Fluency in monologic and dialogic task performance: challenges in defining and measuring L2 fluency

Article

Accepted Version

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Fluency in monologic and dialogic task performance: Challenges in defining and measuring L2 fluency

Abstract

The study reported in this paper challenges current models of measuring second language fluency by comparing monologic versus dialogic task performance, and providing a novel insight into the measurement of the interactive aspects of dialogic performance. The data that constitute 35 monologic and dialogic task performances from second language learners were coded using a battery of established measures known to tap different aspects of fluency, and subjected to statistical analysis to test for overlaps or differences. Interactive aspects of fluency in dialogue, e.g. interruptions, overlap and unclaimed between turn pauses were also investigated to compare with common measures of monologic speech. While the results confirm previous research findings suggesting that performance is in general statistically more fluent in a dialogue in terms of speed, length of pause and repair measures, they indicate that performances in the two modes are not different in terms of number and location of pauses. The analysis of the dialogues indicates that the decisions researchers make about measuring the interactive aspects of fluency would have an impact on the outcome of measurements of fluency. These findings highlight the need for developing a more systematic and reliable approach to measuring second language (L2) fluency.

Key words: L2 fluency, monologue, dialogue

1 Introduction

The extensive work in task-based language teaching research investigating the development of second language (L2) ability in instructed settings suggests that complexity, accuracy and fluency are three principal dimensions that encapsulate language proficiency (see Housen and Kuiken 2009 for a full account). While this body of research recognizes fluency as being an essential component of communicative language ability, as well as an important performance descriptor and a key indicator of L2 development (de Jong et al. 2012; Housen and Kuiken 2009; Kahng 2014; Skehan 2014), a review of the literature in this area highlights three important limitations: a) there are gaps in our understanding of L2 fluency as a construct (Kahng 2014; Prefontaine 2013), b) the findings of fluency research often display mixed results due to the lack of a systematic approach to measuring fluency (Kormos 2006; Skehan 2014), and c) there are concerns about operationalizing and measuring fluency validly and reliably (Housen and Kuiken 2009; Housen et al. 2012). At a theoretical level, research in this area has led to new developments about defining the construct of fluency (Foster 2013; Segalowitz 2010) and offering a more in-depth understanding of how fluency operates in L1 and L2 (de Jong et al. 2012). However, it is important to note that L2 fluency has hardly been researched on its own, since most studies examine fluency as one of several aspects of L2 performance, e.g. accuracy and syntactic and lexical complexity, within a general construct of L2 proficiency (e.g. Kormos and Denes 2004, Skehan and Foster 1996). Acknowledging a separate agenda for researching the kind of discourse-level issues involved in interactional competence e.g. in

1

language classroom settings (Walsh 2013), it is noted that within the cognitive framework of researching L2 speech, L2 fluency has largely been investigated in monologic mode, with limited attention to measuring L2 fluency in task types that involve interaction between speakers. It is also possible to argue that current approaches to conceptualizing fluency are congruent with Levelt's (1989) widely accepted three-stage model of speech production, *Conceptualization*, *Formulation* and *Articulation*, in which language processing and production is defined on the basis of a monologic perspective to performance. Although recent research has shed light on a number of significant aspects of defining and measuring fluent monologic task performance (Kormos and Denes 2004; Segalowitz 2010), little systematic research has been done to discover the way fluent interaction and effective communication can be defined in interactive tasks, or what similarities and differences distinguish fluency of performance elicited by monologic versus dialogic tasks. This lack of understanding and the inconsistency of measurement limit the reliability of models we currently use to discuss core theoretical issues of speech planning and retrieval of linguistic knowledge for speech in real time. This study is therefore an attempt to provide a more in-depth understanding of the nature of fluent performance across the two modes.

2 Defining fluency

One of the earliest definitions of fluency frequently cited is Fillmore (1979) who defined fluency as "the ability to talk at length with few pauses; the ability to fill time with talk; the ability to talk in coherent and semantically dense sentences; the ability to have appropriate things to say in a wide range of contexts; and the ability to be creative and imaginative in the language use (Fillmore 1979: 51). Fillmore's definition, although proposed for L1 fluency, underlined the complex and multifaceted nature of fluency and highlighted the main factors and processes that make fluent speech possible. Further research in this area, e.g. Freed (2000) suggested that fluency was made up of different characteristics of speech ranging from its psychological manifestations, reflections on underlying speech-planning and thinking processes, to speech production, hesitation phenomena, and temporal dimensions of speech. Segalowitz (2000: 202) called for researchers to distinguish between cognitive aspects, i.e. "the efficiency of the operation of the cognitive mechanisms underlying performance" and performance aspects of fluency, i.e. "the observable speech, fluidity and accuracy of the original performance". In a more recent publication, Segalowitz (2010) proposed that L2 fluency comprises three distinct but inter-related concepts: Cognitive, utterance and perceived fluency. While cognitive fluency, in this framework, is concerned with mobilizing and integrating the underlying cognitive processes involved in language production (Segalowitz 2010: 48), utterance fluency refers to the measurable aspects of fluency such as speed, pausing and hesitation, and perceived fluency represents the inferences listeners make about someone's cognitive fluency based on their perceptions of how fluent the speaker is. From a research perspective, it is intriguing to see that definitions of fluency, e.g. Fillmore's, Freed's and Segalowitz's, have conceptualized the construct of fluency as a characteristic of monologic speech without considering or discussing its representation and operation during speech when two or more speakers interact with one another. Therefore a key question the current study seeks to answer is whether the same aspects of fluency unequivocally characterize fluent speech in monologues and dialogues.

