WHITE PAPER

The Research Behind the DRA3 Benchmark Assessments

Overview

Rob Altmann, MA

The Developmental Reading Assessment™, Third Edition (DRA™3) Benchmark Assessments allow teachers in grades K–8 to systematically observe, record, and evaluate changes in student reading performance. DRA3 Benchmark Assessment components provide teachers with information that can be used to determine a student's Independent or Instructional reading level and identify what the student needs to learn next. Teachers can choose from 102 fiction and nonfiction books that are distributed across 23 reading levels. The Benchmark Assessments can be used two or three times a year to document a student's reading progress.

This paper summarizes several research studies that led to the development of the DRA3 Benchmark Assessments, as reported in the DRA3 Technical Manual (Pearson, 2019).

Study 1: Development of the Level Estimator

A new component of the DRA3 is the Level Estimator (LE), which is a word list used by teachers to establish a starting level for the benchmark books when no other information about a student's reading performance is available (e.g., a new student). The LE was developed in two phases. During Phase 1, an initial 70-item word list was created by selecting words with a variety of difficulty levels. One-hundred ninety-four students in Grades 1–5 were asked to read as many of the words on the list as they could and were also asked to read a 1-minute reading passage that was used to evaluate a student's proficiency in oral reading fluency.

Items were scored as correct when the student was able to correctly read the word in a short amount of time (within approximately one second). Item scores were summed to create a total score. A Rasch-based Item Response Theory (IRT) approach was used to calibrate item-difficulty estimates for the LE words. Items were removed from the item pool if they did not adequately measure the overall construct or if they were too similar in difficulty to other items, resulting in a final set of 40 items. The Spearman-Brown formula for split-half reliability was used to measure internal consistency, resulting in $r = .98$, which is considered excellent. The LE total score was correlated with grade and the total score from the 1-minute reading passage, resulting in an $r = .80$ and $r = .94$, respectively. As expected, LE scores increased as grade increased and were strongly correlated with oral reading fluency performance.

During Phase 2 of the LE development, LE scores were compared to scores obtained on DRA3 benchmark assessments. Mean LE scores were computed for students who were rated at the independent level on Oral Reading Fluency across a wide range of DRA3 benchmark assessment reading levels (i.e., Levels 4, 8, 12, 14, 18, 20, 24, 28, 34, 40 and 50). These scores were used to determine starting points for the DRA3 benchmark assessment books based on a range of LE score (scores beyond level 50 were interpolated); the resulting recommendations are provided in Table 1.

Table 1. Recommended DRA3 Benchmark Assessment Levels Based on Level Estimator Score

Level Estimator score	DRA3 level
0–7	A–3
8–10	4–8
11–15	10–12
16–19	14–18
20–21	20–24
22–25	28–30

Level Estimator score	DRA3 level
26–28	34
29–30	38–40
31–32	50
33–34	60
35–37	70
38–40	80

Study 2: DRA3 and DRA2 Scoring Equivalency

A primary goal of the DRA3 project was to improve the usability and the clarity of the Teacher Observation Guides (TOG), which are used by teachers to record, evaluate, and rate student performance. To ensure changes made to the TOGs did not impact scoring, a study was conducted to examine the equivalency of DRA3 and DRA2 scoring criteria. Forty-eight students and eight teachers participated in a study that used benchmark assessment materials from levels 8, 20, 34, and 50. Using videotaped oral reading sessions and copied student booklet materials (i.e., student booklets for levels 34 and 50), teachers rated on average 24 students, using either a DRA3 or a DRA2 TOG (TOG assignments were made using a counterbalanced procedure to minimize potential ordering effects). This resulted in a total of 174 usable teacher ratings across all students and assessment levels. The results of the Oral Reading Fluency and Comprehension total scores are provided in Table 2 (results for individual ratings that comprise each of the total scores are found in the DRA3 Technical Manual). Overall, scores obtained on DRA3 and DRA2 TOGs were not significantly different from each other, providing evidence that the improvements made to the DRA3 TOGs did not negatively impact scoring results.

