSITER, M. J. 2009. Perceptions of L2 fluency by native and non-native speakers of English. *Canadian Modern Language Review/La revue canadienne des langues vivantes*, 65.3: 395-412.
SAITO, K.; AKIYAMA, Y. 2018. Effects of video-based interaction on the development of second language listening comprehension ability: A longitudinal study. *TESOL Quarterly,* 52.1: 163–176.
SAJAVAARA, K. 1987. Second language speech production: Factors affecting fluency. *Psycholinguistic models of production*: 45-65. Ablex.
SCHACHTER, S.; CHRISTENFELD, N.; RAVINA, B.; BILOUS, F. 1991. Speech disfluency and the structure of knowledge. *Journal of Personality and Social Psychology,* 60.3: 362-367.
SCHNADT, M. J.; CORLEY, M. 2006. The influence of lexical, conceptual and planning based factors on disfluency production. In: SUN, R.; MIYAKE, N. (Eds.). *Proceedings of the twenty-eighth meeting of the cognitive science society:* 750-755.
SHOOK, D. J. 1994. FL/L2 reading, grammatical information, and the input-to-intake phenomenon. *Applied Language Learning,* 5.1: 57-93.
SHRIBERG, E. E. 1994. Preliminaries to a theory of speech disfluencies. Unpublished Ph.D. thesis, University of California at Berkeley.

SUH, B-R.; LEOW, R. P. 2020. Second language accuracy development through interaction in video-based telecollaboration and face-to-face contexts and type of assessment task: A curricular approach. *Studies in Foreign Language Education,* 34.3: 1-28.
SWERTS, M. 1998. Filled pauses as markers of discourse structure. *Journal of Pragmatics,* 30: 485-496.
TAVAKOLI, P.; SKEHAN, P. 2005. Strategic planning, task structure, and performance testing. In: ELLIS, R. (Ed.). *Planning and Task Performance in a Second Language*. John Benjamins.
TELLES, J. A. 2015. Learning foreign languages in teletandem: Resources and strategies *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*, 31: 603-632.
TOWELL, R., HAWKINS, R., BAZERGUI, N. 1996. The development of fluency in advanced learners of French. Applied Linguistics, 17.1: 84 119. https://doi.org/10.1093/applin/17.1.84
ULLAKONOJA, R. 2008. Pausing as an indicator of fluency in the Russian of Finnish learners. In: BARBOSA, P. A.; MADUREIRA, S.; REIS, C. (Eds.). *Proceedings of the Speech Prosody 2008 conference:* 339-342. Editora RG/CNPq.
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*, 11.3: 291-319.
VOLSKAYA, N. B. 2004. O pauze i ne tol'ko o ney. In: ZINDERA, L. R.; BONDARKO, L. V. (Eds.). *Foneticheskye chitenya v chest' 100-letya so dnya rozhdenya:* 129-136. Filologicheskiy fakul'tet SPbGU.