A Taxonomy of CBM Oral Reading Coding Errors: Threats to Assessment Fidelity and Data-based Decision Making

Purpose

Given the continued importance of curriculum-based measures of reading (CBM-R) to data-based decision making, this study developed a taxonomy of examiner error types to identify potential avenues for improving the use and interpretation of individually administered reading assessments.

Methods

Three different groups of raters scored the students' CBM-Rs. Group 1 consisted of one doctoral and two Master's students who administered the test at the students' schools, so these were referred to as the *naturalistic raters*. Three research assistants at the second author's institution made up the second group of raters (Group 2). These individuals had at least some prior experience administering and scoring CBM-R. They were referred to as the *expert raters* due to their extensive training and reliability calibration. Group 3 consisted of two Master's students at the first author's institution, each with two years' prior experience administering CBM-R. These individuals were trained to be the *gold standard raters*.

Scoring, determining rater errors, and classifying errors were all conducted while the gold standard raters were working with a given passage reading to ensure careful consideration of all aspects. The two raters worked 30 hours per week for 12 weeks, so an average of 3.33 hours were spent analyzing a single, one-minute passage reading of the 108 unique readings in the dataset.

Results

Across four separate ratings for each of 108 passage readings completed by students in Grades 5 and 6, a total of 929 examiner errors were identified (range = 1-37 errors per student; M = 8.52). A micro-analysis of the errors resulted in a taxonomy of 37 unique error types clustered into 10 categories. In 47% of the passage readings, examiners made errors identifying the total words read. The highest number of combined errors occurred in the category for scoring mispronunciations of words (n = 417 or 45% of the total errors). In 98% of the passages, raters disagreed on the number of words read correctly—the score used for data-based decision making.

Next Steps

- Extend our taxonomy of rater errors by (a) including those not based on counting and (b) distinguishing between rater errors and rater effects
- Determine the strength of the relationship between examiner, student, and school context predictor variables and differential rater function over time (DRIFT)
- Test a conceptual model of DRIFT and its impact on data-based decisions about student placement

Overall Frequency of Error Types by Category

	Error Type	Possible Occurrence Per Passage	Number of Add Errors	Number of Omit Errors	TWR Errors	Total Errors
Test Development: Beginning the Passage (Total Errors in Category = 40)						
1	Student Restarts	S	0	0	5	5
2	Read Title	S	0	0	35	35
Test Development: Word Choices (Total Errors in Category = 44)						
3	Repeated Words Around SC	M	0	3	0	3
4	Rare Word	M	29	11	0	40
5	Hyphenated Word	M	1	0	0	1
Test Development: TWR Determination (Total Errors in Category = 94)						
6	Last Words	S	7	10	81	98
Examiner: TWR Determination (Total Errors in Category = 84)						
7	Unclear When Timer Stops	S	0	0	65	65
8	Examiner Stops Early	S	0	0	15	15
Examiner: Scoring Attempted Self-corrects (Total Errors in Category = 14)						
9	Error on SC	M	0	7	0	7
10	Error Then SC	M	7	0	0	7
Examiner: Scoring Clarifications of Words and Hesitations (Total Errors in Category = 2)						
11	Student Clarifying Word After Reading	M	1	0	0	1
12	Examiner Gives the Word	M	0	1	0	1
Examiner: Nonstandard Pronunciations (Total Errors in Category = 119)						
13	Dialect/Colloquialism	M	9	0	0	9
14	Accent	M	16	3	0	19
15	Stutter	M	5	0	0	5
16	Drawing Out Pronunciation	M	1	5	0	6
17	Pause During Word	M	4	0	0	4
18	Unclear Pronunciation	M	53	2	0	55
19	Words Read Quickly	M	21	0	0	21
Examiner: Incorrect Pronunciations (Total Errors in Category = 417)						
20	Adding –s Ending	M	0	15	0	15
21	Dropping –s Ending	M	0	59	0	59
22	Dropping –ed ending	M	0	7	0	7
23	Derivation of Word	M	0	19	0	19
24	Formed Contraction	M	0	20	0	20
25	Replaced With Other Word	M	0	141	0	141
26	Letter Reversal	M	0	4	0	4
27	Replaced Syllable	M	0	97	0	97
28	Add Phoneme	M	0	4	0	4
29	Left Out Phoneme/Syllable	M	0	9	0	9
30	Saying the Silent Consonant	M	0	14	0	14
31	Alterations of Numbers	M	0	4	0	4
32	Changed Emphasis	M	0	24	0	24
Examiner: Whole Word Insertions and Deletions (Total Errors in Category = 70)						
33	Insert Word	M	4	0	0	4
34	Skipped Word	M	0	64	0	64
35	Changing Order	M	0	2	0	2
Examiner: Other Scoring Errors (Total Errors in Category = 45)						
36	Unclassifiable	M	19	0	0	19
37	Around Error	M	20	6	0	26
	Total Error Counts		197	531	201	929

Deborah K. Reed, School of Teacher Education and Florida Center for Reading Research, dkreed@fcrr.org