The Variability of Fluency in Dialogue and Monologue

Giles Witton-Davies (朱杰佑)
National Taiwan University
giles@ntu.edu.tw

Since fluency is most easily measured when there is only one speaker, fluency research is usually based on monologue data. However, everyday speech usually takes the form of dialogue, where fluency is less easy to measure objectively, as the contributions of each participant have to be separated before any measurements can be made. When it comes to oral assessment, face-to-face language testing usually involves dialogue, while indirect or semi-direct speaking tests usually elicit a series of monologues, which are recorded and assessed later. This study looks at how fluency varies in these two kinds of speech - monologue and dialogue. A group of Taiwan university students were recorded performing two oral tasks: an individual picture story, and a discussion task carried out in pairs. The same tasks were performed at two times – in their first and final semesters at university (in year 1 and at the end of year 8). Transcripts were produced of the recordings and Praat Analysis was carried out, allowing for temporal measurements to be made of such variables as speech rate, repair, mean length of run, and pausing within and between clauses. It was found that speech rates were significantly higher for the dialogue discussion than for the monologue picture stories. Speech rates also improved more for the discussion task than for the story task over the four year period. The implications of this finding for classroom teaching, testing and research into fluency are discussed.

INTRODUCTION

The Study of Fluency

What is fluency in a foreign language? Various definitions have been offered, although there is still no universally accepted answer to the question. Segalowitz (2010) makes a three-way distinction between different possible foci for discussing fluency – (a) utterance fluency, which is often seen as being related to temporal measures of speech rate and hesitations; (b) perceived fluency, which can be assessed though listener judgements of a speaker's utterance fluency; and (c) cognitive fluency, which is hypothesised as the processing speed underlying (a) and (b).

In studies of fluency, the focus has usually been on utterance fluency, and the search for temporal measures that will serve as reliable indicators of speech that is judged as fluent (Dechert & Raupach, 1980). It is even hoped that such research will enable automatic measurements of fluency (de Jong & Wempe, 2009; Ginther, Dimova, & Yang, 2010). However, others have emphasised fluency as a perceptual phenomenon (Freed, 2000; Guillot, 1999). Lennon uses the term perceived fluency, and defines it as “the impression the listener has that the psycholinguistic processes of speech planning and speech production are functioning more or less easily and effortlessly.” In some ways, perceived fluency is simply a judgement of a speaker's utterance fluency. But focusing on the listener encourages us to take into account more than the basic temporal measures of speed and hesitation, and to appreciate the importance of context. As Fillmore (1979) points out, we do not expect the same kind of fluency from a disc-jockey and an academic; we would not criticise the academic for not speaking as fast as the disc-jockey, nor criticise the disc-jockey for being more repetitive or reformulating more than the academic.

Few researchers have discussed cognitive fluency, or the underlying psycholinguistic processes involved in fluency. One of the first to do this was Schmidt (1992), who used the declarative / procedural knowledge distinction to explain how speech becomes more fluent as skills are proceduralised. Towell et al (1996) used this model in their study of fluency development, and sought to distinguish the mere speeding up of processes from changes to the
nature of the processing that was taking place. Segalowitz has focused specifically on measurements of cognitive fluency and how it develops. He has taken particular interest in how simple speeding up of processing through practice can be distinguished from deeper changes in the structure and organisation of processing, which may lead not only to faster processing but also to greater processing stability (Segalowitz, 2000, 2010). If the increase in processing stability is greater than the increase in processing speed, he argues, we can say that cognitive fluency has improved.

The study reported here uses temporal measures of utterance fluency to compare monologue with dialogue, and to see how fluency develops on different tasks over a four year period. The design of the study is longitudinal, with data collected from the same speakers at different times. Before describing this research, I will survey the major cross-sectional and longitudinal studies of fluency, to show what is already known about oral fluency in a foreign language.

Cross-sectional and Longitudinal Studies of Fluency

Cross-sectional studies of fluency usually involve comparisons between independently obtained measures of oral fluency or proficiency, on the one hand, and temporal measures of fluency on the other. Independent measures may come from oral tests (OPI, OEPT, IELTS, TOEFL), from rater judgements, from class grouping, or by comparing learner's L1 and L2 performance.

Riggenbach (1991), examines how high level and low level learners differ, identifying speech rate, silent pauses (especially when not at grammatical boundaries) and “clusters of disfluencies” as key areas that distinguish the two groups. Riazantseva (2001) makes a 3-way comparison between L1 Russian speakers, L2 speakers of English (who are also the Russian speakers), and a second group of L1 English speakers. This allows for comparisons of pausing patterns between groups speaking two different L1s, between L1 and L2 speakers of English, and between the same people speaking their L1 and L2. She finds that L1 Russian features longer pauses than L1 English, but that while low level Russian L2 speakers of English make much longer pauses than L1 English speakers, these pauses become significantly shorter at higher levels. Kormos & Dénes (2004) seek to identify the temporal measures that best predict the fluency ratings of independent judges. They find that speech rate, mean length of run and “pace” (the number of stressed syllables per minute) are the best predictors of fluency ratings.

Iwashita et al (2008) consider the relationship between holistic oral test scores from raters and a variety of measures of grammar, vocabulary, pronunciation, and fluency. With respect to the fluency measures, they find the ones that correlate most strongly with rater scores are speech rate, and the time and frequency of silent pauses. In this article, “speech rate” is in fact pruned speech rate, as all repair and filled pauses were removed before analysis. In a similar study, Ginther et al. (2010) compare oral proficiency test (OEPT) scores with temporal measurements made using Praat software, together with measures of lexis, grammar and pronunciation. Speech rate, articulation rate and MLR (mean syllables per run) are found to correlate most closely with test scores, while silent pauses (but not filled pauses) also correlate significantly. The numbers of participants in these studies vary considerably: 6 in Riggenbach (1991); 30 speakers of Russian and 20 of English in Riazantseva (2001); 16 in Kormos and Denes (2004); 200 for Iwashita et al (2008); and 150 for Ginther et al. (2010).