	DRA3		DRA2				
	Mean	*SD*	Mean	*SD*	*t* value	Standard difference	*SEM*
Level 8							
Oral Reading Fluency	12.8	2.7	12.9	2.4	0.13	-0.02	.50
Comprehension	21.8	3.1	21.8	2.7	0.07	-0.01	.57
Level 20							
Oral Reading Fluency	13.4	1.7	13.3	1.9	0.17	0.05	0.38
Comprehension	20.5	4.1	20.1	2.5	0.36	0.11	0.72
Level 34							
Oral Reading Fluency	13.3	1.8	13.2	1.7	0.24	0.09	0.49
Comprehension	19.1	3.3	20.2	4.1	1.09	-0.30	1.03
Level 50							
Oral Reading Fluency	13.2	1.0	13.3	1.6	.20	-0.11	0.41
Comprehension	18.0	2.7	17.8	3.0	.21	0.06	0.87

Study 3: DRA3 and DRA2 Benchmark Assessment Book Comparability

For the DRA3, 59 (of 102) benchmark assessment books were newly written and were subsequently refined over the course of a two-year period. The process included four primary phases:

Phase 1: Topic Selection and Drafting of Content. The development team (i.e., the test authors, test development experts, professional illustrators, graphic designers, and editors) worked with classroom teachers to identify interesting and age-appropriate topics for new texts. Writers submitted drafts of each text, using existing DRA2 books as a general guide for length, complexity, and amount of graphic support needed for each DRA reading level.

Phase 2: Initial Content Editing and Leveling. Each text draft was rigorously reviewed by the publisher's content development team to ensure accuracy, logical flow, and consistent leveling within DRA2 and DRA3 texts. Consideration was given to things like the total number of words, total number of sentences, average words per sentence, range of sentence lengths, and Flesch-Kincaid grade level. Books were then edited to be as similar as possible to existing DRA2 books.

Phase 3: Content Review and Leveling by DRA Authors. The DRA authors completed a detailed review of the preliminary version of all newly written texts, in an effort to provide feedback on the appropriateness of text content and vocabulary for the designated DRA reading level. They also reviewed each benchmark book to ensure consistent leveling by checking the difficulty level of each 100-word section of text using the Fry readability formula.

Phase 4: Content Refinement and Bias Review. The content of each book was finalized by conducting copyediting rounds, final fact-checking, photo selection or illustration development, creation of graphics, and minor content updates as needed based on bias review feedback.

To evaluate the consistency of reading levels between the newly written DRA3 books and existing DRA2 books, student performance using a subset of the new books was compared within the same level (using levels 8, 20, 34, and 50). Two-hundred thirty-four students in grades 1, 2, 3, and 5 participated in a study to compare oral reading fluency and comprehension performance between newly written and existing books. The administration of newly written and existing texts was counterbalanced to minimize possible order effects on test scores. Tests were administered across two sessions, separated by 1 to 21 days. Table 3 presents the results of the study. Mean performance for both Oral Reading Fluency and Comprehension was very similar between stories for all levels, with a median standard difference (absolute value) of .18, which is considered a small difference (Cohen, 1988). Correlations between scores are stronger for Oral Reading Fluency than Comprehension, which is expected given the more objective nature of categories rated in Oral Reading Fluency (e.g., rate, accuracy). Correlations range from .79–.89 for Oral Reading Fluency and from .63–.81 for Comprehension. Overall, there is a strong relationship between scores obtained from existing (DRA2) and newly written (DRA3) books.

