

## Investigating Teachers' Perceptions of Fluency

Ellen Head

*Miyazaki International College*

### Reference Data:

Head, E. M. (2019). Investigating teachers' perceptions of fluency. In P. Clements, A. Krause, & P. Bennett (Eds.), *Diversity and inclusion*. Tokyo: JALT.

This paper is a report on an investigation into spoken fluency as perceived by EFL professionals in China. Building on Segalowitz's (2010) analysis of fluency into utterance and cognitive and perceived fluency, a study by Dore (2016) was replicated. China-based EFL teachers were asked to rate three speech samples then complete a questionnaire explaining the reasons for their ratings and their beliefs about fluency. The overall questionnaire results were similar to Dore's, but the sample ratings showed greater divergence, with only 56% agreement on which sample exhibited a higher degree of fluency. The majority preferred a performance with fewer hesitations; the others appeared to be influenced by grammatical proficiency and accent. Respondents agreed that effortlessness and coherence were significant components of fluency. Ratings of components of utterance fluency correlated with automaticity, supporting the view that utterance fluency could be predictive of perceived fluency.

本論は、中国のEFLの教官たちに行った、流暢さに関するアンケート調査の報告である。流暢さの要素を、発声、認知、知覚される流暢さに分類したセガロヴィッツ (2010) の考え方を用い、ドーレ (2016) による研究を再現した。計 46 人のEFLの教官たちが3つのスピーチサンプルを評価し、評価の理由と流暢さについての信条を説明する質問事項に答えた。結果は全体的にドーレの研究結果にかなり似たものになったが、評価において、ドーレの結果より大きな分散が見られた。サンプルがより高い程度の流暢さを示した場合でも、わずか56%の一致であった。多くは躊躇のない話し方を好み、文法の習熟と発音に影響を受けたのは少数だった。回答者たちが同意したのは、難なく話すことと首尾一貫性が流暢さの重要な要素であるということだった。発声の流暢さに関する評価は、自動的処理に相関し、発声の流暢さが知覚される流暢さを予測するという見方を支持していた。

**F**luency is a term used in speaking rating scales such as the Common European Framework of Reference for Languages (CEFR), the International English Language Testing System (IELTS), and the Eiken Test in Practical English Proficiency (Eiken). However, Lennon (1990) pointed out it is ambiguous, used by nonspecialists to refer to general speaking ability but also used in more technical contexts such as exam assessment criteria. Research into components of fluency led to a model by Segalowitz (2010) in which the following three aspects were defined: cognitive fluency (mental processing of language), utterance fluency (articulatory and phonological), and perceived fluency (the impression created). According to Segalowitz's model, we can only make inferences about cognitive fluency. Utterance fluency has temporal and acoustic outputs that we can measure objectively. Perceived fluency is the listener's attempt at inferring about cognitive fluency. In this paper, I present a replication of a mixed-methods research study of perceptions of fluency by Dore (2016).

Research into the relationship between utterance fluency and cognitive fluency has found that human ratings of spoken performance align well with acoustic measurements, which are related to utterance fluency (Derwing, Rossiter, Monroe, & Thompson, 2004; Park, 2016; Saito, Trofimovich, & Isaacs, 2017). Saito et al. (2017) also compared perceptions of fluency by experienced and inexperienced raters. They correlated well on global measures, although experienced raters were more accurate on measures like vocabulary range. Kormos and Denes (2004) found that speech rate, mean length of utterance, and number of stressed words per minute could predict human raters' assessments of fluency. De Jong (2018) claimed that automated measures of utterance fluency predict overall fluency. Tests like the Pearson Test of English Academic already use automated scoring. Muller, Adamson, Herder, and Brown (2014) highlighted the need for measures of fluency that are available and useful to teachers. Dore's (2016) study is important as it bridges the gap between technical terms and terms that teachers use. If, as Dore suggested, the local context impacts perceptions of fluency, teachers working in Asia may perceive fluency differently from those working in a European

context. In this paper I report on a replication of Dore's study with teachers in China and discuss whether we can make systematic generalizations about how context influences perceptions of fluency.

### Dore's Study of Perceptions of Fluency

Dore (2016) compared the perceptions of oral fluency of 48 native-speaking EFL teachers at universities in Italy and the United Kingdom. She asked participants to rate three samples of learner narrative recorded by native speakers of Chinese and Korean. Respondents then answered open-ended questions about the ratings and their beliefs about fluency. Finally, she asked respondents to give ratings for components of fluency such as speech rate and accent. They agreed that "effortlessness" and "coherence" were important, but they were divided on accent, accuracy, automaticity, and "colloquialness." Dore reported that differences correlated with country of residence and length of training; Italy-based teachers gave higher importance to accent and UK-based teachers to automaticity. Dore theorized that UK-based university EFL teachers might be more exposed to current ideas in ELT.