Longitudinal studies can show fluency development over time, often by comparing temporal fluency measures with rater judgements. Longitudinal studies focus on fluency development but they also throw light on the nature and measurement of fluency itself. An early example of a study of fluency development is Lennon (1990). The findings are of limited reliability because of the small sample size, but Lennon found that all four of the German speakers of L2 English he studied improved on pruned speech rate, number of filled pauses per unit, and percentage of C-units followed by a silent pause. Freed (1995) studied American
learners of L2 French, some of whom spent a semester abroad, while others stayed at home. The abroad group was, against expectations, not found to show significant differences from the “at-home” group, except for in the measure of speech rate. It was, however, found that lower level learners did make greater overall improvement than the higher level ones. Towell et al. (Towell et al., 1996) studied a group of 12 British students of L2 French on a four year course at a university in the UK. The key measures on which improvements were found after a year abroad were mean length of run and speech rate.

A series of more recent longitudinal studies have come from Derwing et al. (e.g. Derwing et al. 2007, 2009). These studies focus on second language learning of English among immigrants to Canada, representing different L1 language backgrounds. The 2007 paper found that a Slavic-language speaking group made progress in fluency and comprehensibility over a two year period, while a Mandarin-speaking group made little progress over the same time. The 2009 paper found that L1 and L2 fluency correlations were strong at low levels of L2 proficiency, but that they weakened with L2 development. The fact that L1 fluency became less significant as a predictor of L2 fluency as L2 level increased led the authors to hypothesise that fluency was a “state”, which could change over time, rather than a “trait”, which was relatively fixed and would never change. In this study, the measures which correlated best with rater judgements of fluency were speech rate, pruned speech rate and pause frequency (pauses per second), with pruned speech rate showing the strongest correlations for both Mandarin and Slavic speakers.

An issue not addressed in the research cited above is how fluency relates to other proficiency measures. It seems intuitively likely that more complex and accurate speech may at times be achieved at the expense of fluency, while for more fluent speech, accuracy and complexity may need to be sacrificed. This idea of trade-offs has been extensively studied by Skehan and Foster (e.g. Skehan & Foster, 1999). Fluency is divided into three facets: speed fluency (rate of speech), breakdown fluency (pausing) and repair fluency (repetitions, reformulations etc.). Findings relevant to the present research include evidence that mid-clause pausing is related to (lack of) planning, while end-of-clause pausing is not (Skehan & Foster, 2005). A related study by Tavakoli and Skehan (2005) found that repair fluency was distinct and independent from speed and breakdown fluency. This study is also of interest in that it compares fluency on different tasks – using both narratives with a more structured in content and ones with a looser overall structure. It was found that fluency was greater in the more structured narratives.

To summarise the findings of the above research, the measure which is most often found to be a good indicator of fluency is speech rate. Pruned speech rate is less often used as a measure, but when it is it seems to be equally, or even more, reliable. Mean length of run is the other main measure that reflects fluency. The significance of phonation/pause-time ratio is only supported in some studies, and there is evidence that the amount of pausing does not always distinguish between levels of proficiency or fluency. It seems that it is pause location rather than pause time, length or frequency that is the key factor in distinguishing fluent from non-fluent speech (Tavakoli, 2011).

One good reason for the effectiveness of speech rate and, even more, pruned speech rate, as predictors of fluency is that they are combined measures. Speech rate is a combination of articulation rate and pause time, as changes in each variable will also change speech rate. Pruned speech rate combines articulation rate, pause time and also repair. Mean length of run is also a composite, because although it basically measures pause frequency, it is also a measure of chunking in speech. Speakers who have longer runs show that they are able to plan and execute speech more efficiently, with less need for the mid-clause hesitations that less fluent speakers rely on. Ideally, however, mean length of run should take into account filled pauses, rather than just silent pauses, as a “run” that features a filled pause in the middle should really be considered two runs, not one.
Fluency in Monologue and Dialogue

Most of the studies referred to above involved the study of monologue, and there is a good reason for this: it is much easier to obtain fluency measures when there is only one speaker. Speech rate (words or syllables per minute/second), for example, can be calculated for monologue simply by dividing the number of words or syllables by the total speaking time, and can be done without sophisticated equipment or software. In the case of dialogue, however, the turns of different speakers need to be separated from each other and timed before the total number of words/syllables, the total speaking time, and the speech rate for each speaker can be calculated. This requires special software such as Transcriber or Praat. While Praat software can analyse monologue automatically in order to obtain approximate measures of speech rate and pause time (de Jong & Wempe, 2009), this is impossible with dialogue because no current software can separate the turns and pauses of different speakers. Instead, speech and pause segments need to be manually allocated to each speaker and labelled appropriately. Meanwhile each pause also needs to be allocated to one speaker (or to both in the case of pauses between turns) in order to make measurements of the length, frequency and total time of pauses, as well as of speech rate (which includes pause time).

Besides these basic measurement issues, dialogue presents the researcher with complications which do not occur in monologue. Turns of different speakers may be marked by pausing, latching (where there is no pause, one speaker taking over exactly where the other leaves off), or overlapping (where both speakers speak simultaneously). If there is a pause, it needs to be decided which speaker is pausing. Then there is the question of how to deal with back-channelling, where a speaker responds to another without any attempt to construct a new turn of their own, perhaps to express encouragement or agreement, although it may be to indicate disagreement or disapproval. Should backchannels be analysed in the same way as ordinary turns, or differently?

