

A Review of the U.S. Text Level Assessment Process

Connie Briggs, Texas Woman's University; Emily Rodgers, The Ohio State University; and Anne Simpson, Texas Woman's University

Old Man Moss, Mr. Jumble, and the 'hum-an' in the house! These are all familiar characters to Reading Recovery teachers in the United States because they are featured in the Scott Foresman Reading Recovery Testing Pack (Pearson), a collection of stories that are used to assess student progress on text reading. The packet contains 20 leveled stories with standard introductions that teachers administer to students up to five times a year, in accordance with the national plan for the evaluation of Reading Recovery in the U.S. The same packet of stories has been used for the annual evaluation for the past 21 years.

This practice of using a standard set of stories is unique to Reading Recovery evaluation in the U.S. By contrast, teachers in every other country with a Reading Recovery implementation select stories from among the same set of books that they use for instruction.

In 2007, Reading Recovery professionals—citing concerns with the relevancy of some of the passages as well as an apparent unevenness in the relative difficulty of the levels—requested that the North American Trainers Group (NATG) review the testing packet and consider alternatives for assessing progress in text reading. As a result, a pilot study was

undertaken in 2008–2009. In this paper, we describe the results of the pilot and share recommendations for the future.

Development of the Scott Foresman Reading Recovery Testing Pack

The document, *Procedures for Administering Leveled Text Reading Passages* (NATG, 2005) provides an historical overview of the development of the single set of passages for text assessment in the U.S. According to that document, when Reading Recovery was first implemented in 1984, the first text passages to be used were selected from the basal series then in use in Columbus Public Schools in Ohio. These passages were selected and ordered along a gradient of difficulty under the guidance of Dr. Marie Clay and New Zealand Trainer Dr. Barbara Watson, both of whom were in residence at The Ohio State University to assist with the Reading Recovery start up in the U.S. As Reading Recovery expanded in the early years, it soon became apparent that in order to evaluate the effects of the intervention across multiple school districts, a single set of passages— independent of instructional materials and easily available to all Reading Recovery teachers across the country—needed to be developed.

As Reading Recovery expanded in the early years, it soon became apparent that in order to evaluate the effects of the intervention across multiple school districts, a single set of passages needed to be developed.

By 1990, the packet of leveled passages with their introductions was completed and, with the exception of small changes made in 2005 to the texts used for reading below level 1, the leveled stories in the Scott Foresman packet have served as a standard measure of reading progress in Reading Recovery for 21 years.

Designing the Pilot Study

NATG formed a task force in late 2007 to explore the concerns expressed by Reading Recovery professionals about the Scott Foresman packet. Because the school year had already started, it was decided that the period of time between January and June 2008 would be used to trial a design for the study and that, if possible, a more-formal pilot study would be undertaken from 2008–2009.

There was consensus that a new set of stories would have to be highly correlated with the Scott Foresman packet in order to avoid a break in the data that had been collected over the past 20 years. The first task therefore was to decide which set of little books would be trialed as the alternative.

The task force considered testing commercially available leveled text assessments but decided against it because such a change would prove costly to many school districts. The task force also considered writing and trialing a new set of stories just for Reading Recovery evaluation, but concluded that such a project would be costly and take several years to complete. It was finally decided that the task force would compare the texts in teachers' sets of leveled books, the ones used for instruction, to the Scott Foresman passages; the same process used by Reading Recovery teachers in every other country except the U.S.

The task force also agreed that the study would provide a good opportunity to pilot the Slosson Oral Reading Test-Revised (SORT-R) as a standardized external measure that would complement the six tasks of *An Observation Survey of Early Literacy Achievement* (Clay, 2002, 2006).

The SORT-R consists of a series of graded word lists from Primer to Grades 9–12. It is an oral reading assessment; students are instructed to read as many words as they can until they reach a list where they do not recognize any words. The SORT-R was selected because it can be administered quickly and easily and because it has good concurrent criterion validity. The Slosson technical manual reports strong correlations



Teachers in the study followed the data collection points in the national evaluation protocol; each time they gave the Scott Foresman packet, they also selected texts from their set of Reading Recovery leveled books to administer.

between the SORT-R and the Peabody Individual Achievement Test, as well as the Woodcock-Johnson Tests of Achievement — at $r = .90$ and higher (Slosson, 2008).

Participating teachers were also interviewed as part of the exploratory study. They were asked to share their views on using the Reading Recovery leveled books instead of the Scott Foresman set and to provide feedback on the design of the pilot.

Finally, Canadian trainer Irene Huggins assessed 25 Canadian Reading Recovery students in the usual way for Canada with Reading Recovery leveled books and then with the Scott Foresman leveled passages. Huggins found that 92% of students assessed read at the same level or within two text levels on both the Scott Foresman and the Reading Recovery leveled books.

Results from the exploratory study provided the task force with feedback to design the more-formal study that followed. Teachers responded positively to the SORT-R, noting the ease of the administration and the insights that the assessment provided in to how students “took apart” unfamiliar words in isolation. They also said that it was extremely time-consuming to test all of their students so, for the study, each teacher was asked to test a minimum of two students, more if time permitted. Other procedures were refined and a process for testing at the beginning of the year was developed.

Study Methods

Participants

The task force invited participation from all of the university training centers. Eighty-four teachers, 12 teacher leaders, and trainers from Arkansas, Georgia, Kentucky, Ohio,

and Texas volunteered to take part. Data were collected on 340 Reading Recovery and 84 random sample students.