The first research question for this study was adapted from Dore's study, whereas the other questions emerged from the data.

- RQ1. Do teachers in China pay attention to the same aspects of fluency as teachers in Italy and the UK?
- RQ2. Are their speech ratings consistent with their declared beliefs?
- RQ3. Can components be grouped into common factors underlying perceived fluency?

### Rationale for Replication

Saville (2017) suggested that assessment is only meaningful in the context of a community of practice. Testing organizations have a responsibility to provide transparent feedback to develop shared understandings of the terms they use. Although frameworks such as the CEFR supposedly have global reach, local ideas about what constitutes competent performance may vary widely. When I worked in Japan (2000-2014) and in China (2015-2017), my duties included conducting oral assessments, such as in-house speaking tests for placement and grading and competitive tests of suitability for study abroad. Even when working with detailed criteria, disagreements of interpretation were frequent. In high-stakes examinations this issue is particularly

critical. Thus, as the Japanese government announced in 2018 that commercial tests such as IELTS and TOEFL will be accepted as qualifications or part-qualifications for university entrance in Japan ("Eight private English tests accepted," 2018, March 27), it is very important to develop a shared understanding of fluency.

Previous researchers have tried to identify the main components that predict perceived fluency (Saito et al., 2017; Van Moere, 2012). However, they focused on correlations between acoustic features of speech and those noticed by raters, rather than on perceived fluency.

### Researching Perceptions of Fluency Among University Teachers of EFL in China

Muller et al. (2014) argued that fluency research is needed in Asia because most fluency studies have focused on BANA/European areas. They commented, "Similar patterns of low confidence, low motivation and low ability can be seen in numerous contexts in Asia" (p. 1). Although there are signs of reform, rote learning was ingrained in the Chinese and Japanese state education systems for many years. Frequent objective tests with a single correct answer discourage students and teachers from risk-taking. This creates difficulties in relation to developing fluency. Murphey (2011) reported on the problems caused by the Japanese university entrance system in the 1990s. Although the falling birthrate has changed the situation in Japan, the high school system still tends to prioritize accuracy over fluency. In China the situation is different but arguably even more problematic (Cheng, 2008). Thirteen million take the college English test annually (Zheng & Cheng, 2008). Some become discouraged after trying repeatedly to pass various tests, investing large amounts of money, and failing without knowing why (Yu, 2014). My research into perceptions of fluency was carried out among a group of ELT professionals working in China during 2016. Eleven university EFL teachers from Shanghai and Shantou completed the tasks online. A further 35 teachers at an Australian-affiliated university in Shanghai did the survey at a workshop.

The responses from both groups were examined to check if the range and nature of the responses were similar by comparing the range, means, and SDs before combining the results. Three of the respondents to the online survey could not access the audio, but their responses were included in the component survey.

Among the 43 respondents, the average length of teaching experience was 10.7 years, ranging between 1 and over 20 years. Two had PhDs, 19 had an MA in TESOL or applied linguistics, and 14 had CELTA or DELTA. Eleven had experience of examining in in-house or external contexts. Two had BAs with no formal ESOL training.

## Method

Participants listened to three recordings of a story told in English by native speakers of Chinese (Mandarin) and rated their fluency on a scale of 1 to 7. Next, they answered open questions about the recordings and fluency. Finally, they rated the importance of 20 components of fluency to their concept of fluency (see Appendix A). The open questions were the same as Dore's (2016, pp. 22-23), and the components were based on those reported in her paper.

The stimulus was a picture story about winning the lottery. I added the question "What would you do if you won the lottery?" at the end, to round off the interaction. The speakers were a convenience sample drawn from friends of different ages, gender, and English ability. Their speeches are summarized in Table 1. (The first speaker had more preparation time due to an equipment malfunction.) Transcriptions are provided in Appendix B.