Due to these difficulties in dealing with monologue, the overwhelming majority of fluency analyses deal with monologue, typically using picture stories (Derwing et al., 2007, 2009; Kormos & Dénes, 2004; Lennon, 1990) or film retelling (Towell et al., 1996). Below I consider some exceptions, and partial exceptions, to this tendency, where dialogue data is used to study fluency.

Freed (1995) elicited speech from learners of L2 French by means of the Oral proficiency Interview. This involves an interviewer following a structured list of questions, so while there is a semblance of a dialogue, it is intentionally one-sided, with only the interviewee producing turns of any length. Freed's article provides no specific explanation of how dialogue is analysed, but presumably speech rate and pausing were measured once the short interventions by the interlocutor had been removed. Possibly the participants’ turns were treated as a series of monologues.

Tauroza and Allison (1990) carried out a survey of speech rates typical of different speech styles and contexts. This was done with listening comprehension tests in mind, but the figures are also useful to bear in mind when studying fluency. The order of speed, from slowest to fastest, was found to be as follows: lecture → radio monologue → interview → conversation. There was some variability in each category, as is to be expected, and considerable overlap in speech rates between categories, but the overall picture was clear: that dialogue is faster than monologue, and that more spontaneous forms of speech are faster than speech in more formal and planned contexts.

Derwing et al. (2004) compared fluency on three different tasks, and correlated measures with rater judgements. Task one was a picture story, task 2 asked participants to describe the happiest moment in their life, while task 3 was a dialogue with a native speaker interlocutor. Pruned speech rate was found to be the best predictor of rater judgements on all tasks. Fluency was lowest for the picture story, and highest for the dialogue, suggesting that dialogue may be more fluent than monologue. However, the results may instead be a consequence of the more
familiar content of tasks 2 and 3, making them easier for these participants. The authors also express concerns that the native speaker interlocutor may have had an influence on the fluency of participants in the third task. It is interesting to note that tasks 1 and 2 were both narratives, but that the true, personal story (the happiest moment in my life) elicited more fluent speech than the picture story.

Finally I will consider briefly the two studies which best address the issues inherent in fluency analysis of dialogue. Riggenbach (1991), the results of which have already been referred to above, is based wholly on interactive dialogue data. Apart from making the usual temporal measurements of speech rate, pausing and repair, the author also investigates a variety of interactive features, such as backchanneling, echoing, questions, laughter, and repair initiation. As she says, “...approaching the notion of conversational fluency is a different task from using monologue-type data (or intraturn analysis only).” (ibid: 439). Different kinds of turn change (overlaps, latching, gaps between turns) are also identified and qualitatively analysed for their effect on fluency. Factors such as the ability to initiate changes of topic, to show comprehension of the other speaker, and to anticipate the end of turns, are identified as central to conversational fluency. This article provides one of the few examples of how to deal with fluency in dialogue, using only a small number of samples but suggesting ways in which analysis of fluency in dialogue might proceed.

Ejzenberg (2000) makes a four-way task comparison involving dialogue and monologue, in cued and uncued forms. Cued tasks are elicited and structured by means of cues provided by the researcher, while uncued tasks lack this kind of structuring and support. Thus the comparison is between cued and uncued dialogues, together with cued and uncued monologues. Overall it was found that dialogues featured faster speech rates than monologues, but that while uncued dialogues were faster than cued, cued monologues were faster than uncued. The effect of tasks being cued was in fact different for the more fluent speakers as compared to the least fluent, but what is most relevant here is that, both for high and for low levels, dialogue was more fluent.

Why Analyse Fluency in Dialogue?

It may be asked why analysis of dialogue is needed, given the complications described, and the relative simplicity of fluency analysis for monologue. Below I give four reasons.

Firstly, there is evidence that dialogue is different, and it is therefore important to investigate and explore these differences. Tauroza & Allison's (1990) findings have already been mentioned, as have those of Ejzenberg (2000). Both studies found evidence that dialogue may be more fluent than monologue. Apart from possible differences in speech rate, many other factors that affect fluency in different categories of speech should be investigated.

Secondly, conversation is the most basic use of language and the most frequent. All languages are (or have been) spoken, and many have no written form, while the most frequent kind of speech is interactive conversation. Van Lier (2004) has argued strongly on theoretical grounds for a dialogical view of language. Following on from this view, it seems logical that language should be studied in the context of its use in dialogue, where ideas, meanings, feelings are exchanged between speakers. Speech may at times take the form of a monologue, but this is not the most common or typical form. Guillot has also argued that dialogue provides a better representation of speaking ability than monologue (Guillot, 1999: 32). It therefore seems advisable to analyse samples of dialogue (with or without monologue data) in studies of fluency.

Thirdly, we should study fluency in dialogue for the sake of balance. There is already much research based on monologue, but little on dialogue. This means that views on fluency are largely based on monologue, with little information and few figures on dialogue to balance these. In order to have a complete picture of oral fluency, we need to know how fluency measures of monologue and dialogue compare.
Finally, in my experience as a teacher, by far the most widespread aim of learners of English is to achieve the ability to hold a conversation, so as to be able to communicate with foreigners (not necessarily L1 speakers of English) using English. It therefore seems sensible to measure fluency in dialogue rather than (or in addition to) fluency in the telling of a story, which is a more specialised skill that many native speakers to not possess.

This Study

This paper reports interim findings from a long term study of fluency in an EFL context at intermediate to advanced levels, where the participants were university students at a Foreign Languages and Literature Department in Taiwan. Using findings from previous research and also experience gained from the current research project, I have identified certain fluency measures which have been found to be reliable. Using these measures, I assess the differences between the students’ performance on two tasks – a monologue storytelling task, and a dialogue discussion task, at two points in time – the students’ first and final (eighth) semester at university. The following research questions are addressed:

1. Are there differences in fluency measures between the monologue story-telling task and the discussion dialogue? Does one task produce a more fluent performance than the other?
2. Are there differences in fluency development over time between tasks? Does one task produce greater rates of change in fluency measures over four years than the other?

