The Plateau of Oral Reading Fluency Growth: An Initial Recommendation When to Stop Assessing

Chalie Patarapichayatham

Joseph F. T. Nese

Leilani Sáez

University of Oregon

Corresponding Author:
Chalie Patarapichayatham, PhD.
College of Education, BRT,
5262 University of Oregon,
Eugene, OR 97403-5262
Email: chalie@uoregon.edu

Abstract

The use of oral reading fluency (ORF) as a predominant measurement tool for identifying struggling readers has grown exponentially over the past 30 years. ORF growth using curriculum-based measures has also become an important practical and empirical issue influencing the field. Although fluency scores have shown to be reliable predictors of reading performance up through middle school for poor readers, the point at which assessments of ORF fail to provide additional information concerning reading proficiency for typical readers, in general, is currently unknown. The purpose of our paper was to determine a range of fluency scores at which ORF growth plateaus as a potential guide to educators and researchers for ceasing ORF assessments. A sample of 89,465 students was selected to include all Grade 3 – 6 students from 2009 – 2010 and 2011 – 2012. Results showed that the growth trajectory sharply increased in Grades 3 – 5, whereas the growth trajectory only slightly increased from Grade 5 to Grade 6, indicating ORF tests may be less informative as reading skills more fully develop. The range of ORF plateau score range was 137.72 – 170.94 words read correctly per minute (wcpm). It is suggested that student ORF is qualified once scores reach this range.

Introduction

Use of the construct of oral reading fluency (ORF) as a predominant measurement tool for identifying struggling readers has grown exponentially over the past 30 years (e.g., Baker, Smolkowski, Katz, Hank, Seeley, Kame'enui, et al., 2008; Christ, Silberglitt, Yeo, & Cormier, 2010; Deno, Marston, & Tindal, 1985). Some researchers have argued that tests of ORF provide an indication of readers' level of basic competency in elementary school, but may be less informative as reading skills more fully develop (e.g., Fuchs, Fuchs, Hosp, & Jenkins, 2001; Yovanoff, Duesbery, Alonzo, & Tindal, 2005). Strong and stable relations between ORF and other standardized reading achievement measures across Grades 1-6 have been found across studies (e.g., Good, Simmons, & Kame'enui, 2001; Hosp & Fuchs, 2005; Reschly, Busch, Betts, Deno, & Long, 2009; Wood, 2006); however, the relation between ORF and reading comprehension diminishes between Grades 5-8 (e.g., Yovanoff, et al., 2005), as comprehension becomes more important to reading proficiency than decoding skills in these later grades.

In practice, ORF test results are typically used as a decision-making tool for classifying students into different levels of risk for reading difficulties, and are considered a global index of reading proficiency (e.g., Stecker, Lembke, & Foegen, 2008). Fluency scores are also used to inform teachers about the effectiveness of reading interventions through the progress monitoring of reading growth (e.g., Hudson, Lane, & Pullen, 2005). Becoming a fluent reader is an important goal because the fluent reader can recognize the printed word with ease and speed, utilizing few cognitive processes (e.g., La Berge & Samuels, 1974). This facilitates simultaneous word recognition and comprehension while making meaning from connected text (e.g., NICHD, 2000).

Increases in the use of Response to Intervention (RTI) models in schools have led to more systematic ORF assessment (e.g., Shapiro, et al., 2006; Speece, Case, & Molloy, 2003), in which students may be assessed three times a year (i.e., benchmark testing for screening risk status) or more often (i.e., progress monitoring for evaluation of instruction and student growth). Growth in reading fluency using curriculum-based measures of ORF has become both an important practical and empirical issue influencing the field (e.g., Christ, et al., 2010; Deno, Fuchs, Marston, & Shin, 2001; Fuchs, Fuchs, Hamlett, Walz, & Germann, 1993; Hasbrouck & Tindal, 2006; Nese, Biancarosa, Anderson, Lai, & Tindal, 2012). Although fluency scores have shown to be reliable predictors of reading up through middle school for poor readers (e.g., Denton, Barth, Fletcher, Wexler, Vaughn, Cirino, et al., 2011), the point at which assessments of reading fluency fail to provide additional information concerning reading ability, in general, is currently unknown.