Table 1. Characteristics of Speech Samples

Speaker (gender, age)	Prep time (minutes)	Speech rate (syllables per second)	Time (story) (seconds)	Number of words (story)
Speaker 1 (female, 50s)	3	2.54	100	207
Speaker 2 (male, 30s)	1	1.76	45	57
Speaker 3 (male, 11)	1	1.88	49	78

## Results

### *Global Ratings for the Speakers*

Twenty-six respondents rated Speaker 1 as the most fluent. However, the overall scores for S1 ranged from 2 to 7, as shown in Table 2. Thirteen responders rated S3 as the most fluent and four rated S3 and S1 equal. Thus, there was only 60% agreement as to the most proficient speaker. This was very different from Dore's study: Her respondents were nearly 90% in agreement.

Table 2. Global Ratings for Each Speaker (N = 43)

Speaker	Mean	SD	Range
1	5.16	0.83	2 - 7
2	3.44	0.79	2 - 4.5
3	4.85	0.92	3 - 6

Note. Scale: 1 (*very fluent*) to 7 (*not at all fluent*).

### *Quantitative Survey of Components of Fluency*

#### *Research Question 1: Do Teachers in China Pay Attention to the Same Aspects of Fluency as Teachers in Italy and the UK?*

The ratings for the components are similar to those in Dore's study, as can be seen in Table 3. The highest-rated components were the same in both surveys: effortlessness and coherence. Accent and colloquialness were rated low in both surveys. However, formulaic sequences were considered important by Dore's respondents but not by mine. Although the mean ratings for most components were between 3 and 4 in both Dore's and my studies, the range of scores and variance were much greater in my study. Scores ranged from 1 to 5 for all components and standard deviations were generally high.

Table 3. Ranking of Fluency Variables (N = 46)

Variable	China-based respondents		Dore (2016) respondents	
	Mean	SD	Mean	SD
<b>FLUENCY</b>				
Effortlessness	4.0	1.27	4.08	0.71
Number and length of pauses	3.72	1.31	3.83	0.75
Automaticity	3.60	1.52	3.79	0.898
No. of hesitations	3.93	1.19	3.79	0.898
Length of runs	3.5	1.56	3.73	.676
Rate of speech	3.28	1.38	3.69	.748
Reformulations	3.52	0.99	3.48	.850

Head: *Investigating Teachers' Perceptions of Fluency*

Variable	China-based respondents		Dore (2016) respondents	
	Mean	SD	Mean	SD
Fillers	3.72	1.10	3.38	.890
Repetition	3.37	0.96	3.02	.934
<b>COMPLEXITY</b>				
Complexity of topic content	3.06	1.11	3.08	1.127
Variety of vocabulary	3.17	1.11	3.02	1.120
Complexity of structures	3.06	0.96	2.79	1.110
Accuracy	3.09	1.06	2.69	1.095
<b>PHONOLOGY</b>				
Native-like rhythm	3.76	1.11	3.48	1.271
Accent	2.48	1.29	2.48	1.052
<b>GLOBAL ASPECTS</b>				
Coherence	4.04	1.04	4.02	.956
Coping in social situations	3.52	1.12	3.69	1.133
Global proficiency	3.06	1.43	3.60	1.026
Formulaic Sequences	2.78	1.27	3.50	1.185
Colloquialness	2.89	1.10	2.69	1.095

Note. Likert-scale ratings: 5 = very important to fluency; 1 = not important.

In addition to the statistical analysis as done by Dore, factor analysis was performed to investigate factors underlying the components for the whole sample and also for the subgroups of participants who rated S1 or S3 highest. XLSTAT software (2019) was used to make a Spearman correlation matrix and chart; see Figure 1. Effortlessness, coherence, and automaticity loaded together with acoustic components (pause, rate, hesitation, run-length, rhythm, reformulation, fillers), accounting for 27.31% of the variability as F1 (Cronbach's alpha ( $\alpha$ ) = .867). The second factor, accounting for 17.55% of the variability, consisted of mainly nonacoustic components: topic complexity, vocabulary variety, coping skills, structural complexity, accent, and grammatical accuracy ( $\alpha$  = .790).

Figure 2 shows the analysis of the responses of those who preferred S3 or rated S3

and S1 equally. The pattern is similar to Figure 1, with the difference that colloquialness loads onto Factor 1. Cronbach's alpha is .926 for F1 and .889 for F2. Figure 3 shows the responses of those who preferred S1. Factor 1 is similar to that for the whole group with the addition of vocabulary ( $\alpha$  = .718). Factor 2 contains a positive value for topic complexity correlated negatively with hesitation, run-length, and automaticity. Teachers who preferred S1 found topic complexity more important and acoustic factors less important than those who chose S3. However, a large part of the variance remains unaccounted for, as will be discussed later.