2.1 Studying fluency in monologic and dialogic task performance

Monologue and dialogue are two frequently used modes of oral language in both real life and pedagogic contexts. Although it has been argued that dialogues, given their interactive nature, represent language more authentically and naturally (Guillot 1999; Van Lier 2004), research in SLA has predominantly focused on measuring monologic performance elicited by tasks such as oral narratives (Skehan and Foster 1996; Tavakoli, 2011), short talk (de Jong and Perfetti 2011), and answering-machine message leaving tasks (Mehnert 1998). Frequent use of monologues in L2 fluency research can be attributed to a number of factors including a) the degree of control associated with a monologic task performance (i.e. simpler pragmatic demands for speech planning), b) predictability of the outcome of the performance, and c) clarity and ease of the procedures for measuring language produced in a monologic task. On the other hand, measuring fluency in dialogic tasks can prove difficult not only because of the complex pragmatics involved in dialogue, leading to a less controlled and less predictable nature of performance in this mode, but more importantly because of the difficulty associated with measuring the interactive aspects of dialogues, e.g. overlap, unclaimed between-turn pauses, and the interdependence of the interlocutors' performances.

A limited number of studies have used dialogic tasks to investigate L2 fluency, and only few have examined the differences in the same speakers' fluency when they perform a monologue and a dialogue. Michel (2011), in a between-participant design, examining the effects of task complexity, i.e. "the level of challenge that a task is likely to contain" (Skehan 1998: 134), and interaction on L2 performance, used the same decision making task in two modes: a) a monologic answering machine message leaving task, and b) a dialogic telephone conversation task. The results of her study indicated that the dialogic mode elicited language of higher fluency in terms of speed, pausing and repair measures. The differences between the two modes of performance for all measures of fluency in Michel's study (2011) reached statistically significant levels, with noticeable effect sizes observed for repair and pausing behaviour. The surprisingly non-significant results from the effects of task complexity on fluency can perhaps be explained in terms of how task complexity was operationalized in this study, i.e. since both tasks involved a comparable degree of reasoning and required a decision to be made they did not vary adequately in terms of complexity. With regards to the effects of mode, Michel (2011) argues that speakers may find dialogues cognitively less demanding to perform, not because they find interactive dialogue easier to engage with fluently in terms of pragmatic or task complexity, but because speakers can use the interlocutor's turn to plan their own subsequent performance.

In a study focusing on the development and measurement of fluency in monologue and dialogue, Witton-Davies (2014) used a picture story retelling monologue and a discussion dialogue to investigate the development of fluency in L2 speakers over a four-year period. His findings supported previous research and confirmed that performance in dialogues was consistently more fluent than that in monologues, with higher speech rates, less pausing and fewer repair words being the key characteristics of dialogic performance. Neither of the above studies explored whether the choice and operationalizations of fluency measures in a dialogue had an impact on the results; this is what the current study aims to shed light on.

