

CRITIQUE—DRA2

**Written Critique of an Existing Measure**  
Developmental Reading Assessment (DRA2)  
Donna Newberg-Long  
QRM 4900  
Fall 2007

## **Identification of the Measure**

This critique of the revised *Developmental Reading Assessment (DRA2)* will provide an in depth review of this reading assessment tool recently used in many schools across the United States.

The most recent editions include: the DRA2 for grades kindergarten through third, published in 2006; the DRA2 for grades fourth through eighth, published in 2003; and the DRA Word Analysis assessment published in 2005. In addition, a Spanish version has been published entitled *Evaluacion del desarrollo de la lectura (EDL2)* meaning *Reading Development Evaluation*, in English.

## **Authors**

The DRA2 was developed by Joetta Beaver and a group of primary teachers in Upper Arlington School District in Ohio. With over thirty years experience as an educator, Beaver also conducts workshops on topics such as developmental assessment, supporting emergent readers and writers in the classroom over time in reading and writing. Because of the DRA K-3 success, Joetta Beaver and Mark Carter, Ph.D. teamed up and expanded the program to include an intermediate grade assessment, DRA 4-8.

Mark Carter, PhD has focused much of his professional work on assessment. He served as a coordinator of assessment for Upper Arlington Schools (where he currently teaches fifth grade), conducted numerous seminars, and co-authored DRA2 4-8, DWA, and Portfolio Assessment in the Reading Classroom. He received his doctorate of philosophy from Ohio State University where he also taught graduate courses in education as an adjunct professor.

The impetus that led Beaver to develop the DRA2 was a desire to find a way to help teachers better evaluate students' reading and writing skills. The DRA2 is published and sold by Pearson Education, Incorporated, and is located at 299 Jefferson Road, Parsippany, New Jersey 07054. Their phone number is 1-800-321-3106, and e-mail address is [www.pearsoned.com](http://www.pearsoned.com) or [www.pearsonlearning.com](http://www.pearsonlearning.com).

The DRA2 is a set of individually administered, criterion-referenced reading assessments that provides a method for evaluating and documenting student progress in reading. The intention is that the test be scored and interpreted by classroom teachers. Both the DRA2s for primary and intermediate grades measure students' independent reading levels based on criteria including student engagement, oral reading accuracy, fluency, and comprehension. It is also intended to assist teachers in planning what students need to learn next.

The DRA Word Analysis accompanies the DRA2 and will not be the focus of this critique, however, it is important to recognize that it exists and assists in evaluating student reading skills. Its purpose is to evaluate phonics skills and phonemic awareness in kindergarten and first grade students, as well as these skills in students who are below grade level in grades two through five.

The DRA2 consists of two kits, one for K through 3 and another for 4<sup>th</sup> through 8<sup>th</sup> grades. The components of the DRA2 kit for K—3 include a spiral-bound teacher guide, training DVD, forty-five Benchmark Books for K-1, twenty-two leveled books for 2—4, spiral-bound black-line masters book, Teacher Resource Guide, Assessment Forms and black-line masters on CD, laminated assessment procedures overview card, timing clipboard with calculator, Word Analysis Teacher Guide and Assessment Book, 46 hanging file folders, and

thirty student assessment folders, all in a colorful storage box. For fourth grade through eight, the same is provided except, there are three each of twenty eight leveled texts and twenty additional texts for on-level readers in Grade four through eight and eight bridge pack texts for intermediate and middle school students reading at a second or third grade level. It is possible to purchase an On-Line Management System that includes a digital pen for scoring.

The DRA2 for 4—8 differs from the K—3 only in that students respond in writing rather than retelling beginning at level 28. There is also the addition of a vocabulary item to the comprehension questions.

Both DRA2 for K—3 and 4—8 include two basic components; a student reading survey and a set of leveled paperback books, each including two fiction texts and two additional non-fiction texts at levels sixteen, eighteen, thirty-eight and forty. The DRA Word Analysis includes forty tasks organized into five skill areas. These areas are the five essential components of reading identified by the National Reading Panel in 2000.