Studies of reading fluency among typical readers suggest weaker predictive utility as reading skill develops (e.g., Nese, Park, Alonzo, & Tindal, 2011; Wayman, et al., 2007; Yovanoff, et al., 2005). Particularly for RTI practice, in which ORF measurement is the primary method for quickly evaluating reading skill, it would be worthwhile to investigate the possibility of an ORF plateau to signal a stopping point for administering ORF screening assessments. In other words, ORF performance beyond this plateau would signify that student ORF ability is qualified, and that screening or monitoring performance on other reading measures may be more informative for school practice.

Oral reading fluency is the academic construct most often assessed as part of an RTI model and most often studied by researchers (e.g., Shapiro, Keller, Lutz, Santoro, & Hintze, 2006; Speece et al., 2003; Wayman, Wallace, Wiley, Tichá, & Espin, 2007; Wood, 2006). The purpose of our study was to determine the range of fluency scores at which ORF growth plateaus as a potential guide to educators and researchers to cease assessing ORF. In seeking such a plateau here, we referred to a range of words read correctly per minute (wcpm), not a particular grade level, as we anticipate that students will arrive at this number at different points in maturational development.

This study has the potential to be useful across educational stake-holders. For educators administering ORF benchmark assessments as part of an RTI model, an ORF plateau range may save instructional time and school resources. This would result in more meaningful assessments for students, who would not be subject to redundant ORF assessments and could move to more appropriate reading tests such as comprehension. And this effort could potentially move researchers toward similarly applied studies of within- and across-year student growth.

Methods

Sample

The sample included 6 cohorts of students across grades 3 – 6 drawn from the easyCBM® assessment system, an online benchmarking and progress monitoring system used as part of an RTI framework across the United States. The full assessment system contains a variety of K – 8 reading measures; for this study, we focused on the oral reading fluency assessment only. A convenience sample of 89,465 students was selected to include all grade 3 – 6 students from 2009 – 2010 and 2011 – 2012, so that the sample combined 6 cohorts, making it longitudinal and cross-sectional. Three within-year testing occasions (i.e., fall, winter, and spring) were used at each of the four grades, for a total of 12 testing occasions (time points).

Measure

The easyCBM® ORF measures of grades 3 – 6 were used in this current study. The easyCBM® ORF measures, approximately 250 words in length, were written according to word count and grade-level guidelines and reviewed by researchers as well as experienced teachers. The ORF measures are individually-administered by classroom teachers and instructional assistants trained in standardized test administration. On these measures, students read aloud for 60 seconds from a grade-level appropriate original work of narrative fiction while test administrators followed along on their own copy of the material. Self-corrections were counted as correctly read words, and any word a student skipped or read incorrectly is counted as an error. After one minute the assessor calculated the total number of words read correctly to arrive at the student's score, words read correctly per minute (wcpm).

Analysis

To determine the ORF plateau range, we began by determining the functional form that best fits our data, either linear growth model or quadratic growth model. The Deviance (DIC) was used for selecting the best model, in which a smaller value indicates a better fit. All parameters were estimated under a structural equation modeling framework using the M*plus* 7.0 software (Muthén & Muthén, 1998 – 2012) using the Bayesian estimator. With relatively large sample sizes and 12 time points across four years, it might be too heavy for maximum likelihood estimator to obtain reasonable results. Also, the maximum likelihood estimator usually takes longer computation time than a Bayesian estimator (e.g., Muthén, 2010). We decided to use a Bayesian in our study. Based on previous research that has demonstrated the nonlinear trajectory of ORF (e.g., Christ, et al., 2010; Nese et al., 2012), we expected that a quadratic growth model would best fit the data.

