Fluency measures and L2 speaking proficiency levels

May Y. Wu, Rasmus Steinkrauss, and Wander Lowie
University of Groningen

The relationship between automatic fluency measures, including speed, breakdown and repair measures (e.g., articulation rate, silent pause ratio, the number of repairs per clause), and overall proficiency levels has gained rising interest. So far, only a few studies have investigated which fluency measures can distinguish between proficiency levels (e.g., Tavakoli, Nakatsuhara, & Hunter., 2020; Tavakoli, Kendon, Mazhumaya, et al., 2023), and found that several temporal fluency measures (e.g., articulation rate, speech rate) and some breakdown fluency measures (e.g., frequency of mid-clause/end-clause pauses, length of mid-clause/end-clause pauses) can do so but repair fluency measures cannot.

The current study advances this line of research by testing which fluency measures can distinguish between IELTS speaking proficiency levels (Understanding and explaining IELTS scores n.d.). Using a data set taken from a larger project investigating the reliability of commonly used automatic complexity, accuracy and fluency (CAF) measures in L2 English oral data (Wu et al., 2023), we examined if the 27 automatic fluency measures obtained by Praat (de Jong et al. 2021) and CLAN (MacWhinney 2000) would differ across IELTS band levels 5.0, 5.5, 6.0 and 6.5 (CEFR higher B1 to B2 levels) in the IELTS Speaking Test of 49 L1 Chinese learners of English. These band scores were based on speaking samples collected from a mock test using published IELTS Speaking Test materials (e.g., Cambridge University Press, 2016), which were rated by two expert raters who had over ten years of experience in IELTS training.

Our analysis (see Table 1) showed that two temporal and two breakdown fluency measures (articulation rate, length of run, silent pause ratio, and mid-clause silent pause ratio) could differentiate between lower IELTS levels (5.0 and 5.5) and the highest level (6.5) (c.f. Tavakoli, Kendon, Mazhumaya, et al., 2023). For the remaining measures, no comparisons between levels reached significance (effect sizes ranging from .01 to .21). Overall, none of the measures can differentiate neighbouring IELTS proficiency levels.

Keywords: fluency; assessing fluency; L2 speaking; proficiency levels; IELTS

Reference

Cambridge University Press (2016) Cambridge IELTS 11 Academic Student's Book with Answers. Cambridge: Cambridge University Press.

De Jong, N. H., Pacilly, J., & Heeren, W. (2021). PRAAT scripts to measure speed fluency and breakdown fluency in speech automatically. *Assessment in Education: Principles, Policy & Practice* 28(4). Routledge: 456–476.

MacWhinney B (2000) *The CHILDES Project: Tools for Analyzing Talk*. Mahwah: NJ: Lawrence Erlbaum Associates.

Tavakoli, P., Nakatsuhara, F. & Hunter, A-M. (2020). Aspects of fluency across assessed levels of speaking proficiency. *Modern Language Journal*, 104(1), 169–191.

Tavakoli, P., Kendon, G., Mazhurnaya, S. & Ziomek, A. (2023) Assessment of fluency in the Test of English for Educational Purposes. *Language Testing*, 40(3), 607–629.

Understanding and explaining IELTS scores (n.d.). Retrieved March 25, 2024, from https://takeielts.britishcouncil.org/teach-ielts/test-information/ielts-scores-explained

Wu, M. Y., Steinkrauss, R. & Lowie, W. (2023, July 17–20). *The reliability of quantitative complexity, accuracy and fluency measures in L2 English oral production*. The 20th World Congress of International Association of Applied Linguistics (AILA), Lyon, France. https://aila2023.dryfta.com/

Table 1

Fluency measures that can differentiate IELTS levels

Measures	*p*-value	Effect size	Significant group(s)
Articulation rate (w)	0.000895*	0.30	6.5 - 5.0
Mid-clause silent pause ratio	0.0014*	0.29	6.5 - 5.0; 6.5 - 5.5
Length of run (w)	0.00156*	0.29	6.5 - 5.0
Silent pause ratio	0.00201*	0.28	6.5 - 5.0
Speech rate (w)	0.0127	0.21	-
The ratio of all pauses (w)	0.0166	0.20	-
Length of run (syll)	0.0237	0.19	-
Length of mid-clause silent pauses	0.0239	0.19	-
Speech rate (syll)	0.0375	0.17	-
Length of mid-clause pauses	0.0399	0.17	-
Phonation time ratio	0.0422	0.16	-
Mid-clause silent pause frequency	0.0435	0.16	-

Note. The significant level for the ANOVA test was set as 0.01, following Tavakoli, Kendon, Mazhumaya, et al. (2023). This table also reports the *p*-values and effect sizes of the eight measures falling within the *p*-value range of .01 to .05. However, the significant groups that can be differentiated by the eight measures are not included as they did not meet the significance threshold set by our study.