2.2 Characteristics of a dialogue

While a monologue involves production of sequences by one speaker, a dialogue is "prototypically a joint enterprise involving more than one person" (Cameron 2001: 87), with the speakers taking turns to talk. Edwards (2008) reports that categories that make a dialogue different from a monologue include between-turn pauses, interruptions by the second speaker, and simultaneous talk. Turn-taking seems to be of central importance in a dialogue (Cameron 2001; Edwards 2008) since for a conversation to work in an ideal manner at any single moment one speaker's talk is followed by a short silence before the next speaker takes the turn to speak. However, real life dialogues are normally far from ideal in terms of the turn taking principles. In the case of simultaneous talk, i.e. overlap, one speaker will normally win the floor and therefore the other becomes silent (Cameron 2001). And after a period of silence, normally one of the speakers breaks the silence. Turn taking is not planned in advance in a normal dialogue; rather, it develops when the speakers engage in the conversation (Cameron 2001; Wilson and Zimmerman 1986). Sacks et al. (1974) proposed a 3-step procedure for turn taking in English in which a) current speaker chooses the next speaker; b) next speaker(s) self select themselves; and c) current speaker may continue with their turn after the silence. Although it seems simple and straight forward, the structured procedure may not be observed in everyday conversations all the time. Highlighting turn taking as one of the most salient features of social interaction, Wilson and Zimmerman (1986), among others, argued that it should not be viewed as a simple exchange of stimulus and response. Rather, turn taking is fundamentally a collaborative activity that develops in a less structured manner. Research in conversation analysis has shown that a key mechanism in the organization of turns in a conversation is the ability to anticipate the moment of completion of a current speaker's turn, what is known as *projection* (Lerner 2003; Schegloff 2000, Schegloff 2001). Whereas previous research (e.g. Caspers 2003) emphasised the role of intonation in projection, De Ruiter et al. (2006) argue that knowledge of lexicosyntactic content, i.e. lexical and syntactic characteristics, of an utterance is necessary and perhaps sufficient for both predicting projection and regulating conversational turn taking. Such findings imply that second language learners whose L2 knowledge of lexicosyntax is not yet adequately developed may find it difficult to anticipate projection. It is necessary to note that turn-taking and projection are reported to be culturally shaped and determined, and therefore it may vary across different discourse communities (Cameron 2001; De Ruiter 2006; Edwards 2008). The data analysis (Section 4.3) will examine turn taking patterns, interruptions and overlap speaking time in the dialogue data.

2.3 Measuring fluency

In an attempt to create a more systematic approach to measuring fluency, Skehan (2003), and Tavakoli and Skehan (2005) suggested that fluency should be measured with regard to its three main characteristics: a) speed fluency, i.e. speed with which speech is performed, b) breakdown fluency, the pauses and silences that break down the flow of speech, and c) repair fluency, hesitations, repetitions and reformulations that are used to repair speech during the production process. Following from this, Skehan (2014) suggests that when measuring fluency composite measures that blend *speed* and *flow* of speech, e.g. phonation time and length of run should also be considered.

Recent research findings suggest that some measures of fluency are internally related and, if not chosen carefully, one measure may overlap with others (Kormos 2006; Skehan 2014; Tavakoli and Skehan 2005). Identifying the best measures of fluency that can reliably encapsulate L2 utterance fluency and minimize the possible overlap between different measures, Witton-Davies (2014) and Mora and Valls-Ferrer (2012) suggest that pause length, pause frequency, pause location, mean length of run, speech and articulation rates, phonation time ratio, and a selection of repair measures are the most reliable measures of utterance fluency. Prefontaine (2013) reports that mean length of run and average pause time are two measures of utterance fluency that most strongly relate to self-perceptions of fluency. Kahng (2014: 810) reports that speech rate and mean length of run are strongly associated with both L2 oral proficiency and perceived fluency, whereas articulation rate and repair measures are not. It is beyond the scope of this article to reflect on how these studies illuminate all aspects of the speech planning process as well as performance in real time, but we can pick out that certain measures are crucial in understanding what is going on when speakers are engaging in dialogue and how we can reliably measure their speech fluency. In order to fulfil the aims of this study, most relevant of the fluency measures identified from previous research and some novel measures were selected:

Speed
- Articulation rate: mean number of syllables per minute divided by mean amount of phonation time (excluding pauses)
- Speech rate: mean number of syllables per minute divided by total time (including pauses)

Break down
- Mean length of pauses per 60 seconds
- Mean number of pauses per 60 seconds (clause-internal versus clause-external)

Repair
- Repair measures: mean number of partial or complete repetitions, hesitations, false starts and reformulations
- Mean number of filled pauses, e.g. em and er

Composite
- Mean length of run: the mean number of syllables between two pauses[1]
- Phonation time ratio: time taken to perform the task (excluding pauses)

Dialogue only measures (not previously investigated)
- Number of turns and number of interruptions

A key contribution of this study is using dialogue-only measures which will not only provide an insight into aspects of dialogic performance, but allow for a comparison of fluency across both modes.

3 Research questions:

The following research questions guided the study.

1. Does L2 speakers' fluency remain the same in monologic and dialogic task performance? If not, what are the differences between the two modes of performance?

---

[1] It should be noted that following from de Jong et al., (2012) a pause is an unfilled silence of longer than 0.25 a second.

2. Which aspects of fluency in dialogic task performance are affected by the way measurement of fluency, in terms of turns, pauses and overlaps, is operationalized?

