Bypassing Transcription: Efficacy of a Tally-Based System for Disfluency Analysis of Preschool Children Who Stutter

Jean SAWYER
Illinois State University, Normal, Illinois
jsawyer@ilstu.edu

1. Introduction/Purpose

Efforts to quantify stuttered speech have generally involved two different directions. The first relies on the categorical judgment of disfluent events as stuttered or not stuttered, for example, the number of stuttered words per 100 words or syllables (Ingham & Riley, 1998; Onslow et al., 1990). One of the weaknesses of this approach is the considerable level of listener disagreement about what is perceived as stuttering (Martin & Haroldson,1981).

The second approach to quantifying stuttering is a descriptive analysis, which involves identification of *a priori* defined speech characteristics, or disfluencies. Yairi and Ambrose (1999) have delineated two categories of disfluencies, which are stuttering-like disfluencies, and other disfluencies. Stuttering-like disfluencies are comprised of part- and single-syllable word repetitions and dysrhythmic phonation, which include prolongations and blocks. Other disfluencies include interjections, revisions, and phrase repetitions/multisyllabic word repetitions. Both stuttering-like and other disfluencies occur in anyone's speech, but stuttering-like disfluencies characterize the speech of people who stutter. Advantages of counting disfluencies are that it eliminates the demand of making perceptual judgments of stuttering and can differentiate children who stutter from typically fluent children (Yairi & Ambrose, 2005). Additionally, it can be a prognostic indicator for continued stuttering (Yairi & Seery, 2022). Close to the onset of stuttering, disfluencies cannot differentiate children who will develop chronic stuttering from those who will recover, but a downward trend in stuttering-like disfluencies over the first year of stuttering is seen as a "strong sign for eventual recovery" (Yairi & Seery, 2011, p. 364).

Disfluency analysis has been criticized in part because there have been several different schema of disfluencies described (Einarsdóttir & Ingham, 2005). Ingham and colleagues (2012) have argued that because not all disfluencies are stuttered, measures of stuttering, rather than disfluencies, are more appropriate to measure the stuttering disorder. Additionally, disfluency analysis is time-consuming, as the listener needs to categorize the disfluencies, in addition to counting them (Bothe, 2008).

Disfluency analysis involves coding each type of disfluency produced. Transcribing the speech, a time-consuming process, makes the coding easier, but disfluency analysis can be done in real time. Yaruss et al. (1998) described a real-time analysis that delineated two types of disfluencies, "more typical" and "less typical." Several clinicians produced transcripts from 50 short speech samples of 200 words in length. One person did a real-time analysis of the samples, and the authors reported no significant differences found between transcript-based and real-time analysis. The authors did not, however, examine a more nuanced profile of disfluencies.

To date, no reliability studies exist for off-line disfluency analysis. The goal of the current study was to determine the efficacy of a tally-based method to calculate frequency and type of disfluencies in the stuttered speech of preschool children.

2. Method

A graduate student working with the author completed 30 hours of supervised training in disfluency analysis, transcribing the speech of adults and children who stutter. She subsequently analyzed 47 samples of preschool children who stuttered using the tally-based method. The speech samples for these children had previously been collected and transcribed, and each sample was collected during a 15-minute play session with the child and an adult caregiver (Sawyer et al., 2017). The student listened to a speech sample and marked the disfluency types and calculated the total of each type. Stuttering-like disfluencies (SLD) included part-word repetitions (PW), single-syllable word repetitions (WW), and dysrhythmic phonation (DP). Other disfluencies (OD) were also calculated, and included revisions (R), phrase repetitions and multi-syllabic word repetitions (MP), and interjections (I). The student also calculated the average length of repetitions (repetition units; RU) while marking disfluency types. The sample

was listened to a second time to determine the number of syllables. Severity was calculated using a weighted scale (Yairi & Ambrose, 1999), which used the disfluency calculations tabulated. This weighted scale gives more weight to units of repetition and dysrhythmic phonation, which perceptually are more likely to indicate stuttering than other measures. Results calculated using the tally-based method were compared to the previously analyzed transcripts of the children's speech. The reliability of the transcript-based analysis across examiners had previously been determined (Cohen's Kappa at .91 for interrater reliability, Sawyer et al., 2017).
The difference between speech samples analyzed using the tally-based method and the transcript method were determined by calculating mean differences and using paired samples t-tests, along with the Wilcoxon's signed rank test, as the samples were not normally distributed.