METHODOLOGY

Learner Data

The data comes from speech samples taken from 17 Taiwanese university students studying English language and literature at a foreign languages department in Taiwan. Data was collected on two occasions: the first was during their first semester at university, while the second was during their last semester, in year four of their course. On each occasion, students worked in pairs, telling a story to their partner based on a set of pictures, after which they completed a collaborative discussion task. The total time allocated for both tasks was 8-10 minutes, and each pair was given 10 minutes of preparation time. Thus in total each student produced two narratives and two discussions. The narrative picture stories were taken from Fletcher and Birt (1992) and Jones & Baeyer (1983) while the discussions typically involved reaching agreement on a set of priorities (e.g. the characteristics most wanted in friends), a desirable order (e.g. the stages of a relationship), or a list of items to select from (e.g. products that should or should not be advertised). Students used the same picture story on each occasion, while the discussion task was the same for 15 out of 17 students. Student pairings were the same where possible (10 students had the same partner), but in year four 7 students had new partners, and there was one group of three students where a pair from year 1 was joined by a third partner.

Processing & Analysis

The output from the native speaker and non-native speaker participants was transcribed, following conventions described in Foster et al. (2000). Temporal measurements were obtained using “Transcriber” and later “Praat”, both of them open-source software that can be used used to obtain more accurate speech rate, pause and filled pause data. The following temporal measures of fluency were used:

1. **Articulation rate (AR):** words per minute of speaking time, not including pause time.
2. **Speech rate (SR):** words per minute (including silent pause time). All words, filled pauses and repair are included.
3. **Pruned speech rate (PSR):** pruned words per minute (including silent pause time),
4. **Phonation-time ratio (PTR):** the percentage of time that is spent speaking, as opposed to pausing. This is therefore a measure of pause time, and is the converse of pause time ratio.

5. **Mean length of run (MLR):** the mean number of words between silent pauses. This is effectively a measure of pause frequency.

6. **External pause ratio (ePTR):** this is the percentage of time used in external pauses, which come between, rather than within, clauses. Pauses between clauses are thought to not detract from fluency.

7. **Internal pause ratio (iPTR):** this is the percentage of time used in internal pauses, which come within, rather than between, clauses. These pauses are thought to detract from fluency, and are sometimes called “disfluent pauses.”

8. **Percentage of pruned words:** this measures repair by giving the percentage of words that remain after pruned words are discarded. A higher score means less repair, and fewer filled pauses.

9. **Words per clause:** a measure of complexity, as longer clauses will have more complex constituents – for example there may be longer noun or verb phrases.

This is the only measure here that is not a measure of fluency.

Individual and mean figures were thus obtained for all temporal measures on each task. Thus the differences could be identified, and trends of change (or lack of change) identified. The individual and mean figures gave an initial indication of where there were differences, and t-tests were used only on a few selected variables in order to see whether these were significant.

**Issues in the Analysis of Dialogue**

As has been mentioned earlier, special problems arise in the transcription and analysis of dialogue. Such matters are usually explained only in very general terms, if at all, in published articles, which makes understanding (not to mention replication) of studies difficult, and means that each researcher has make their own decision about how to deal with the problems. For this reason, I will describe the problems, and the solutions arrived at, in some detail.

One special feature of dialogue is the occurrence of overlapping turns, when both speakers speak at the same time. To fully account for each speaker’s contribution, it is necessary to have two separate temporal records, each one giving a particular speaker’s complete turns. This means that the combined speaking time for both speakers will often be more than the total time when either participant is speaking). Using Praat (de Jong & Wempe, 2007) it is possible to separate and label the speaker’s turns and pauses manually, and to create two “tiers” so as to reflect independently each speaker’s contributions to the dialogue. In S1’s tier, S1’s words are recorded in full, while S2’s words that overlap with S1 are ignored. In S2’s tier, the reverse is true.

The second major problem when dealing with fluency in dialogue is the attribution and categorisation of pauses. In the present study, pauses in the storytelling monologue are broadly divided into those falling between clauses (external) and those occurring within clauses (internal). This is because it has been found that learners do not necessarily pause more than native speakers in terms of frequency or overall pause time: the principal difference lies in the location of pauses – learners pausing more within clauses than do native speakers (Tavakoli, 2011). However, in the discussion dialogue there is the additional problem of how to allocate pauses between turns. In this study, pauses between turns by the same speaker are considered external (between clause) pauses, just as they would be in monologue. Pauses between complete turns by different speakers (i.e. where the speaker changes) are labelled “turn pauses”. At such moments either speaker could have spoken, so the pause is “shared” between the two speakers, with each speaker being allocated half of the pause time in that place. To allocate the
whole pause to both speakers would exaggerate the total amount of pausing, while if the pauses were not attributed to either speaker, the overall pause rate and time would be kept artificially low.

Turn pauses are therefore, in common with pauses between clauses by the same speaker, considered external pauses, as they come between complete turns, and so between complete clauses. However, in some cases there is a pause between turns by different speakers where the first turn (and the last clause in that turn) is incomplete. In such cases, the pause is allocated to the first speaker and designated as an internal (within-clause) pause, on the grounds that the speaker is clearly pausing due to an inability to complete the clause without hesitation. The other speaker starts speaking before the first speaker can decide how to complete the utterance.