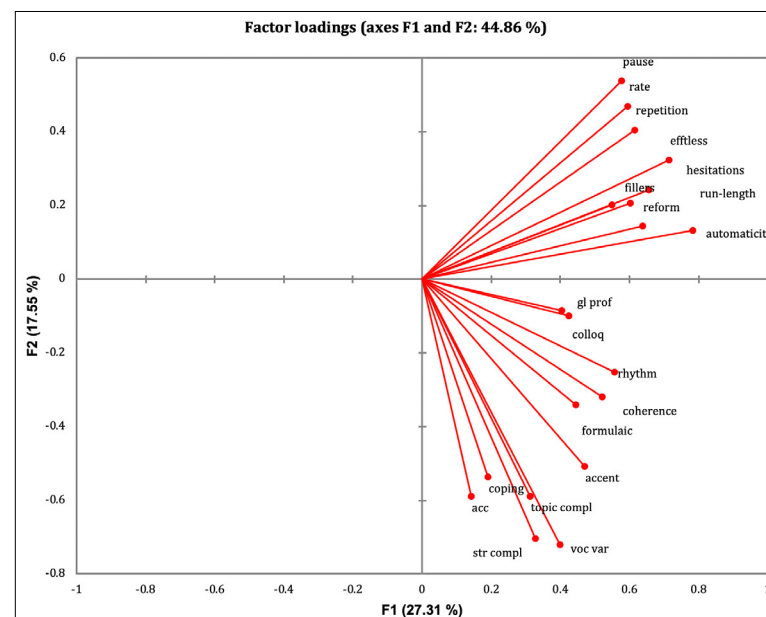


Figure 1. Factor loadings for fluency variables (all respondents).

Head: Investigating Teachers' Perceptions of Fluency

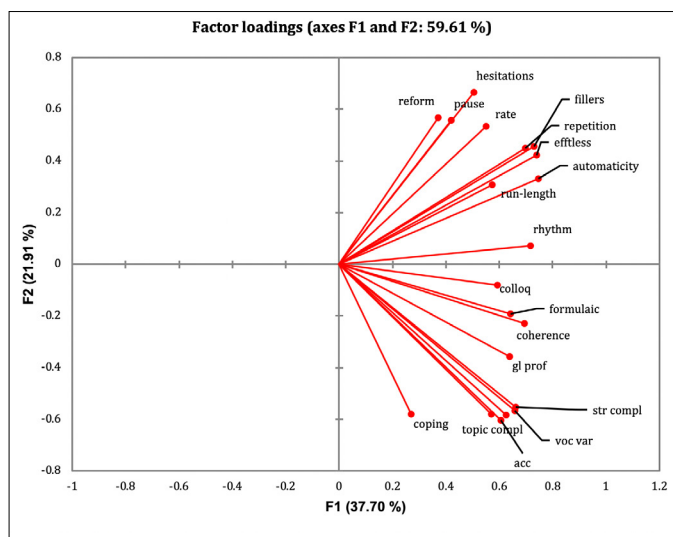


Figure 2. Factor loadings for fluency variables (group preferring S3).

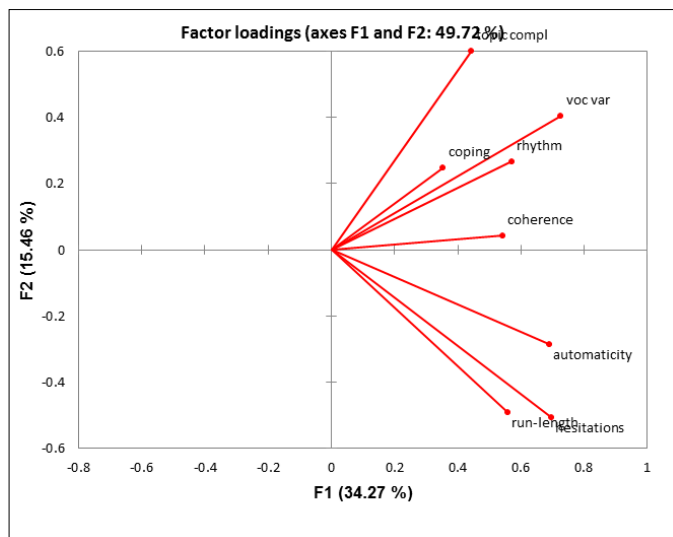


Figure 3. Factor loadings for fluency variables (group preferring S1).