The fixed effects (e.g., intercept, linear, and quadratic means) of the best model were then used to compute the plateau of the ORF trajectory using the following formula:

$$\text{ORF plateau} = \text{intercept} + (\textit{time})(\text{linear}) + (\textit{time}^2)(\text{quadratic}) \tag{1}$$

The peak ORF score was that at which the difference between the current and previous time point was zero. In order to construct a range around the plateau, we computed the standard error of the measurement (*SEM*) using the following formula:

$$\text{SEM} = SD\sqrt{1 - r_{xx'}} \tag{2}$$

where *SD* represents the standard deviation of the observed test values, and $r_{xx'}$ represents the reliability of the test. In order to derive the test reliability, we used the unique measure residual variances for each time point to first derive the unique reliability for each time point (i.e., $\frac{\text{true score variance}}{\text{observed score variance}}$; Yeo, Kim, Branum-Martin, Wayman, & Espin, 2011). This allowed for the test reliability to vary across time, which is generally more tenable than assuming it is a fixed function of the test. We then averaged the reliabilities of the 12 time points for an estimate of the test reliability, and used the *SEM* to construct a 95% confidence interval around the ORF plateau.

Results

Our study demonstrated two main findings; the students' growth across years and the ORF plateau. Table 1 shows mean and standard deviation of observed score in each time point. For grade 3, the mean increased from fall to winter but slightly decreased from winter to spring, whereas the mean increased from fall to spring for Grades 4 to 6. The means of time points 4 – 9 (across Grades 4 and 5) constantly increased (positive growth trajectory), indicating students continued to develop their ORF skill in Grades 4 and 5.

The average observed scores across three time points for each grade were 108.63, 125.57, 152.60, and 153.81 wcpm for Grades 3 to 6, respectively (see Figure 1). The averages scores sharply increased from Grades 3 to 5, indicating students constantly improved their ORF scores when they were in Grades 3 to 5. The difference between the average observed scores for Grade 5 and Grade 4 was larger than the difference between the average observed scores for Grade 4 and Grade 3 (27. 03 vs. 16.94), indicating students developed 27 wcpm on average from Grade 4 to Grade 5. Students developed 17 wcpm on average from Grade 3 to grade 4. The average observed scores for Grade 5 and Grade 6 were almost the same (152.60 wcpm vs. 153.81 wcpm for grades 5 and 6, respectively). In other words, students developed only 2 wcpm on average from Grade 5 to Grade 6. It can be interpreted that students did not improve their ORF much from Grade 5 to Grade 6. In other words, ORF may be less informative for examining growth in passage reading speed. Our finding was consistent with previous research (e.g., Fuchs, Fuchs, Hosp, & Jenkins, 2001; Yovanoff, Duesbery, Alonzo, & Tindal, 2005): ORF tests may be less informative as reading skills more fully develop. Our finding could then be suggested that the ORF skill is fully developed once students are in Grade 5.

Regarding the ORF plateau, as expected, a quadratic model (DIC = 2,770,858) fit the data better than the linear model (DIC = 2,802,232). Results from quadratic growth model were used in this paper. The means of intercept, linear, and quadratic term were 81.13, 11.62, and – 0.46, respectively. Finally, the derived ORF plateau was 154.33 wcpm, and with an *SEM* of 16.61 wcpm (SD = 41.53, $r_{xx'}$ = .84), the range of ORF plateau scores was 137.72 – 170.94 wcpm. This ORF plateau range may be a potentially useful starting point for considering when to stop assessing ORF for the purpose of indicating reading skill growth because it indicates a more stable reading rate level, in which less information about reading skill growth can be obtained from the measurement. In other words, our findings suggest that for students who can read approximately 138 – 171 wcpm, less and less growth will be observed over time (as a result of steady and well-developed passage reading ability), rendering the regular assessment of ORF less informative for making instructional decisions.