Apart from these problems about how to allocate turns and pauses, there are further issues that need to be resolved when transcribing and analysing dialogue. A typical feature of dialogue, and especially of everyday conversation, is what is called “backchannelling”. This is the term used to describe listener responses to another speaker, where there is no attempt to intervene with a complete clause or turn, but instead the listener wishes to show that he/she is paying attention, and perhaps agreeing (or disagreeing) with what the speaker is saying. The most typical backchannels are “uhun”, but “mnn”, “yes” and “really” are other frequent examples. Such “backchannels” need to be distinguished from what are simply short responses, as when speakers respond to an intervention with a simple “yes” or “OK”, thus declining to produce a longer turn of their own.

If backchannels were all included in fluency analysis in the same way as other turns, some speakers’ measures would be affected. For example, a frequent backchanneller would have a lower MLR (mean length of run) than they would have if the backchannels were missing or ignored, as backchannels are typically only one or two syllables in length. On the other hand, if all one-word responses were ignored, some speakers’ MLR might be inflated, as frequent undeveloped one-word turns would be ignored in this analysis. The solution adopted here is to take non-verbal listener responses (“uhun or mnn”) as backchannels, while all verbal listener responses (yes, oh, or really) are taken to be short responses, on the grounds that it is impossible to distinguish in any systematic way between short responses and verbal backchannels. A further difficulty is that it is often difficult to distinguish nonverbal backchannels from filled pauses purely on the basis of the sound made, as this may be the same for both – e.g. mm, er, uh. The difference lies in the use to which they are put and where they are located – filled pauses coming as part of a turn by the same speaker, while backchannels come during an interlocutor’s turn.

One final feature that occurs more frequently in dialogue than in monologue is laughter. In the data studied in this investigation, laughter occurs only rarely in monologue, but is frequent in dialogue, especially for certain pairs of students. Laughter serves a variety of purposes in speech (Jefferson, 1984), including embarrassment, hesitation, nervousness and as an alternative to backchannelling, apart from the most obvious reason – reacting to humorous actions, words or situations. Due to its multiple functions, laughter cannot be equated with a single temporal variable (such as pausing, filled pausing or speaking), but it can play an important role in qualitative analysis, where it may be seen to interact with other features of fluency such as speech rate and silent or filled pausing. Speakers who laugh more may need to pause less, for instance.

RESULTS AND DISCUSSION
Results
The key finding from the present data is that speech rates for the discussion dialogue were faster than those for the picture story narrative, when comparing the two different task performances by the same speaker at the same time (i.e. both in year 1 or both in year 4). It was also found that these students increased their speech rate more, on average, for the dialogue
than for the story task over the four year period. The complete individual results are presented in the appendix, where each student's mean figures for all measures are shown. Below I compare the overall mean results.

Table 1 shows how the mean figures for fluency measures compare on the two tasks. The table also shows mean differences between the tasks (when performed at the same time – in year one or year four). On all measures except one, the discussion dialogue produced faster or less hesitant speech: AR, SR and PSR are higher, there is less repair (the percentage of pruned words is higher), pause time, including both external and internal pause time, is less, and the mean length of run between pauses is longer. Only for the complexity measure – pruned words per clause – is the mean score for the story higher. It should be noted, however, that only the measures for SR, PSR and external (between clause) pausing show a significant difference. Both internal and external pausing are lower for the dialogue, although only external pausing is significantly lower.

Table 1. Mean Scores for Temporal Measures on Story Monologue and Discussion Dialogue

<table>
<thead>
<tr>
<th></th>
<th>AR</th>
<th>SR</th>
<th>PSR</th>
<th>PTR</th>
<th>MLR</th>
<th>%pw</th>
<th>pwpc</th>
<th>ePTR</th>
<th>iPTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Story monologue</td>
<td>181.2</td>
<td>121.4</td>
<td>94.8</td>
<td>67.8</td>
<td>3.9</td>
<td>77.3</td>
<td>5.9</td>
<td>16.7</td>
<td>15.5</td>
</tr>
<tr>
<td>Discussion dialogue</td>
<td>191.6</td>
<td>138.2</td>
<td>111.8</td>
<td>72.7</td>
<td>4.1</td>
<td>80.6</td>
<td>5.6</td>
<td>13.7</td>
<td>13.6</td>
</tr>
<tr>
<td>% difference</td>
<td>5.7</td>
<td>13.9***</td>
<td>17.9***</td>
<td>7.2</td>
<td>3.1</td>
<td>4.2</td>
<td>-4.7</td>
<td>-18.0*</td>
<td>-11.9</td>
</tr>
</tbody>
</table>

Data compared: 34 story monologues and 34 discussions from 17 students. Tasks were compared for the same year – e.g. story in year 1 with discussion year 1.

• = sig. p<0.05, ** = sig. p<0.01, ***= sig. P<.001. T-test 1-tailed.

The narrative monologue, meanwhile, does produce more complex speech than the discussion – as indicated by the pruned words per clause measure, although this difference is not significant. Complexity is not a fluency measure, so this finding is not unexpected, as more complex speech is likely to be less fluent, requiring as it does more planning and monitoring.

The differences between the two tasks raise the question of whether performances on one task may change more than on the other. Possibly, for example, the picture stories might become more fluent, as the year one scores are lower, thus allowing more room for improvement. Alternatively, it might be that the difference between tasks becomes wider, and that the discussion dialogue is the task that provides optimal conditions not only for fluent performance, but also for fluency development. The second comparison in this results section therefore is one between the scores in year one and year four, taking each task separately. The results are shown in table 2.