Components of Fluency Data Analyzed by Group, “Preference for S1” and “Preference for S3”

The component ratings were reanalyzed to check for differences between those who selected S1 as the most fluent and those who selected S3. The greatest differences were found for topic complexity, structural complexity, and accent, with those who preferred speaker 3 or rated 3 and 1 the same rating all three components as more important; see Table 4.

Table 4. Differences in Component Rankings (Divided by Preference for S1 or S3)

Component	Rated speaker 1 more fluent (n = 26)		Rated speaker 3 more fluent (n = 13) or equal (n = 3)	
	Mean	SD	Mean	SD
Topic complexity	2.64	0.971	3.63	0.99
Structural complexity	2.87	0.79	3.38	1.37
Accent	2.036	1.085	3.19	1.38

A Mann-Whitney U test was carried out to investigate the statistical significance of these differences, as the data is ordinal. The difference was statistically significant for topic complexity ( $W = 105.0, p = .024$ ) and accent ( $W = 134.5, p = .049$ ) but not for structural complexity ( $W = 114.0, p = .16$ ).

Open-Ended Questions and Responses

The survey included four open-ended questions:

1. What were your reasons for choosing the most fluent speaker?
2. What are the key characteristics of fluent speech?
3. What do you think are the underlying causes that make speech disfluent?
4. To what extent do you think speech fluency is “in the ear of the listener”?

Regarding her open-ended responses, Dore (2016) commented, “Grouping the variables was not straightforward, and some of them could arguably be categorized differently” (p. 34). Dore analyzed her open-ended answers as follows: 64% of responses

Head: *Investigating Teachers' Perceptions of Fluency*

related to “fluency,” 54% related to “phonology,” and 19.8% related to complexity. Given that all questions addressed fluency, using this term as a subcategory is problematic. In addition, the inclusion of acoustic variables (pauses, hesitation) within this category obscured the relative importance of acoustic versus cognitive factors. Therefore, Segalowitz’s (2010) three categories—cognitive, utterance, and perceived fluency—were used here. The following definitions were used:

1. Any acoustically measurable variable was categorized as *utterance fluency*.
2. Pragmatic and cultural factors were interpreted as impacting *perceived fluency*.
3. Comments referring to linguistic knowledge and complexity, coherence, or linking were categorized as related to *cognitive fluency*.

Commonly recurring concepts were grouped together; for example “connecting ideas,” “logical cohesion,” and “using connectors” became “connecting ideas.” Some words such as “natural” were difficult to classify. Given that a “natural” impression could be bolstered by interpersonal behaviors such as gestures, this was categorized as perceived fluency.

### *Analysis of Open-Ended Survey Responses*

Coherence was rated important on the quantitative survey and aspects related to coherence were also mentioned in 15 open-ended answers: connectors, sense groups, and topic development. There was some inconsistency regarding accent. Accent was rated relatively low on the quantitative survey, but intonation was mentioned 16 times. Although intonation is not synonymous with accent, it can arguably be considered a part of the accentual system because it is related to meaning/expressivity (Wells, 2006). Tallies were divided on the basis of which speaker was selected as most fluent (see Tables 5 and 6). The comments of the teachers who selected S1 emphasized cognitive fluency, whereas those preferring S3 made more comments related to utterance fluency.

Table 5. Key Words Derived From Open-Ended Answers of S1 Preference Group, by Category of Fluency

Cognitive		Utterance		Perceived	
Comment	No.	Comment	No.	Comment	No.
Grammatical accuracy	2	Intonation	9	Engagement	1
Coherence	7	Speed/automaticity	5	Easy to understand	3
Connecting ideas	3	Fillers	1		
Longer sense groups	3				
Vocabulary	10				
Complexity	3				
Topic development	5				
Totals	33		15		4

Table 6. Key Words Derived From Open-Ended Answers of S3 Preference Group, by Category of Fluency

Cognitive		Utterance		Perceived	
Comment	No.	Comment	No.	Comment	No.
Grammatical accuracy	2	Pause placement	7	Native-like	3
Longer sense groups	7	Intonation, accent	8	Appropriate	1
Vocabulary	4	Natural	5	Lazy	1
		Speed	5		
Totals	13		25		5

### *Open-Ended Survey Question 3: What Do You Think Are the Underlying Causes That Make Speech Disfluent?*

Most of the responses from teachers in China related to cognitive factors: “insufficient knowledge,” mentioned by 14 (29%), the inability to process language rapidly by 14 (29%), with other common responses including language interference (12 people, 25%). Lack of vocabulary was mentioned by 11 respondents and grammar by three. Pronunciation, intonation, and flat tone (“utterance fluency”) appeared in six answers. Personality,

affective factors, and confidence were considered important by nine respondents (mentioned in 31% of answers). In contrast, in Dore's survey, a much higher proportion mentioned social/affective factors (50%). Lack of practice was mentioned by 21% of Dore's respondents but not by mine.