3. Results

The tally-based method did not differ from the transcript-based method in more global-based measures of means for SLD counts (p=.065), nor in the average length of repetition (p=.154). OD means were different (p=000), as were the mean number of syllables counted (p=.000). Table 1 shows the results for SLD, OD, RU, and syllable counts.

Table 1. Comparison of tally- and transcript-based calculations for SLD, OD, RU, and Syllable Counts

	SLD tally	SLD transcript	OD tally	OD transcript	RU tally	RU transcript	Syllables tally	Syllables transcript
Mean	6.01	6.39	5.99*	5.38*	1.41	1.46	555.17	651.68
SD	5.87	4.34	2.57	1.80	0.32	0.89	174.09	169.34

Notes: * = statistically significant, $p < .05$; SLD is stuttering-like disfluency, OD is other disfluency, and RU is repetition units

Examining individual disfluencies, the tally-based method did not differ from the transcript-based method in counts of part-word repetitions (p=.051), single-syllable word repetitions ($p > .05$), and multisyllabic word/phrase repetitions ($p > .05$). The means for those variables can be found in Table 2.

Table 2. Comparison of tally- and transcript-based calculations for PW, WW, and MP Counts

	PW tally	PW transcript	WW tally	WW transcript	MP tally	MP transcript
Mean	1.94	2.21	2.74	2.47	1.14	.087
SD	1.66	1.44	2.53	2.78	1.31	0.61

Note: PW is part-word repetition, WW is single-syllable word repetition, and MP is multisyllabic word and/or phrase repetition

There were some individual disfluency differences between the two methods. The methods yielded means that were significantly different for dishryhmic phonation (p=.043), interjections (p=.003), and revisions (p<.001. The means for those disfluency types are found in Table 3.

Table 3. Comparison of tally- and transcript-based calculations for DP, I, and R Counts

	DP tally	DP transcript	I tally	I transcript	R tally	R transcript
Mean	1.38*	1.62*	1.69*	1.89*	1.35*	2.40*
SD	2.33	1.65	1.98	1.33	0.76	0.93

*= statistically significant, $p < .05$

Note: DP is dysrhythmic phonation, I is interjection, and R is revision

The severity of the speech samples was assessed using the weighted scale. Most of the speech samples, 38/47, showed agreement between the transcript- and tally-based methods. Twenty-six of these were rated mild, 11 moderate, and 1 severe. There were nine disagreements. Five of the tally-based samples were rated mild, but these were rated moderate with the transcript-based method. Three of the tally-based samples were rated moderate, and

these were rated mild with the transcript-based method. Finally, one tally-based sample was rated as severe using the tally-based method and moderate with the transcript-based method. The mean weighed scale for the tally-based method was 10.14 (SD=13.21), and for the transcript-based method it was 9.83 (SD 7.71). The difference between these means was not statistically significant (p=.184).

4. Discussion

The tally-based system was successful calculating global totals of stuttering-like disfluencies, as there were no statistically significant differences between the tally-based system and the transcript-based system in total SLD. It was not as accurate as the transcript-based method in calculating individual disfluencies that comprise SLD, however, as there were significant differences found in calculations of dysrhythmic phonation (DP). DP play a role in assessing severity of stuttering in the weighted scale, and the weighted scale differences in the two methods were not statistically significant. The tally-based system appears to be successful for more global measures of disfluency, such as SLD counts and calculations of severity using the weighted scale.

The tally-based system was less accurate when calculating other disfluencies and syllable counts. The student coder calculated 21% fewer other disfluencies than were calculated via transcription. Note that, because other disfluencies are not indicators of stuttered speech, that measure has less importance than the measurement of stuttering-like disfluencies for clinical work (Conture, 2001; Yairi, 1996). In addition, the student coder calculated 15% fewer syllables by tally than were calculated by transcription. It may have been the case that syllables were not as salient to a coder as disfluencies were, resulting in fewer syllables calculated by tally.