Table 2. Change on Temporal Measures for Story and Discussion Separately

<table>
<thead>
<tr>
<th></th>
<th>AR</th>
<th>SR</th>
<th>PSR</th>
<th>PTR</th>
<th>MLR</th>
<th>%pw</th>
<th>pwpc</th>
<th>ePTR</th>
<th>iPTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>story1</td>
<td>176.8</td>
<td>118.4</td>
<td>89.8</td>
<td>66.8</td>
<td>4.0</td>
<td>74.6</td>
<td>5.5</td>
<td>17.7</td>
<td>15.6</td>
</tr>
<tr>
<td>story2</td>
<td>185.6</td>
<td>124.5</td>
<td>99.8</td>
<td>68.8</td>
<td>3.9</td>
<td>80.1</td>
<td>6.3</td>
<td>15.7</td>
<td>15.4</td>
</tr>
<tr>
<td>change</td>
<td>5.0</td>
<td>5.1</td>
<td>11.1***</td>
<td>3.0</td>
<td>-3.2</td>
<td>7.5</td>
<td>12.8</td>
<td>-11.2***</td>
<td>-0.8</td>
</tr>
<tr>
<td>discussion</td>
<td>179.7</td>
<td>128.9</td>
<td>103.6</td>
<td>72.8</td>
<td>3.9</td>
<td>79.8</td>
<td>5.5</td>
<td>15.0</td>
<td>12.4</td>
</tr>
<tr>
<td>disc2</td>
<td>203.5</td>
<td>147.6</td>
<td>119.9</td>
<td>72.7</td>
<td>4.2</td>
<td>81.3</td>
<td>5.8</td>
<td>12.4</td>
<td>14.9</td>
</tr>
<tr>
<td>change</td>
<td>13.2**</td>
<td>14.5*</td>
<td>15.7</td>
<td>-0.2</td>
<td>9.4</td>
<td>1.9</td>
<td>5.5</td>
<td>-17.0</td>
<td>20.0</td>
</tr>
</tbody>
</table>

Data from 17 students. Comparisons between stories and discussions in year 1 and in year 2.
The Variability of Fluency in Dialogue and Monologue

* = sig. p<0.05, ** = sig. p<0.01, *** = sig. P<.001. T-test 1-tailed.

Here it can be seen that AR, SR and PSR, and also MLR, all increase more for the discussion than for the story task, although the increase in PSR is large and significant for both tasks. On the other hand, the percentage of pruned speech and pruned words per clause increase more for the story task. It is interesting to note that MLR was actually shorter on the story task in year 4 as compared to year 1. On pause location there is no consistent trend: external pausing decreases more for the discussion, but internal pausing increases considerably for the discussion, while not changing for the story task. To sum up, there is more change on the discussion tasks for rate of speech and MLR, while for the story task there are greater changes in the amount of repair and in one measure of complexity - the number of pruned words per clause. There is certainly no sign of convergence - of picture story fluency “catching up” with discussion fluency.

Discussion

Referring back to the research questions posed earlier, we can now say that, for the group of learners under study:

1. The discussion dialogue task produced higher rates of speech (AR, SR PSR) than the story monologue. However, clausal complexity (as measured by the number of words per clause) was higher for the story monologue.
2. Increases of speech rate measures and length of run were higher for the discussion dialogue task than for the story monologue. On the other hand, the story monologues showed a greater reduction in the amount of repair and a more sizeable increase in clausal complexity.

There is confirmation therefore of previous findings that dialogue is more fluent, in that it produces faster speech rates, than monologue. Part of the difference in speech rates is due to reduced pause time in the dialogues, which may be a result of turn taking – longer hesitations are avoided because when one speaker hesitates, another steps in. Thus dialogues may naturally be more fluent because of faster turn-taking. Nevertheless, the figures above show that articulation rate is also faster in dialogue, so the difference in pause time is not wholly responsible for any improvements.

It may be that monologue, particularly of the kind used in speaking tasks by learners of English, encourages slower delivery. One reason for this could be that the lack of an interlocutor in monologue makes the speaker more relaxed about time and speed of speech, there being no pressure to speed up delivery, no threat of interruption, and no need to “hold the floor”. Another reason could be that in monologue speakers do more monitoring of their own speech as they seek to improve aspects of their language such as accuracy or complexity. In dialogue there is little attention available for monitoring, as attention has to be paid to what the interlocutor is saying, and how to respond to it. The cognitive demands of dialogue simply do not allow for much monitoring. In monologue, the reverse is the case – the attention that would, in dialogue, be paid to the interlocutor is used, instead, to monitor the speaker's own speech, possibly producing greater accuracy but acting as a restraint on fluency. The greater clausal complexity for story monologues is consistent with this explanation of the data, as it suggests that these monologues are more carefully planned.

CONCLUSION

This research lends support to the idea that dialogue is more fluent than monologue, in so far as speech rates in dialogue are faster. This is of interest to teachers who teach speaking classes, to testers of spoken English, and to researchers wishing to investigate fluency.
For teachers, it is important to be aware of the distinct characteristics of different kinds of speech, and that the same speech rates may not be appropriate, or even possible, in different contexts. Teachers need to provide practice in a variety of speaking activities in order to help learners cope with the special requirements of each type. Both monologic and dialogic speech should be practised in class. In my own experience, however, some teachers focus almost exclusively on the oral production of memorised texts in the form of speeches, recitations or presentations, neglecting to provide practice with the more spontaneous kind of speech that occurs in conversation and discussion.

For oral examiners, the message is that the highest levels of fluency are usually reached only in dialogue, especially in informal conversation. This makes tests that are based on examiner–candidate interaction, or where two candidates speak to each other, more likely to allow candidates to show their best levels of fluency than tests where candidates speak to a computer.