#### *Open-Ended Survey Question 4: To What Extent Do You Think That Speech Fluency Is "in the Ear of the Listener"?*

Nearly two thirds of respondents (28 out of 43) agreed that perception was very or somewhat important, which was very similar to Dore's participants. The reasons given by those who said fluency was subjective in my survey generally had to do with training, cultural background, and personal preferences of the listener. Only four respondents in my survey (9%) denied fluency was subjective because training and experience created a degree of objectivity. One of these commented, "As a trained teacher we have a very good shared understanding of what is and what isn't fluent." Another said, "There is a subjective element of course, but also a noticeable clustering of opinion."

Cultural familiarity and familiarity with a particular variety of English was mentioned by seven people including, "Expats should be able to understand more than someone who has not been in a language learning environment" and "[Those of the] same culture understand each other."

#### **Discussion**

The wide divergence in the choice of most fluent speaker shows that even within a similar working context there can be a range of opinions about fluency. The component survey indicates some agreement on aspects that are moderately relevant to fluency (repetition, reformulation, and structural complexity share low standard deviation and were rated about 3). The components that drew the highest and lowest ratings were similar to those in Dore's study: Coherence and effortlessness were rated important and accent was least important in terms of averages. Yet the variance for these three was much higher than in Dore's study.

The teachers in China rated grammatical accuracy and structural complexity more highly than those in the UK and Europe. Although this effect did not reach statistical significance, responses to the open-ended questions also suggested that respondents considered fluency to be related to knowledge more than practice. It would be interesting to explore this in a future study.

#### *Research Question 2: Are Speech Ratings Consistent With Declared Beliefs?*

Comparing the open-ended responses of those who selected S1 and those who selected S3, we can see their definitions of fluency overlapped. However, the terms were interpreted differently. Many listeners gave higher ratings to S1, who produced more language but was less native-like than S3. Speaker 1 made her speech longer by joining many simple phrases together with "then." The positive evaluation of the more prolific speaker may reflect acceptance of an Asian variety of English being used as a lingua franca. As one wrote, "It is less important to sound like a native, more important to speak a global variety of English."

Qualities that are strongly related to perceived fluency such as "engagement" were mentioned to justify the selection of S1 versus "lazy" S3. Those who preferred S3 tended to have higher ratings for topic complexity and accent. Transcripts show that S3 manipulated complex sentence structures (which/that) with more density than did S1, despite long pauses. S3 also had a US accent. This suggests that the raters were somewhat consistent with their beliefs.

#### *Research Question 3: Can Components Be Grouped Into Common Factors Underlying Perceived Fluency?*

This question could not be answered definitively. For the group who preferred S3, the two factors identified through factor analysis accounted for just under 60% of the variation; for the whole group the figure was 40.58%. For the group who preferred S1 it was only 33.35%. One reason the component analysis did not account for all the variation may be ambiguity in the phrasing of the survey. Another possible reason is that some components of perceived fluency were not captured in the survey. In particular, the components did not cover interaction fluency, although the speech samples included about one minute of interaction at the end. Personal preferences for aspects such as accent and gender may also be important. Furthermore, those who chose S1 may have done so simply because this speaker talked for longer.

The results for the whole group provide a clearer picture. Topic complexity, accuracy, structural complexity, variety of vocabulary, colloquialness, and other components loaded onto Factor 1 when the analysis was performed for the whole group. Most of these components seem to reflect cognitive fluency.

Factor 2 seems to relate partly to elements of utterance fluency (pausing, speech rate, hesitation). Although Factor 2 also includes some components that are arguably

Head: *Investigating Teachers' Perceptions of Fluency*

manifestations of cognitive fluency (run-length, reformulation), these components are all manifested in sound. Automaticity also loads with these utterance fluency elements in the analysis for the whole group. This accords well with findings by others (De Jong, 2018; Park, 2016). Recent evidence from research into acoustic/temporal aspects of fluency suggests there are threshold levels for acoustic factors such as speech rate and length or number of pauses, which need to be achieved if a speaker is to be considered fluent (Park, 2016). Park found that the ratio of unexpected pauses to total number of pauses provided a reliable measure of fluency. As mentioned, De Jong (2018) claimed that acoustic factors correlate well with overall fluency and global speaking proficiency.