Table 3. Within-Level Comparisons Between New DRA3 and Existing DRA2 Texts

	DRA3		DRA2				
	Level	Mean	*SD*	Mean	*SD*	Corrected *r*	Standard difference
Oral Reading Fluency	8	10.9	4.1	10.6	4.5	.89	-0.07
	20	12.2	3.9	12.4	0.07	.89	0.06
	34	13.0	2.8	12.4	0.17	.79	-0.22
	50	12.3	3.3	12.8	0.36	.87	0.16
Comprehension	8	17.7	4.5	18.0	0.24	.63	0.05
	20	19.8	5.0	17.9	1.09	.65	-0.39
	34	18.8	4.0	19.8	.20	.63	0.22
	50	16.3	4.9	15.2	.21	.81	-0.22

Study 4: Internal Consistency and Inter-Rater Reliability Study

Reliability is an estimate of the consistency or stability of test scores. For DRA3 benchmark assessments, teachers use the Teacher Observation Guides (TOGs) to rate a student's performance on oral reading fluency, printed language concepts (levels A–3), and comprehension (levels 4–80). Specifically, teachers rate performance using the Continuum found at the end of each TOG. While each benchmark book has its own corresponding TOG (102 in total), there are fewer unique Continuums; a total of 14 Continuums are used across all TOGs. In order to estimate the reliability of DRA3 benchmark assessment scores, a sampling of TOGs was used, including Continuums 3 (Fiction, levels 4–12), 5 (Fiction, levels 18–24), 6 (Fiction, levels 28–38), 7 (Fiction, levels 40–80), 8 (Nonfiction, levels 4–12), 9 (Nonfiction, levels 14–16), 10 (Nonfiction, levels 18–24), and 12 (Nonfiction, levels 40–80). Internal consistency (coefficient alpha) reliability estimates were computed for all TOGs included in the studies discussed previously. Table 4 presents the reliability estimates by continuum. Oral Reading Fluency estimates range from .84 to .94 (median of .88); Comprehension estimates range from .78 to .88 (median of .82). Overall, coefficient alpha estimates are considered good to excellent.

Table 4. DRA3 Benchmark Assessment Books: Internal Consistency (Coefficient Alpha) Reliabilities, Oral Reading Fluency and Comprehension

		Alpha	
Continuum	N	Oral Reading Fluency	Comprehension
3	344	.91	.83
5	195	.90	.88
6	360	.87	.81
7	66	.85	.80
8	77	.94	.84
9	86	.84	.79
10	161	.87	.78
12	185	.89	.87

The scores obtained during the DRA2 and DRA3 scoring comparison study (Study 3) were also used to estimate another type of score reliability: the consistency of scores across different raters, known as inter-rater reliability. For a test like the DRA3, it is important to demonstrate that a student's performance isn't the result of peculiarities in the way a particular teacher rates a student's reading ability. As described earlier, this study involved multiple raters using videotaped test administrations to rate a student's performance. For this study, inter-rater reliability was assessed by determining the similarity of scores assigned to the student's performance. On the DRA3, Oral Reading Fluency and Comprehension are assigned a rating of Intervention/Emerging, Instructional/Developing, Independent, or Advanced. Table 5 provides the level of agreement (percent) among raters in assigning performance levels to the same students. Overall, there was considerable agreement between raters; agreement for Oral Reading Fluency ranged from 76%–91%, and agreement for Comprehension ranged from 69%–80%.

Table 5. Inter-rater Agreement of Oral Reading Fluency and Comprehension Category Ratings

	Percent Agreement	
Level	Oral Reading Fluency	Comprehension
8	77	79
20	76	80
34	76	69
50	91	69

Study 5: Correlations With Other Measures

Two-hundred thirteen students in Grades 2–5 who completed a DRA3 benchmark assessment in Study 3 also completed the aimswebPlus Oral Reading Fluency test, a widely used brief measure designed to assess a student's oral reading performance. DRA3 Oral Reading Fluency and Comprehension scores were compared to the aimswebPlus Oral Reading Fluency score. Table 6 presents the correlations between scores. As expected, DRA3 Oral Reading Fluency scores show a strong, positive correlation with aimswebPlus scores across all grade levels, ranging from r = .72 to r =.79. DRA3 Comprehension score correlations are somewhat lower, as expected, ranging from r = .62 to r = .67.

Table 6. Correlations With aimswebPlus Oral Reading Fluency, by Grade

	aimswebPlus Oral Reading Fluency	
Grade	DRA3 Oral Reading Fluency	DRA3 Comprehension
2	.78	.62
3	.77	.65
4	.72	.67
5	.79	.67

Pearson