In the field of research into oral proficiency, fluency and task based language learning, picture stories are widely used as an elicitation tool in such research, and monologue is generally preferred to dialogue because it is more convenient to measure and analyse for fluency. Yet researchers need to be aware that monologues are not the only form of speech that needs to be investigated, and that tasks such as picture stories have peculiarities and limitations that affect their validity as tests of the ability to communicate in a foreign language. Dialogue tasks seem to elicit more fluent speech, and findings based on dialogue may reflect everyday speaking ability better than those based on monologue. Story telling tasks can elicit different levels of performance from speakers who have different attitudes to the task: some may use careful speech, seeking precision and accuracy, while others may speak more spontaneously and try to make their stories entertaining. There is therefore a need for more research into fluency to be based, at least partly, on dialogue, and on task types that reflect everyday speech more closely.

More data, based on larger samples of dialogue and monologue, is needed to confirm and extend these findings. With more studies that tackle the difficulties of analysing dialogue, hopefully a consensus will emerge on the ways in which the various methodological issues should be solved, allowing for cross-study comparisons. It would also be interesting to make additional measures of linguistic accuracy and complexity, in order to see whether there is any trade-off effect between fluency, on the one hand, and accuracy and complexity on the other. It may be that dialogue allows for greater fluency at the expense of accuracy and complexity, while monologue encourages greater accuracy and complexity at the expense of fluency. It may also be that different types of monologue and dialogue elicit different levels of fluency, as was found by Derwing et al. (2004), where participants spoke more fluently about the happiest moment of their lives than on a picture story. For comparisons across studies, it is essential that task types be comparable, and that the significance of any differences between tasks should be clear. It would also be interesting to investigate learners from other universities in Taiwan, with lower and higher levels of students, as such findings may vary. Universities in Taiwan have a clear “pecking order”, and this research was carried out at a middle-upper ranking one (not NTU).

Meanwhile, a note of caution on the use of temporal variables is required. No single temporal measure should be used on its own to gauge fluency, as perceptions of fluency seem to depend on a range of factors, it being the combination of these that produces the overall impression. That is why measures that combine other measures, of which Pruned Speech Rate is the best example, are the most reliable ones. Speakers may achieve a fast Articulation Rate yet make long pauses between utterances; or they may produce a fast Speech Rate while making frequent repetitions, false starts and reformulations. In such cases, the overall impression may not be of fluent speech, despite the high AR or SR. PSR will be a better indicator in such cases, measuring as it does a combination of articulation rate, pausing and repair.
REFERENCES


The Variability of Fluency in Dialogue and Monologue

Appendix: Table 3. Overall results for 17 individual students, with mean figures for each measure. Students are given numbers instead of their names. The letter refers to the task – S = story monologue, D = discussion dialogue. The final number refers to the year – 1 = 1st year, 2 = 2nd year, 3 = 3rd year, 4 = 4th year. So SD4 means student 5, discussion, 4th year.