The cost of each comprehensive kit is as follows:

- K—3 = \$323.95
- 4—8 = \$323.95
- DRA2 On-line Management System for Grades K-3
  - \$87.95 per teacher per year subscription
  - Starter Kit includes digital pen and 6 pack of score sheets for every title
    - \$178.95 per Kindergarten
    - \$278.95 per teacher for grades 1-3
    - Refill Packs (\$9.95 per 6 pack) (Replace the Blackline Masters and hence lowers copying expense, but the expense would be great each year)

## **Description of the Measure**

The purpose of the DRA2 is to analyze student-reading performance within a literature-based reading program. With results of the DRA2, you can set instructional goals specific to each child, document student progress over time and match readers to appropriate reading levels and books. The DRA2 helps teachers identify students' strengths in reading and to drive effective instruction. The assessments are conducted during one on one student/teacher conferences. Each student is able to choose from a few texts, fiction or non-fiction, so that the text is of interest to the student. Use of student teacher conferences to discuss reading strengths and goals is the heart of the implementation of this assessment.

A Teacher Observation Guide is customized for each Benchmark Assessment Book. Each book has a corresponding continuum that scaffolds the teacher's analysis of the students level of performance in Reading Engagement, Oral Reading Fluency and Comprehension. Descriptions are provided for a range of possible behaviors and responses across four performance levels. These include: Emergent levels A—12 entitled Intervention levels 14—40; Developing levels A—12 entitled Instructional levels 14—40; Independent; and Advanced. After scoring each area being assessed (Reading Engagement, Oral Reading Fluency and Comprehension), A total score is calculated providing useful information for teachers to drive instruction.

The last page of each assessment is entitled the DRA2 Focus for Instruction. This tool allows teachers to plan for instruction and to group students for small group instructions or interventions based on individual needs.

A Student Reading Survey encourages students to take more responsibility in setting personal reading goals that are relevant to them. It also informs the teacher about the students' interests and level of reading success based on student choices for reading and self assessment.

### **Administration Credentials**

The DRA2 was developed for classroom teachers to enable them to evaluate student progress in reading. The credentials required to administer the DRA2 then must include teaching credentials of highly qualified staff. DRA2 training is also required before anyone can administer the assessment to students.

### **Research**

Between 1988 and 1996, the original DRA was field tested in the Upper Arlington City School District of Ohio. Changes to the forms, procedures and Benchmark Books used for testing have been made over the years in response to feedback from the field tests done with classroom teachers. Additional field tests were held in May and September of 2000 with school districts in Canada and the United States. The DRA Word Analysis was added in 2004. It was field tested in 2003 and revisions were made based on input from participating class teachers.

DRA2 was field tested in fall of 2004 and spring of 2005 and was published for use in 2006. The selection of targeted skills and assessment methods is based on current research in reading from sources such as the *National Reading Panel Report* (2000) and *Reading for Understanding* (Rand Education, 2002).

There is evidence to support validity, reliability, predictive usefulness, and correlation to other known measures. For example, evidence demonstrates a dramatic drop in students reading below grade level in classroom that use DRA as a classroom assessment tool. Interrater reliability was demonstrated when reliability was analyzed. Positive outcomes of several validity studies also support use of the DRA and EDL.

### **Time**

The DRA2 can be time consuming. For levels A through 28, testing takes anywhere from thirty to forty-five minutes based on the skill of the reader. Administering the Word Analysis tasks can take extended periods of time when a student continues to pass the tasks. This, however, can lead to a thorough analysis of the students' phonemic awareness and related word skills. For levels 30 and up, testing takes about thirty minutes for each student. Since students at levels 28 and above read independently and respond in writing to questions in the Student Booklet, less time is demanded from the teacher for each student.

At level forty, students must complete a student booklet that includes prediction, text features, summarizing and comprehension questions. It generally takes thirty-five to forty-five minutes for students to complete the Student Booklet on their own.