The aspects of speech that are important in assessing stuttering, however, were similar across methods. In addition to collecting similar counts of stuttering-like disfluencies, the tally-based system matched that of the transcript-based system in calculating repetition units, so severity of stuttering based on the weighted scale was similar across both methods.

This was the first study to examine specific disfluency types in non-transcript based coding, and the accuracy of the tally-based system as compared to the transcript-based system was mixed. The tally-based system compared favorably to the transcript-based system in calculating part-word and single-syllable word repetitions, and multisyllabic word and phrase repetitions, but was not as accurate in determining dysrhythmic phonation, interjections, and revisions.

5. Conclusion

Disfluency analysis based on transcription is a time-consuming process. The current study has shown that a tally-based system is a viable way to calculate stuttering-like disfluencies and assess severity using the weighted scale.

Future studies may want to examine the effectiveness of the method using more coders. It may also help to do more training in calculating syllable counts and in identifying disrhythmic phonation, interjections and revisions. More training may help improve the accuracy of the tally-based method.

Disclosures

The author has no financial disclosures regarding this manuscript. Her nonfinancial disclosure is that she is an employee of Illinois State University and receives a salary from that organization.

References

Bothe, A. K, (2008). Identification of children's stuttered and nonstuttered speech by highly experienced judges: binary judgments and comparisons with disfluency-types definitions. *Journal of Speech, Language, and Hearing Research, 51*, 867-878.

Conture, E. (2001). *Stuttering: Its nature, diagnosis, and treatment.* Needham Heights, MA: Alllyn & Bacon.

Einarsdóttir, J. & Ingham, R, J. (2005). Have disfluency-type measures contributed to the understanding and treatment of developmental stuttering? *American Journal of Speech-Language Pathology, 14*, 260-273.

Ingham, R. J., Ingham, J. C., & Bothe, A. K. (2012). Integrating functional measures ith treatment: A tactic for enhancing personally significant change in the treatment of adults and adolescents who stutter. *American Journal of speech-Language Pathology, 21*, 464-277.

Ingham, J. C., & Riley, G. (1998). Guidelines for documentation of treatment efficacy for young children who stutter. *Journal of Speech Language, Nd Hearing Research, 41*, 753-770.

Martin, R., & Haroldson, S. R. (1981). Stuttering identification: Standard definition and moment of stuttering. *Journal of Speech and Hearing Research, 15*, 743-752.

Onslow, M., Costa, L., & Rue, S. (1990). Direct early intervention with stuttering: Some preliminary data. *Journal of Speech and Hearing Disorders, 35*, 405-416.

Sawyer, J., Matteson, C., Ou, H., & Nagase, T. (2017). The effects of parent-focused slow relaxed speech intervention on articulation rate, response time latency, and fluency in preschool children who stutter. *Journal of Speech, Language, and Hearing Research, 60*, 794-809.

Yairi, E. (1996). Applications of disfluencies in measurements of stuttering. *Journal of Speech and Hearing Research, 39*, 402-104.

Yairi, E., & Ambrose, N. (1999). Early childhood stuttering: I. Persistency and recovery rates. *Journal of Speech, Language, and Hearing Research, 42*, 1097-1112.

Yairi, E., & Ambrose, N. G. (2005). *Early childhood stuttering*. Austin: Pro-ed.

Yairi, E., & Seery, C. H. (2015). *Stuttering: Foundations and clinical applications*. Boston, MA: Pearson Allyn & Bacon.

Yairi, E., & Seery, C. H. (2022). *Stuttering: Foundations and clinical applications* (3rd ed.). Upper Saddle River, NJ: Pearson.

Yaruss, J. S., Max, M., Newman, R., & Campbell, J. (1998). Comparing real-time and transcript-based techniques for measuring stuttering. *Journal of Fluency Disorders, 23*, 137-151.