4 Methodology

4.1 Participants, procedures and tasks

The participants were 35 EAP students enrolled on a pre-sessional course at a university in the UK. They were at B2 level (CEFR) and were placed on their course based on their IELTS Score (5.0 or 5.5). They were aged between 22 and 35, and had a range of diverse L1s including Arabic, Chinese, Kurdish, Russian and Thai. The participants had been on their EAP courses for four weeks when the experiment took place. For the purpose of the study, they performed a monologue and a dialogue in one of their speaking classes. The choice of the tasks was guided by three main criteria: a) the tasks were in line with the course objectives, i.e. improving the learners' speaking and listening ability, b) the instructors considered the task as interesting, relevant and at the right level, and c) the task types were familiar but the topics had not been covered in the course before so that any possible practice effect can be avoided. The monologue was a retelling of a recent personal shopping experience, for which they had 1 minute to plan and 1 minute to perform the task. The dialogue involved a discussion task that required the participants to present the case for or against a particular topic e.g. which is better: watching a movie at home or in the cinema. They had 1 minute to plan and 3 minutes to perform the task. The tasks, planning time and time on task were piloted with a different group of learners before the experiment took place[2].

4.2 Data analysis

Task performances were digitally recorded, transcribed and coded for a range of fluency measures (Section 2.3 above). While many of the participants spoke longer than one minute for each task, for the purpose of the analysis, all temporal measures were based on the first 60 seconds of their performance beginning when they actually started speaking. PRAAT (Boersma and Weenink 2013) software was used to measure temporal aspects of fluency e.g. phonation time, length of pause and articulation rate. For PRAAT measurement, intra-rater reliability was used for a 10% sample of the data and coefficient measures of above 95% were achieved. For the rest of the measures, e.g. the number of filled pauses and repairs, coding was done manually with a researcher reading through the text and coding the transcripts. To ensure reliability of the coding of these measures, initially a 10% sample of the data were coded by a second researcher. This was repeated for a second and sometimes a third time until a 90% inter-rater reliability was obtained.

---

[2] It is important to note that in a within-participant design, it is very difficult to use the same topic in different modes without a practice effect. As such, task topic and task mode are typically confounded in research of this type, and therefore, the results obtained may at least be partly attributed to the interaction between topic and mode.

4.3 Measuring fluency in a dialogue

As noted above, analyses of number of turns and number of interruptions were carried out on the dialogue data. Careful examination of the data suggests that in contrast to a monologue, in a dialogue the speakers' fluency depends at least to some extent on the interlocutor's conversational skills in using English in an international setting, e.g. turn-taking and, their willingness to communicate (Cameron 2001). Therefore, factors such as how dominant, passive, or involved (Edwards 2008, Tannen 1994) the interlocutors were, the number of turns taken by each speaker, the interruptions and overlap in speaking time, and the unclaimed between-turn pauses affected different aspects of fluency in the dialogues. For the purpose of measuring fluency in the dialogues, a number of steps were taken. First, to ensure that tasks were performed interactively, any data from interlocutors who had a very unbalanced dialogue, with one speaker dominating for a long period (e.g. more than 70% of the time) or the other remaining quiet for an extended period were excluded. One minute of each speaker's performance was used for the analysis of temporal aspects, i.e. the analysis of a dialogue included two 60 seconds performances. For features such as number of turns, interruptions and repair measures the whole performance was examined. The data were checked to make sure each participant had at least two turns in the 60 seconds of their performance in a dialogue. To interpret silence, particularly in the between-turn pauses, the context in which it had occurred was also considered, e.g. whether the speaker paused to look for a lexical item or to signal the end of their turn (Kurzon 2013).

In terms of interruptions, overall there were not many interruptions in the data. Most participants did not interrupt their interlocutor frequently (mean=4); only three participants interrupted their interlocutors more frequently (mean=13). A qualitative examination of the data suggested that interruptions were mainly a sign of 'high involvement' (Tannen, 1994) rather than dominance in the dialogue. A large number of these interruptions (70%) were non-lexical filled pauses (mhm) or short back channels like, 'yeah'. The rest of the interruptions (30%) were either lexical phrases like 'you're right,' or the beginning of a new turn such as ' but when you …'. All the lexical interruptions or short back channels were considered as overlapping speaking time which was considered to belong to both participants, and as such were included in the measurement of fluency for both speakers.

Kurzon (2013) argues that there are three types of conversational silence in informal situations: a) a short silence or pause where the speaker does not respond immediately, b) a voluntary silence where one of the speakers is asked a question but s/he intentionally keeps quiet, and c) a silence in a multiparty conversation where one of the participants chooses to remain silent while others are conversing. A qualitative examination of the data suggested that the unclaimed between-turn pauses did not belong to any of Kurzon's (2013) categories. These were typically long pauses after one speaker came to the end of their utterance and remained silent, while the other speaker also kept quiet because they were either uncertain if the first speaker's turn was complete, or not prepared to speak. Given that it was difficult to attribute such pauses to individual speakers, one way to deal with the unclaimed between-turn pauses was to exclude them from the analysis. A

second option, deemed suitable for this study, was to divide the pauses equally between the two speakers[3]. The data analysis below includes data from both ways of measuring these pauses.