Another area is time intensive for teachers. That is time to score, based on the rubric of descriptors for the Continuum and to determine the Focus for Instruction page of the assessment. Teachers need planning time amounting to approximately two days to complete their entire class and additional time to score and analyze student outcomes. Another step that can promote student success in reading is time to meet with colleagues in order to group students for interventions and focus on particular skills.

**Reliability of the DRA2**

Reliability of the DRA2 is reported differently depending on which report you read. According to E. Jane Williams, Ph.D., inter-rater reliability between the originator and second rater is strong, while, when using a third rater, it was not so strong. Inter-rater reliability between three raters was .74, while only two rated at .80.

According to Popham (2006), "...whereas standardized test developers strive for .80 to .90 reliability coefficients when dealing with stability...they may be well satisfied with a solid .70 or so when they combine stability and alternate-form reliability." (p. 108) There is a problem with the DRA2 report by DRA2 On-line Interim Overview by Pierson Learning that reports "...the measure of reliability was (greater than) .75. Given the small set of items, (either 6 or 7 depending on the text) this represents adequate reliability." (p.4) This problem lies in the fact that the measure was not given to all benchmark texts being used. Also the study did not include an equitable number of students from all ethnic and socio-economic backgrounds. Therefore, I believe there is a misrepresentation of the stated reliability coefficient by Pierson Learning.

Popham states, "In order for a test to yield valid score-based inferences, it must be reliable...A test can be reliable without necessarily leading to valid score-based inferences." (p 100). Since stability of a test's reliability is based on an assessment's consistency over time, there is a question about the stability of the DRA2 due to the fact that the scoring from teacher to teacher has been found to be inconsistent.

Reliability comes into question due to numbers of areas that may be troublesome according to Natalie Rathvon, Ph.D. In her review of the DRA2, insufficient numbers of

students are sampled at several of the grade levels and for each Benchmark text. She found that the field test was also done only with students who were reading on grade level which “highly inflated mean scores and reduced variance.” (p. 8) Greatly underrepresented were students from low socio-economic, ethnically diverse backgrounds and those with disabilities. This skews the reliability of the measure.

So much of the scoring and administration is left to the judgment of the teachers, inter-rater reliability will be difficult to attain. On the other hand, since this is not used for high-stakes consequences, the instrument is valuable in providing teachers with data, however skewed, to assist them in developing instruction to meet the needs of their students.

The Rasch analyses determined that the reliability for the summed score for individual teacher rating was high, while that of rater reliability was low. Standard deviations were large. There is no reporting of mean scores and standard of errors for the alternative texts at each level which poses a problem for both reliability and validity of the instrument.

According to the DRA2 Interim On-line report provided by its publisher, Pearson Learning, Dr. Clement A. Stone, a psychometrician at University of Pittsburgh, provided analyses during the process of development of the DRA2 and is currently conducting reliability and validity analyses for additional evidence of such. This report states that reliability evidence has determined that the texts used for testing are equal in reading ability at each level and that “... comprehension assessment results are consistent.” Cronbach’s alpha reliability was used to measure consistency of the comprehension items and found the measure of reliability to be greater than 75%, which is adequate according to the interim

report. Internal consistency was strong according to E. Jane Williams. Item separation reliability using Cronbach's alpha was .98 while test separation was .97.

### **Validity of the DRA2**

According to Popham, "...tests themselves are never valid. Rather, the concept of validity is linked to the interpretations that educators draw based on the use of tests." (p 80) Validity of the DRA2 is questioned although input from teachers was used to revise the instrument. The concern is raised that no review was done by external reviewers in developing the assessment instrument. According to Natalie Rathvon, Ph.D., "No evidence of its criterion-related or construct validity is currently available, such as information concerning its relationship to other validated reading measures, teacher ratings, or school grades; utility in predicting future reading achievement or response or response to intervention; or ability to distinguish between poor or proficient readers."