<table>
<thead>
<tr>
<th>Year</th>
<th>AR</th>
<th>SR</th>
<th>PSR</th>
<th>PTR</th>
<th>MLR</th>
<th>%spm</th>
<th>ePTR</th>
<th>ipPTR</th>
<th>pwpC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1ST</td>
<td>158.6</td>
<td>116.6</td>
<td>101.1</td>
<td>77.5</td>
<td>4.3</td>
<td>4.3</td>
<td>17.6</td>
<td>8.9</td>
<td>5.8</td>
</tr>
<tr>
<td>1ST</td>
<td>160.9</td>
<td>121.8</td>
<td>82.4</td>
<td>4.3</td>
<td>4.3</td>
<td>11.7</td>
<td>5.9</td>
<td>4.2</td>
<td></td>
</tr>
<tr>
<td>2ST</td>
<td>182.5</td>
<td>98.1</td>
<td>54.7</td>
<td>3.6</td>
<td>3.7</td>
<td>16.3</td>
<td>9.0</td>
<td>5.6</td>
<td></td>
</tr>
<tr>
<td>2ST</td>
<td>154.4</td>
<td>99.3</td>
<td>79.0</td>
<td>64.3</td>
<td>3.2</td>
<td>3.2</td>
<td>16.3</td>
<td>19.4</td>
<td>6.1</td>
</tr>
<tr>
<td>2ST</td>
<td>159.0</td>
<td>148.9</td>
<td>97.7</td>
<td>73.4</td>
<td>4.2</td>
<td>4.2</td>
<td>9.8</td>
<td>16.8</td>
<td>5.8</td>
</tr>
<tr>
<td>2ST</td>
<td>176.0</td>
<td>103.9</td>
<td>57.6</td>
<td>5.7</td>
<td>5.7</td>
<td>14.0</td>
<td>7.5</td>
<td>4.2</td>
<td></td>
</tr>
<tr>
<td>3ST</td>
<td>170.0</td>
<td>70.0</td>
<td>45.8</td>
<td>3.7</td>
<td>3.7</td>
<td>14.0</td>
<td>3.1</td>
<td>3.1</td>
<td></td>
</tr>
<tr>
<td>3ST</td>
<td>205.6</td>
<td>122.7</td>
<td>66.8</td>
<td>59.7</td>
<td>2.8</td>
<td>2.8</td>
<td>17.7</td>
<td>22.6</td>
<td>4.9</td>
</tr>
<tr>
<td>3ST</td>
<td>184.4</td>
<td>82.3</td>
<td>55.6</td>
<td>4.3</td>
<td>4.3</td>
<td>29.9</td>
<td>26.4</td>
<td>4.8</td>
<td></td>
</tr>
<tr>
<td>4ST</td>
<td>197.8</td>
<td>127.2</td>
<td>91.3</td>
<td>66.3</td>
<td>3.6</td>
<td>3.6</td>
<td>60.1</td>
<td>12.1</td>
<td></td>
</tr>
<tr>
<td>4ST</td>
<td>171.3</td>
<td>111.2</td>
<td>92.3</td>
<td>64.9</td>
<td>3.3</td>
<td>3.3</td>
<td>17.9</td>
<td>20.7</td>
<td>6.4</td>
</tr>
<tr>
<td>5ST</td>
<td>154.1</td>
<td>60.7</td>
<td>72.0</td>
<td>62.8</td>
<td>3.5</td>
<td>3.5</td>
<td>20.7</td>
<td>16.6</td>
<td>6.0</td>
</tr>
<tr>
<td>6ST</td>
<td>170.2</td>
<td>115.7</td>
<td>88.3</td>
<td>67.4</td>
<td>3.9</td>
<td>3.9</td>
<td>23.0</td>
<td>9.5</td>
<td>5.4</td>
</tr>
<tr>
<td>7ST</td>
<td>170.2</td>
<td>115.2</td>
<td>84.7</td>
<td>4.7</td>
<td>4.7</td>
<td>23.0</td>
<td>9.5</td>
<td>5.4</td>
<td></td>
</tr>
<tr>
<td>8ST</td>
<td>158.6</td>
<td>85.8</td>
<td>55.4</td>
<td>5.9</td>
<td>5.9</td>
<td>23.0</td>
<td>9.5</td>
<td>5.4</td>
<td></td>
</tr>
<tr>
<td>9ST</td>
<td>125.2</td>
<td>70.9</td>
<td>56.4</td>
<td>58.9</td>
<td>2.3</td>
<td>2.3</td>
<td>24.3</td>
<td>16.8</td>
<td>5.1</td>
</tr>
<tr>
<td>10ST</td>
<td>193.8</td>
<td>128.2</td>
<td>95.6</td>
<td>66.2</td>
<td>3.7</td>
<td>3.7</td>
<td>18.0</td>
<td>15.9</td>
<td>5.2</td>
</tr>
<tr>
<td>11ST</td>
<td>159.3</td>
<td>155.8</td>
<td>117.4</td>
<td>76.3</td>
<td>4.0</td>
<td>4.0</td>
<td>12.5</td>
<td>11.3</td>
<td>4.9</td>
</tr>
<tr>
<td>12ST</td>
<td>173.8</td>
<td>126.4</td>
<td>114.2</td>
<td>74.7</td>
<td>4.2</td>
<td>4.2</td>
<td>15.4</td>
<td>11.7</td>
<td>6.3</td>
</tr>
<tr>
<td>13ST</td>
<td>176.0</td>
<td>81.4</td>
<td>71.2</td>
<td>74.3</td>
<td>2.2</td>
<td>2.2</td>
<td>12.2</td>
<td>13.5</td>
<td>5.8</td>
</tr>
<tr>
<td>14ST</td>
<td>211.2</td>
<td>142.6</td>
<td>130.5</td>
<td>67.5</td>
<td>4.5</td>
<td>4.5</td>
<td>21.7</td>
<td>10.8</td>
<td>5.3</td>
</tr>
<tr>
<td>15ST</td>
<td>192.3</td>
<td>128.9</td>
<td>114.1</td>
<td>60.9</td>
<td>4.7</td>
<td>4.7</td>
<td>24.9</td>
<td>8.1</td>
<td>6.1</td>
</tr>
<tr>
<td>16ST</td>
<td>161.4</td>
<td>95.0</td>
<td>46.7</td>
<td>58.0</td>
<td>2.7</td>
<td>2.7</td>
<td>15.6</td>
<td>56.4</td>
<td>3.7</td>
</tr>
<tr>
<td>17ST</td>
<td>162.0</td>
<td>104.2</td>
<td>76.1</td>
<td>63.1</td>
<td>2.9</td>
<td>2.9</td>
<td>12.4</td>
<td>24.5</td>
<td>5.5</td>
</tr>
<tr>
<td>18ST</td>
<td>128.2</td>
<td>129.4</td>
<td>88.4</td>
<td>70.5</td>
<td>4.0</td>
<td>4.0</td>
<td>14.7</td>
<td>7.5</td>
<td>5.3</td>
</tr>
<tr>
<td>19ST</td>
<td>183.7</td>
<td>175.0</td>
<td>91.0</td>
<td>51.8</td>
<td>6.8</td>
<td>6.8</td>
<td>5.8</td>
<td>5.8</td>
<td>5.7</td>
</tr>
<tr>
<td>20ST</td>
<td>193.0</td>
<td>178.1</td>
<td>140.7</td>
<td>91.8</td>
<td>6.8</td>
<td>6.8</td>
<td>5.8</td>
<td>5.8</td>
<td>5.7</td>
</tr>
<tr>
<td>21ST</td>
<td>184.2</td>
<td>128.6</td>
<td>90.2</td>
<td>69.8</td>
<td>4.2</td>
<td>4.2</td>
<td>21.2</td>
<td>9.0</td>
<td>4.7</td>
</tr>
<tr>
<td>22ST</td>
<td>180.9</td>
<td>126.9</td>
<td>98.3</td>
<td>78.4</td>
<td>3.1</td>
<td>3.1</td>
<td>15.5</td>
<td>8.0</td>
<td>3.2</td>
</tr>
<tr>
<td>23ST</td>
<td>167.3</td>
<td>112.0</td>
<td>85.1</td>
<td>67.0</td>
<td>3.8</td>
<td>3.8</td>
<td>14.7</td>
<td>18.3</td>
<td>5.7</td>
</tr>
<tr>
<td>24ST</td>
<td>177.2</td>
<td>115.2</td>
<td>95.0</td>
<td>65.0</td>
<td>3.8</td>
<td>3.8</td>
<td>14.7</td>
<td>18.3</td>
<td>5.7</td>
</tr>
</tbody>
</table>

The future results for all measures: S = story monologue, D = discussion dialogue. The final number refers to the year – 1 = 1st year, 2 = 2nd year, 3 = 3rd year, 4 = 4th year. So SD4 means student 5, discussion, 4th year.