5 Results

In order to see whether there were significant differences between fluency measures in monologic and dialogic task performance, a number of t-tests were run to compare the participants' fluency elicited by the monologue and dialogue tasks, and Cohen's (1998) definitions of effect size used. Table 1 shows the results of the t-tests for these comparisons. The measures shown here, as noted in the Methodology Section above, are identified by the relevant literature as the core measures tapping into utterance fluency.

Table 1: Results of t-tests comparing fluency measures in monologic and dialogic performance (between-turn pauses excluded)

Temporal Measures	Monologue mean (SD)	Dialogue mean (SD)	*T*	*P*	*Cohen d*
Articulation rate	192.65 (23.96)	225.08 (26.77)	6.81	.001*	1.27
Speech rate	143.25 (23.33)	175.23 (22.16)	8.40	.001*	1.37
Mean length of pause (in seconds)	.67 (.15)	.51 (.08)	6.35	.001*	1.33
Number of pauses clause-internal	13.85 (5.27)	13.46 (6.68)	.29	.773	.06
Number of pauses clause-external	8.86 (3.26)	8.37 (3.47)	.73	.471	.15
Repairs	7.34 (3.60)	5.48 (4.11)	2.19	.04	.48
Filled pauses	12.80 (4.80)	15.88 (8.82)	2.01	.05	.43
Mean length of run	6.62 (2.61)	7.95 (2.74)	2.99	.005*	.50
Phonation time ratio	74.48 (6.93)	80.37 (8.17)	3.82	.001*	.78

All temporal measures are per 60 seconds of each speaker's talk.

The results indicate that there were significant differences between many of the fluency measures in the two modes of task performance favouring dialogic speech with the participants producing longer runs (*t*=2.99; *p*= .005; *d*=.50), shorter pauses (*t*=6.35; *p*= .001; *d*=1.33), higher phonation time ratios (*t*=3.82; *p*= .001; *d*=.78), and faster articulation rate (*t*=6.81; *p*= .001; *d*=1.27) and speech rate (*t*=8.40; *p*= .001; *d*=1.37) in dialogic task performance. All these significantly different measures showed medium to large effect sizes (Cohen 1998). The participants also produced more filled pauses (*t*=2.01; *p*= .05; *d*=.43) and fewer repair measures (*t*=2.19; *p*= .04; *d*= .48) in the dialogic task with noticeable effect sizes (Cohen, 1998). However, these two *p* values should be interpreted with care because when a Bonferroni adjusted alpha level is considered a significance level of *p*< .005 is set, suggesting *p* values of .04 and .05 should not be taken as statistically significant in this context. As for number of pauses, although there were more pauses both in the middle of clauses and at clause boundaries in monologic task performance, the differences between the two modes were

---

[3] One may argue that a third option was to use the principles of Conversation Analysis to analyse the data to determine who the unclaimed pauses could be attributed to. However, given the lack of a shared L1 among the participants and the possibility of different conversational norms and practices among them, this option was ruled out.

generally small and negligible. Apart from pausing, the results clearly support existing findings that speakers are more fluent in dialogue than in monologue, using standard measures.

As discussed above, there are a number of measurement decisions that can affect the results of fluency measures in dialogic task performance. The first important aspect that can affect measurement of fluency is the way the between-turn pauses are operationalized, i.e. a) whether the pauses are included in the measurement of different aspects of fluency, and b) if so, who is responsible for the pauses between two speakers' turns. This is an aspect of fluency measurement that has not been discussed in fluency studies before. In the analysis presented in Table 1 above, the between-turn pauses were excluded from the measurement. This is to say, all the silent pauses made between the different turns were not attributed to either of the two speakers engaged in the dialogue.

The current study was concerned that the significant differences observed in different fluency measures between the two modes could (at least partly) be explained by excluding the between-turn pauses. As such, alternative analyses were needed in which a) the pauses were included in the new measurement, and b) the pauses were equally divided between the two speakers. Consequently, a second set of measurements was used to calculate all temporal aspects of fluency with the pauses divided between the two speakers and included in the measurement. For example, if there was a two-second unclaimed pause between two speakers' turns, each speaker was credited for one second of the pause. It is important to note that when these pauses are included in the measurement, mean length of run and articulation rate will not be affected as they do not involve pause duration or pause frequency. Given that the between-turn pauses happen at clause boundary, the number of clause-internal pauses will not change either. Hence, these measures are excluded from the next set of t-tests. Table 2 demonstrates the differences between the temporal measures of fluency where pauses were excluded from the analysis as shown in Table 1 above compared with the new measures in which they are included in the analysis and divided between the speakers.