### Conclusion

The survey results in this study were similar to Dore (2016), placing a relatively high value on coherence and effortlessness and a low value on accent. However, there was a fundamental divergence between teachers who preferred a smoother but more prepared performance and those who preferred a more spontaneous but shorter one. This difference was reflected both in speaker ratings and in the components of fluency that were considered most important. There were bigger differences between members of the group in China than between groups in the UK and Italy.

Adding factor analysis to Dore's study showed that components relating to utterance fluency were generally associated with automaticity by the China-based respondents. In future research, Dore's component list could be revised to include intelligibility. Utterance components rated similarly (such as fillers and reformulation) could be consolidated. Further investigations into perceptions of fluency are needed in order to promote shared understandings between students, teachers, employers, and agencies responsible for evaluation, particularly in relation to English as a lingua franca.

### Bio Data

**Ellen M. Head** has worked at Miyazaki International College since 2017. Her research interests include learner autonomy inside and outside the classroom and the use of the CEFR in Japan. <ehhead@sky.miyazaki-mic.ac.jp>

### References

Cheng, L. (2008). The key to success: English testing in China. *Language Testing*, 25(1), 15-37. <https://doi.org/10.1177/0265532207083743>

- De Jong, N. H. (2018). Fluency in second language testing: Insights from different disciplines. *Language Assessment Quarterly*, 15(3), 237-254. <https://doi.org/10.1080/15434303.2018.1477780>
- Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thompson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language Learning*, 54(4), 655-679. <https://doi.org/10.1111/j.1467-9922.2004.00282.x>
- Dore, C. (2016). Perceptions of fluency (MA thesis, Winner of British Council Masters Award). Retrieved from [https://englishagenda.britishcouncil.org/sites/default/files/attachments/dissertation\\_design\\_for\\_publication\\_2016\\_reading\\_university\\_cecilia\\_dore.pdf](https://englishagenda.britishcouncil.org/sites/default/files/attachments/dissertation_design_for_publication_2016_reading_university_cecilia_dore.pdf)
- Eight private English tests accepted for future university admission system (2018, March 27). *The Mainichi*. Retrieved from <https://mainichi.jp/english/articles/20180327/p2a/00m/0na/013000c>
- Kormos, J., & Denes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32(2), 145-164. <https://doi.org/10.1016/j.system.2004.01.001>
- Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40, 387-417. <https://doi.org/10.1111/j.1467-1770.1990.tb00669.x>
- Murphey, T. (2011). *The tail that wags*. Japan: Perceptia Press.
- Muller, T., Adamson, J., Herder, S., & Brown, P. (2014). *Exploring EFL fluency in Asia*. <https://doi.org/10.1057/9781137449405>
- Park, S. (2016). Measuring fluency: Temporal variables and pausing patterns in L2 English speech. *Open Access Dissertations*, 692. [https://docs.lib.purdue.edu/open\\_access\\_dissertations/692](https://docs.lib.purdue.edu/open_access_dissertations/692)
- Saito, K., Trofimovich, P. & Isaacs, T. (2017). Using listener judgments to investigate linguistic influences on L2 comprehensibility and accentedness: A validation study. *Applied Linguistics*, 38(4), 439-462. <http://dx.doi.org/10.1093/applin/amv047>
- Saville, N. (2017, June). *Setting and monitoring standards within a professional community of practice*. Paper presented at Academic Forum on English Language Testing in Asia (AFELTA), Taipei, Taiwan.
- Segalowitz, N. (2010). *The cognitive bases of second language fluency*. <https://doi.org/10.4324/9780203851357>
- Van Moere, A. (2012). A psycholinguistic approach to oral language assessment. *Language Testing*, 29, 325-344. <https://doi.org/10.1177/0265532211424478>
- Wells, J. C. (2006). *English intonation*. Cambridge, England: Cambridge University Press.
- Yu, Q. (2014). Various items causing IELTS test-taker's low performance in mainland China: An international joint education program solution. In *Proceedings of the 14th International Conference on Global Economy, Finance and Humanities Research*. <https://doi.org/10.2991/gefhr-14.2014.7>
- XLSTAT [Computer software]. (2019). New York, NY: Addinsoft.
- Zheng, Y., & Cheng, L. (2008). Test review: College English Test (CET) in China. *Language Testing*, 25(3), 408-417. <https://doi.org/10.1177/0265532208092433>



## Appendix A

### Replication Study and Survey: Fluency and Perceived Fluency

Please listen to the three samples of speakers telling a picture story.