When we examined the percent of students who reached our estimated ORF plateau level, we found that approximately 9% of Grade 3 students scored within or above this ORF range, as did 15% of Grade 4 students, 22% of Grade 5 students, and 16% of Grade 6 students. Again, our results demonstrated that ORF increased most while students were in Grades 3 through 5. The development of fluency appears to be active for Grade 3 students because so few of them (i.e., less than 10%) were able to reach our estimated ORF plateau range. It other words, it is probably too early to cease ORF assessing average students in Grade 3. Regarding Grades 4 – 6, between 78 – 85% (depending upon the grade) of students have not met the plateau in each of these grades. This finding implies that there is still plenty of room to grow their ORF for most students. Thus, this would seem to suggest that ORF assessing should continue up through Grade 6.

Conclusion

Our study demonstrated a range of fluency scores at which ORF growth plateaus as a potential guide to educators and researchers to cease assessing ORF, as we assume that the ORF indicates stable reading fluency levels and this plateau can be found in any of these grades. The range of ORF plateau scores was 137.72 – 170.94 wcpm, representing a potential reading rate level in which the continued assessment of ORF may become less meaningful. Also, our finding demonstrated the growth trajectory sharply increased in Grades 3 – 5, whereas the growth trajectory only slightly increased from Grade 5 to Grade 6. Although our study suggested that the ORF skill is fully developed once students are in Grade 5, a majority of these particular students have not reached this plateau yet. We could not say that we should stop accessing ORF at either grade5 or Grade 6. It is more reasonable to identify each student separately. Once a student reaches this plateau, he/she could move to more appropriate reading tests such as comprehension.

As preliminary analysis, we used both a maximum likelihood estimator with robust standard errors and a Bayesian. Both estimator obtained similar results, that is, the quadratic growth model fit better than linear growth model. However, the ORF plateau ranges from both estimators were quite different. The maximum likelihood estimator with robust standard errors obtained the derived ORF plateau of 174.86 wcpm with a range of 155.35 – 194.36, approximately 20 words more than results obtained by Bayesian approach. We thought a Bayesian approach obtained more reasonable results because 174.86 wcpm is probably too high for students to reach within a minute. It would be interesting to further study the estimation approach for both particular estimators.

Our study is able to provide useful information in terms of ORF plateau range by using a very large sample size with four grade levels. We also used a longitudinal and cross-sectional design, a limitation of which is the assumption that data missing by cohort is similar to data present for other cohorts. It would be interesting for future studies to include Grades 1 – 8, and to explore growth mixture models (GMM) to support a plateau range separate from grade level and more specific to fluency. While it is beyond the current study to offer a definitive ORF plateau, this study marks an initial estimate at which ORF assessment can be stopped for average readers.

References

Baker, S. K., Smolkowski, K., Katz, R., Hank, F., Seeley, J. R., Kame'enui, E. J., et al. (2008). Reading fluency as a predictor of reading proficiency in low-performing, high-poverty schools. *School Psychology Review*, *37,* 18–37.

Christ, T. J., Silberglitt, B., Yeo, S., & Cormier, D. (2010). Curriculum-based measurement of oral reading: An evaluation of growth rates and seasonal effects among students served in general and special education. *School Psychology Review*, *39,* 447-462.

Deno, S., Fuchs, L., Marston, D., & Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review, 30,* 507-524.

Deno, S., Marston, D., & Tindal, G. (1985). Direct and frequent curriculum-based measurement: An alternative for educational decision making. *Special Services in the Schools, 2,* 5-27.

Denton, C.A., Barth, A.E., Fletcher, J.M., Wexler, J., Vaughn, S., Cirino, P.T., Romain, M., & Francis, D.J. (2011). The relations among oral and silent reading fluency and comprehension in middle school: Implications for identification and instruction of students with reading difficulties. *Scientific Studies of Reading*, *15*(2), 109-135.

Fuchs, L., Fuchs, D., Hamlett, C., Walz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review, 22,* 27-48.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5,* 239-256.