Table 2: Results of t-tests comparing fluency measures in dialogic performance with the between-turn pauses included and excluded

Temporal Measures	Dialogue measures when pauses excluded (as in Table 1 above) mean (SD)	Dialogue measures when pauses included and divided between speakers mean (SD)	T	P	Cohen d
Speech rate	175.23 (22.16)	170.62 (22.16)	5.45	.001*	.21
Mean length of pause (in second)	.51 (.08)	.54 (.10)	3.47	.001*	.33
Mean number of pauses	21.83 (8.19)	23.54 (7.93)	6.48	.001*	.21
Number of pauses clause-external	8.37 (3.47)	10.09 (3.72)	6.48	.001*	.48
Phonation time ratio	80.37 (8.17)	76.65 (7.72)	5.56	.001*	.47

All temporal measures are per 60 seconds of each speaker's talk.

The results indicate that when the between-turn pauses are included in the analyses and divided between the two speakers, performance in dialogic tasks, previously seen as more fluent on most measures than monologic performance, became significantly poorer, or less fluent, compared to

dialogic scores in the previous analysis – i.e. when unclaimed pauses between the two speakers' turns were excluded. This implies that the way fluency measures are defined and operationalised has had significant effects on the results obtained from fluency measures. This may mean that the main differences achieved between monologic and dialogic task performance (Table 1 above) might have in effect been caused by the way they are measured. Therefore, a new set of t-tests were run to find out whether with this new measurement, i.e. including the between-turn pauses in the analysis, the significant differences between fluency in monologic and dialogic modes still persisted.

Table 3: Results of t-tests comparing fluency measures in monologic and dialogic performance (between-turn pauses divided between the speakers and included in the analysis)

Temporal Measures	Monologue mean (SD)	Dialogue mean (SD)	T	P	Cohen d
Articulation rate	192.65 (23.96)	225.08 (26.77)	6.81	.001*	1.28
Speech rate	143.25 (23.33)	170.62 (22.16)	7.61	.001*	1.20
Mean length of pause (in second)	.67 (.15)	.54 (.10)	4.97	.001*	1.02
Number of pauses clause-internal	13.85 (5.27)	13.46 (6.68)	.29	.773	.06
Number of pauses clause-external	8.86 (3.26)	10.09 (3.72)	1.91	.064	.35
Mean length of run	6.62 (2.61)	7.95 (2.74)	2.99	.005*	.50
Phonation time ratio	74.48 (6.93)	76.65 (7.72)	1.60	.117	.30

All temporal measures are per 60 seconds of each speaker's talk.

The results demonstrate that even when the between-turn pauses are included in the analyses, there are still significant differences between monologic and dialogic task performance with most temporal measures demonstrating more fluent performance in a dialogue. Phonation time ratio is the only temporal measure in which the difference does not reach a statistically significant level with the new analysis (t=1.60; p= .117; d= .30). Although the differences between the two modes are now larger for the number of clause-external pauses, they are not statistically different in monologic and dialogic task performance (t=1.91; p= .06; d= .35). These results suggest that participants' speech in dialogues was more fluent than that in monologues for some, but not all, fluency measures.

6 Discussion

The study confirms findings of previous research (Michel 2011; Witton-Davies 2014) that performance in dialogic tasks is more fluent than that in monologic tasks in terms of speed, length of pause and repair measures. The analysis indicates that composite measures, e.g. mean length of run and speech rate are also higher in dialogic task performance. With regard to break-down aspect of fluency, while mean length of pause is statistically shorter in the dialogues, there is little difference between performance in the monologues and dialogues in terms of number and location of pauses. In line with the findings of previous research (Tavakoli 2011), the participants in this study paused more frequently in the middle of clauses consistently across the two modes. These findings suggest that while being engaged in a dialogue encourages speed and shortens length of pause, it has little impact on how often and where L2 speakers pause. This finding suggests that speakers use pauses to monitor their speech production process (de Jong et al. 2013; Kormos 2006; Michel 2011), and to pay attention to form in terms of accuracy of their performance (Tavakoli et al. in print).