How fluent is each speaker on a scale of 1 to 7 where 7 means very fluent, and 1 means not at all fluent.

1. Sample 1
2. Sample 2
3. Sample 3.
4. What were your reasons for choosing the most fluent speaker?
5. What are they key characteristics of fluent speech?
6. What do you think are the underlying causes that make speech disfluent?
7. To what extent do you think that speech fluency is “in the ear of the listener”?

Questions 8 to 20 refer to the following question:

*What aspects of fluency do people pay attention to when rating nonnative speaker fluency?*

Please rate the importance of each aspect out of 5. 5 means “very important to fluency” and “1” means “not important.” Even if the aspect is a negative one, such as “hesitation,” you should give it 5 if you think it is important.

	1	2	3	4	5
Number of hesitations (few hesitations = fluent)					
Length of runs (how long do they keep going without a pause)					
Complexity of topic content					
Variety of vocabulary					
Native-like rhythm					
Coherence					
Coping in social situations					
Automaticity (words appear to come to mind automatically)					

Complexity of structures					
Accent					
Global proficiency					
Using formulaic sequences					
Colloquialness					
Pausing: length of pauses, number of pauses					
Effortlessness (can she/he speak without effort?)					
Rate of speech					
Accuracy (grammatical accuracy)					
Repetition (she repeats herself to gain thinking time)					
Using fillers appropriately					
Reformulation (she repeats the same meaning in different words as if struggling to find the right word)					

Please add any comments you wish. Thank you.

## Appendix B

### Transcription

#### Speaker 1

- Total time: 2.28 minutes
- Story: 100 seconds; 207 words
- Rate: 2 words per second

John works in an office. / He usually works from Monday to Friday / from 9 am to 5 pm. / He doesn't like his job very much./ It is Saturday morning. It is 7 o'clock. The alarm clock wakes him up./ He feels very tired because he watched the football game /

Head: *Investigating Teachers' Perceptions of Fluency*

until midnight the day before./ Then he went to the dining room, had some breakfast. He has coffee and bread and yogurt and some fruit./ Then he went to his ^a study./ He turned the computer on and watched the lotteries news. Then he made some phone calls to order the lottery. And then he dressed up and drive a car to the lottery office and ^a bought the lottery that he ordered from the telephone turned the computer on and ^a watched the lotteries news. Then he went to the gym, did some exercise / and then meet some friends./ At night / round about 8 o'clock he arrived a home/ waiting for the news./ Wow. He won the lottery prize, he was SO excited. /The following day, he went to the beach, thinking of changing/ er a new job, maybe bigger apartment, maybe travelling around the world./ He is very happy for this change.

Interviewer: What would you do if you won the lottery?

If I won the lottery/ I think I will go to work in the country/ to build a new school and buy lots of books and dictionaries/ for the children there/ so that they ^ every children could go to school. /Because in the country some children quit/ er from going to school because ^ er they are not yet rich. They have to help the family work / in the farm.

*Speaker 2*

- Total time: 2 minutes 50 seconds
- Story: 45 seconds; 57 words
- Rate: 1.267 words per second

Mr. Wang wakes up in half-past 7. / After that he has breakfast and leaves home at about 8 o'clock./ He works very busy the whole day./ And usu'^ he usually will buy a ticket in the lottery./ One day/ "Takarakuji o atarimashita." /So he get a lot of money/ and he is planning to have a one week holiday in Hawaii./ An' he did it.

Interviewer: What would you do if you won the lottery?

What ^ would I do? Maybe I will want to buy a house in countryside, and/ will go for a trip/ with my parents,/ because we don't have time to stay/ with each ^ each other very often/ so I think I will do that.

*Speaker 3*

- Total time: 1 minute 25 seconds
- Story: 78 words; 49 seconds
- Rate: 1.51 words per second
- Follow-up: 47 words

At 8.30 he woke up/ and / he got dressed up and at 8.45 he finished his breakfast and he ~went out for to work. / After 8 hours he go to the lottery and buy a ticket./ And at ^^8. 30 pm he ^ was sitting on his couch ^ couch and watching TV and^ was hoping that his ticket that he won, which he did /, and then / he took a vacation to the beach.

Interviewer: If you won the lottery what would you do?

I would buy new things to replace the old ones.

Interviewer: What kind of things?

Like computers or TVs.

Interviewer: More than one?

Yeah.

Interviewer: Why?

Cos some of them is like very slow and,/ like, their battery are very low/ so if you, like, the new ones, can ^ um make it faster, / stronger.