Good, R. H., Simmons, D. C., & Kame'enui, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading, 5,* 257-288.

Hasbrouck, J., & Tindal, G. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. *The Reading Teacher, 59,* 636-644. doi: 10.1598/RT.59.7.3

Hosp, M. K., & Fuchs, L. S. (2005). Using CBM as an indicator of decoding, word reading, and comprehension: Do the relations change with grade? *School Psychology Review, 34,* 9–26.

Hudson, R., Lane, H., & Pullen, P. (2005). Reading fluency assessment and instruction: What, why, and how? *The Reading Teacher, 58,* 702-714.

LaBerge, D. and Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6,* 292-323.

Muthén, B.O. (2010). Bayesian analysis in mplus: a brief introduction.

Muthén, L.K. and Muthén, B.O. (1998-2012). Mplus User’s Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: an evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups* (NIH Publication No. 00-4754). Washington, DC: U.S. Government Printing Office.

Nese, J. F. T., Biancarosa, G., Anderson, D., Lai, C. F., & Tindal, G. (2012). Within-year oral reading fluency with CBM: A comparison of models. *Reading and Writing.* doi: 10.1007/s11145-011-9304-0

Nese, J. F. T., Park, B. J., Alonzo, J., & Tindal, G. (2011). Applied curriculum-based measurement as a predictor of high-stakes assessment: Implications for researchers and teachers. *Elementary School Journal*, *111*, 608-624. doi: 10.1086/659034

Reschly, A., Busch, T., Betts, J., Deno, S. L., & Long, J. D. (2009). Curriculum-based measurement oral reading as an indicator of reading achievement: A meta-analysis of the correlational evidence. *Journal of School Psychology*, *47*, 427-469.

Shapiro, E. S., Keller, M. A., Lutz, J. G., Santoro, L. E., & Hintze, J. M. (2006). Curriculum-based measures and performance on state assessment and standardized tests: Reading and math performance in Pennsylvania. *Journal of Psychoeducational Assessment, 24*, 19-35.

Speece, D. L., Case, L. P., & Molloy, D. E. (2003). Responsiveness to general education instruction as the first gate to learning disabilities identification. *Learning Disabilities Research & Practice, 18*, 147-156.

Stecker, P., Lembke, E., & Foegen, A. (2008). Using progress-monitoring data to improve instructional decision making. *Preventing school failure, 52*, 48-59.

Wayman, M. M., Wallace, T., Wiley, H. I., Tichá, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *The Journal of Special Education, 41*, 85-120.

Wood, D. E. (2006). Modeling the relationship between oral reading fluency and performance on a statewide reading test. *Educational Assessment, 11*, 85-104.

Yeo, S., Kim, D., Branum-Martin, L., Wayman, M. M., Espine, C. A. (2011). Assessing the reliability of curriculum-based measurement: An application of latent growth modeling. *Journal of School Psychology, 50,* 275–292.

Yovanoff, P., Duesbery, L., Alonzo, J., & Tindal, G. (2005). Grade-level invariance of a theoretical causal structure predicting reading comprehension with vocabulary and oral reading fluency. *Educational Measurement: Issues and Practice, 24,* 4–12.

Table 1

Mean and standard deviation of observed score for each time point

Time point	Grade	Season	n	Mean	SD
1	3	Fall	24,505	88.42	38.22
2	3	Winter	25,010	119.69	43.22
3	3	Spring	25,733	117.11	42.87
4	4	Fall	26,234	108.92	36.77
5	4	Winter	26,298	129.88	38.30
6	4	Spring	27,791	137.90	43.43
7	5	Fall	26,043	143.50	41.56
8	5	Winter	25,958	150.87	41.61
9	5	Spring	27,151	163.44	41.04
10	6	Fall	21,658	141.71	39.92
11	6	Winter	16,512	154.18	42.18
12	6	Spring	20,101	165.55	49.22

Figure 1

Average observed score for each grade