It is possible to argue that the collaborative and interactive nature of a dialogue makes it more convenient for the L2 speakers to produce a more fluent performance. Previous research reports that the collaborative nature of a dialogue allows interlocutors to use their partner's turn to plan for their utterances (Lochbaum 1998; Webber 2008). This may mean that having "listening time" helps speakers with what Levelt (1989) considers the *conceptualisation phase* (where the preverbal message is generated) and *reformulation phase* (where the preverbal message is converted into a phonetic plan for speech) of speech production. But beyond this individual aspect of improving performance, the interactive nature of having a partner in dialogue may genuinely encourage speakers to show greater willingness to communicate interactively, and take the interlocutor's needs into account by producing fewer hesitations and repetitions and faster speech. The frequent use of filled pauses in the dialogues, compared to those in monologues, is an interesting example of taking the partner's needs into account. More qualitative research, e.g. stimulated recall protocols, is needed to provide a better understanding of factors that encourage a more fluent speech in an interactive task such as a dialogue versus a monologic performance such as retelling of a story.

Another important finding of the study is that the decisions researchers make about the measurement of fluency in a dialogue may affect the different temporal aspects of L2 fluency. For instance, the data analysis on Table 2 demonstrates that the way between-turn pauses are defined and operationalised affected measures of phonation time ratio, speech rate, mean length of pause and number of pauses. This may mean that the reliability of such measures across the two modes depends on the way they are operationalised. Whereas previous research (Tavakoli et al. in print; Mora and Valls-Ferrer 2012; Witton-Davies 2014) suggests that pause length, pause location, mean length of run, speech and articulation rates, phonation time ratio, and repair measures are the most reliable measures of fluency in a monologue, this study indicates that the same measures may not be the most reliable representatives of fluency in a dialogue since they may be affected by the decisions about overlap speaking time and between-turn pauses. Further research is needed to investigate reliability of the measures SLA researchers employ when examining fluency in a dialogic mode.

Finally, in respect to the cognitive construct of fluency itself, de Jong et al. (2013) reported that while all measures of utterance fluency were at least to some extent related to cognitive fluency, measures such as mean length of pause were only marginally linked to cognitive fluency implying that pausing can be better explained by other factors, e.g. individual differences (de Jong et al. 2013). The findings of the current study about the differences between speakers' speaking patterns in a dialogue suggest that further research is also needed to examine the extent to which task mode and speaker's conversational skills can explain cognitive fluency.

7 Conclusion

Measuring fluency in monologic and dialogic L2 performance seems to raise both theoretical and methodological issues. At a theoretical level, the findings of the current study raise the question of whether fluency represents the same construct in a monologue compared with a dialogue. In a monologue, where the sole responsibility of speech is on one speaker, fluency is predominantly characterised by the flow of speech, its speed and a lack of dysfluency measures. In a dialogue where speakers rely on one another to produce connected speech, however, fluency is more than

flow and speed, as the interlocutor's conversational skills and taking care of their needs gain significance. For this reason, the speakers' attention is not focused on flow and speed only, but it is distributed to other factors such as producing fewer repairs and more filled pauses to provide a smoother interaction with the interlocutor. This may often be demonstrated through turn-taking, overlap, negotiating meaning and other characteristics of interactive discourse. The findings of this study encourage researchers to take the important aspects of interactive speech into account not only in defining and conceptualising fluency but in modelling speech production. At a methodological level, using the same measures may not be the most reliable and effective way of measuring fluency in both modes. If careful decisions are not made about operationalisation of between-turn pauses, any measure that includes those pauses may be challenged for their consistency and reliability.

References

Boersma, Paul & David Weenink. 2013. *Praat: doing phonetics by computer.* Version 5.3.51, retrieved 2 June 2013 from http://www.praat.org/

Cameron, Debora. 2001. *Working with spoken discourse*. London: Sage.

Caspers, Johanneke. 2003. Local speech melody as a limiting factor in the turn-taking system in Dutch. *Journal of Phonetics* 31(2).251–76.

de Jong, Nel & Charles Perfetti. 2011. Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning* 61(2). 533-568.

de Jong, Nivja, Margarita Steinel, Arjen Florijn, Rob Schoonen & Jan Hulstijn. 2012. facets of speaking proficiency. *Studies in Second Language Acquisition 34*(1). 5–34.

de Jong, Nivja, Margarita Steinel, Arjen Florijn, Rob Schoonen & Jan Hulstijn. 2013. Linguistic skills and speaking fluency in a second language. *Applied Linguistics* 34(4). 893–916.

De Ruiter, Jan, Holger Mitterer & Nick Enfield. 2006. Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language* 82(3), 515–535.

Edwards, Jane. 2008. The transcription of discourse. In Deborah Schiffrin, Deborah Tannen & Heidi Hamilton (eds.), *The handbook of Discourse Analysis*, 321-348. London: Routledge.

Fillmore, Charles. 1979. On Fluency. In Charles Fillmore, Daniel Kempler & William Wang (eds.), *Individual differences in language ability and language behavior,* 85–102. New York: Academic Press.

Foster, Pauline. 2013. Fluency. In Carol Chapelle (ed.),*The encyclopaedia of Applied Linguistics*. London: Wiley-Blackwell.