Design

Two formal research questions guided the inquiry:

1. What are the correlations between Scott Foresman leveled passages and Reading Recovery leveled books?
2. Does the Slosson correlate with Scott Foresman texts? Does it correlate with Reading Recovery leveled books?

Teachers followed the data collection points in the national evaluation protocol; each time they gave the Scott Foresman packet, they also selected texts from their set of Reading Recovery leveled books to administer. All teachers gave the SORT-R to their students. The order of assessments varied across the teachers so that no one assessment was always given first.

Students' scores on the three assessments were correlated for three points in time: at entry to and exit from Reading Recovery lessons and at year-end.

Findings

Correlations between the leveled texts

In response to the first question, the correlations between the Reading Recovery leveled books and the traditional Scott Foresman packet were moderate at entry ($r = .69$), strong at exit ($r = .79$) and stronger at end of year ($r = .84$) for Reading Recovery students. All correlations were significant at the .01 level.

Correlations with the SORT-R

The traditional Scott Foresman passages correlated better with the external measure, the SORT-R, at all three points in time than did the Reading Recovery leveled books ($r = .61, .62, .68$ with the SORT-R compared to $r = .59, .57$ and $.64$ with the leveled books). This trend was the same with both the Reading Recovery and the random sample students. Based on these findings, it would appear that the traditional Scott Foresman passages correlated better with the external measure, the SORT-R, than the Reading Recovery leveled books.

Eighty-four teachers, 12 teacher leaders, and trainers volunteered to take part in the study. Data were collected on 340 Reading Recovery and 84 random sample students.

These better correlations between the Scott Foresman packet and the SORT-R should be interpreted with caution, however, for at least two reasons. First, the Scott Foresman packet contained less variation in titles than the set of Reading Recovery leveled books. There was only one title at each Scott Foresman level to use, whereas teachers could choose from among many titles at each level from their set of Reading Recovery leveled books. In fact, a frequency count showed that more than 300 different Reading Recovery leveled books were selected by teachers, com-

pared to the 30 possible titles available in the Scott Foresman packet.

The greater variability in Reading Recovery titles within levels likely increased measurement error. In our analysis we did not estimate the error in order to correct the correlations. However, we can say with some confidence that when corrected, the correlations between Reading Recovery leveled books and the SORT-R would likely be stronger. We make this prediction because, even unadjusted for the error, the correlations are strong.

In addition, the Scott Foresman packet contained standard book introductions, whereas for the Reading Recovery leveled books teachers were directed to only read the title of the book and not to provide an introduction. The lack of introduction for the Reading Recovery books was intended to reduce the variation in how books were introduced and therefore have some assurance that a text's level of difficulty would be unchanged by a book's introduction. We have no way of knowing, however, how much error resulted because of this difference in how the two sets of stories were introduced.

Recommendations

As a result of this exploratory work, the task force recommended to NATG that teachers continue to use the Scott Foresman leveled passages to assess students' progress on the text level measure. The findings did not reveal adequate evidence to recommend a transition. In addition, the strong correlations between the traditional Scott Foresman packet and the SORT-R, as well as the desire to maintain continuity in

the data set across the years, strengthened the case to keep the traditional passages.

The task force also noted that to make a change in assessment practices, another study would be needed to ensure that significant shifts in the longitudinal data did not occur. It was also noted that this process would require a considerable amount of time and financial support.

Thus the NATG task force recommended the continued use of the Scott Foresman assessment passages for the current time and, at their April 2010 meeting, that recommendation was accepted by NATG. At the same time, it was agreed that the NATG Research Committee would review and recommend an external measure to supplement our current assessment instrument.

The task force did not address the concern expressed at the outset about the apparent unevenness in the difficulty of the levels. This question will be addressed in an investigation using item response theory methods that is currently underway by Jerry D'Agostino, the International Data Evaluation Center's director of research.

Finally, while these decisions do not address concerns about the appearance of the Scott Foresman passages or the relevance of the texts, teachers and teacher leaders can feel more confident about their students' outcomes and the continued reliability of Reading Recovery's results.

The authors wish to thank the teachers, teacher leaders, and trainers who participated in the two studies. Thank you also to Billie Askew, professor emerita, Texas Woman's University, who served as an advisor to the project; and to Jerry D'Agostino, professor in the College of Education and Human Ecology, The Ohio State University and director of the International Data Evaluation Center; and Rene Paulson, biostatistician at Texas Woman's University, who provided support with analysis.

References

- Clay, M. M. (2002, 2006). *An observation survey of early literacy achievement* (2nd ed., rev. 2nd ed.). Portsmouth, NH: Heinemann.
- North American Trainers Group (2005). *Procedures for administering leveled text reading passages*. Columbus, OH: Author.
- Scott Foresman Reading Recovery Test Pack (1979). New York: Pearson.
- Slosson, R. L. (2008). *Slosson Oral Reading Test-Revised*. East Aurora, NY: Slosson Educational Publications.

About the Authors



Connie Briggs is a Reading Recovery trainer and professor in the Department of Reading at Texas Woman's University.



Emily Rodgers is an associate professor in the College of Education and Human Ecology at The Ohio State University and a Reading Recovery trainer.



Anne Simpson is a professor in the Department of Reading and directs the Reading Recovery university training center at Texas Woman's University. She teaches literacy courses and works with the K-12 reading specialist masters program. Prior to coming to TWU, Anne served as a district language arts and Reading Recovery site coordinator for 19 years. She currently serves as president of the North American Trainers Group.