Freed, Barbara. 2000. Is fluency, like beauty, in the eyes (and ears) of the beholder? In Heidi Riggenbach (ed.), *Perspectives on Fluency,* 243–265. Ann Arbor: University of Michigan Press.

Guillot, Maureen. 1999. *Fluency and its teaching*. Clevedon: Multilingual Matters.

Housen, Alex & Fulkert Kuiken. 2009. Complexity, accuracy and fluency in second language acquisition. *Applied Linguistics* 30(4). 461–473.

Housen, Alex, Fulkert Kuiken & Ineke Vedder. 2012. *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*. Amsterdam: John Benjamins

Hymes, Dell. 1972. On Communicative Competence. In John Pride and Janet Holmes (eds.), *Sociolinguistics, Selected Readings,* 269-293. Harmondsworth: Penguin.

Khang, Jimin. 2014. Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning* 64(4): 809-854.

Kormos, Judit, 2006. *Speech production and second language acquisition*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Kormos, Judit & Marian Dénes. 2004. Exploring measures and perceptions of fluency in the speech of second language learners. *System 32*(2). 145–164.

Kurzon, Dennis. 2013. Analysis of silence in interaction. In Carol Chapelle (ed.),*The encyclopaedia of Applied Linguistics*. London: Wiley-Blackwell.

Levelt, Willem. 1989. *Speaking: from intention to articulation*. Cambridge: Cambridge University Press.

Lerner, Gene. 2003. Selecting next speaker: The context-sensitive operation of a context-free organization. *Language in Society* 32(2). 177-201.

Lochbaum, Karen. 1998. A collaborative planning modal of intentional structure. *Computational Linguistics* 24(4): 525-72.

Mehnert, Uta. 1998. The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition 20*(1). 83–108.

Mora, Joan & Margalida Valls-Ferrer. 2012. Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly 46*(4). 610-41.

Prefontaine, Yvonne. 2013. Perceptions of French fluency in second language speech production. *Canadian Modern Language Review* 69(3), 324-348.

Sacks, Harvey, Emmanuel Schegloff & Gail Jefferson. 1974. A simplest systematics for the organisation of turn taking for conversations. *Language* 50(4). 696-735.

Schegloff, Emanuel. 2000. Overlapping talk and the organization of turn-taking for conversation. *Language in Society* 29(1).1–63.

Schegloff, Emanuel. 2001. Accounts of conduct in interaction: Interruption, overlap, and turn-taking. In Jonathan Turner (ed.), *Handbook of sociological theory,* 287–321. New York: Kluwer Academic.

Segalowitz, Norman. 2000. Automaticity and attentional skill in fluent performance. In Heidi Riggenbach (ed.), *Perspectives on Fluency*, 200–219. Ann Arbor: University of Michigan Press.

Segalowitz, Norman. 2010. *The cognitive bases of second language fluency*. New York: Routledge.

Skehan, Peter. 1998. *A cognitive approach to language learning*. Oxford: Oxford University Press.

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

Skehan, Peter. 2009. Modelling second language performance: Integrating complexity, accuracy, fluency and lexis. *Applied Linguistics* 30(4). 501-532

Skehan, Peter. 2014. Limited attentional capacity, second language performance, and task-based pedagogy. In Peter Skehan (ed.), *Processing Perspectives on Task Performance,* 211–260. Amsterdam: John Benjamins.

Tannen, Deborah. 1994. *Gender and discourse*. Oxford: Oxford University Press.

Tavakoli, Parvaneh, Colin Campbell & Joan McCormack. In print. Development of speech fluency over a short period of time: Effects of pedagogic intervention. *TESOL Quarterly*.

Tavakoli, Parvaneh. 2011. Pausing patterns: Differences between L2 learners and native speakers. *ELT Journal* 65(1): 71-9.

Tavakoli, Parvaneh & Peter Skehan. 2005. Strategic planning, task structure and performance testing. In Rod Ellis (ed.), *Planning and task performance in a second language*, 239-277. Amsterdam: John Benjamins.

Van Lier, Leo. 2004. *The ecology and semiotics of language learning: A sociocultural perspective*. Boston: Kluwer Academic.

Webber, Bonnie. 2008. Computational perspectives on discourse and dialogue. In Deborah Schiffrin, Deborah Tannen & Heidi Hamilton (eds.), *The Handbook of Discourse Analysis*, 798-817. London: Routledge.

Witton-Davies, Giles. 2014. *The study of fluency and its development in monologue and dialogue*. Unpublished Doctoral Thesis, University of Lancaster.

De Jong, Nivja & Ton Wempe. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior research methods*, 41 (2). 385 - 390.