Because of these things, she does not recommend the DRA2 for screening, diagnosis, progress monitoring or intervention planning purposes. I believe that Ms. Rathvon is overly critical in her assessment and that the DRA2 is helpful to teachers in planning for instruction since it is not considered a high stakes assessment. Indeed, the DRA2 could provide more assistance to teachers in regards to more clarity of scoring guidelines and examples of acceptable and unacceptable student responses for many of the items.

Popham refers to two basic strategies that can be applied to determine validity via content-related evidence. He states the use of, "...additional content experts as necessary..." can be of great assistance in determining the validity of an instrument. (p 84) The developers

of the DRA2 did not seek the assistance of outside experts. Therefore, the validity of the instrument becomes more questionable. Popham states that "...content-related evidence of validity should be a matter of paramount importance...make sure that the description of what the test measures is sufficiently accurate so that precise interpretations can be drawn from a student's test performance." (p 87) The DRA2 is used to determine the focus for instruction for students. The concern arises with content-validity due to the fact that teacher judgment varies on the DRA2. However, since the DRA2 is not used as a high stakes measure, it is, in my opinion, useful in assisting teachers in creating instruction groups and focusing instruction of these groups on areas of need.

There were no studies done on bias and difficulty of items in the test. According to Rathvon, Caucasian students had higher increases than African American and Hispanic students in reading level. These ... "effect sizes accounted for less than 1% of the variance of change." (p. 13) She also suggests that there is a high degree of variability within and between the DRA levels, leading to lower construct validity.

Williams also believes that because the DRA can be strongly correlated to the Iowa Test of Basic Skills, and strength is added to the case for DRA as a valid measure to assess student's ability to decode and comprehend text.

### **Personal Experience with DRA2**

As the Site Assessment Leader at my school, I have overseen all testing in order to meet CBLA testing requirements. We assess students' reading ability through DRA2/EDL2 assessments, given three times per year. The DRA2 is time intensive yet yields useful data for teachers. This tool is well organized and somewhat complex, demanding that teachers

take the training and study the manuals to become familiar with scoring rubrics, Word Analysis tasks and how to administer the test to their students and analyze results.

Criticism of this assessment tool includes the need for more time for teachers to do testing. Two days is requested for testing by teachers, while a substitute teaches the class, so they can focus quality time on assessing their students. It generally transpires that teachers have to assess students, one or two each day over a month, while the other students are doing independent reading. Teachers would prefer to get the testing done while a substitute take the class, rather than try to fit it in when they are also responsible for instruction and classroom management. Another concern is that there is no rubric for all Benchmark texts. The rubric is a great help to teachers in understanding what is expected to score proficiently on the Continuum.

One wonderful aspect of the DRA2 is the On-Line Management System for K-3 that includes a digital pen for scoring. Purchase of this added feature would save time and assist teachers in forming student groups based on needs. However, few schools actually purchase it for their teachers because of the expense. The price is monumental for many public schools. Another criticism is that there is no option for On-line Management for grades above third.

Purchase of this feature would simply make this assessment much more user friendly and would save the time that teachers complain they need to get it all done. This on-line system groups students for instruction, creates useful reports for teachers, and charts progress. I would suggest that this is essential to the program's success.

References

Beaver, Joetta M. (2006) *DRA2 Developmental Reading Assessment: Teacher's Guide*.

Celebration Press, Pearson Learning Group, pp 4—141.

DRA2 On-line Interim Report Overview; Pearson Learning 2006.

[www.pearsonlearning.com/content/File/DRA/DRA2\\_TM\\_Interim\\_Report.pdf](http://www.pearsonlearning.com/content/File/DRA/DRA2_TM_Interim_Report.pdf)

Popham, James W. (2006) *Assessment for Educational Leaders*: Pearson Education, Inc.,

pp 100—116 and 79—95.

Rathvon, Natalie, Ph.D. (2006) Developmental Reading Assessment Review

[Natalierathvon.com/images/DRA](http://Natalierathvon.com/images/DRA)

Williams, E. Jane, Ph.D.(1999) Developmental Reading Assessment Reliability Study.

[www.pearsoned.com/RESRPTS\\_FOR\\_POSTING/READING\\_RESEARCH\\_STUDIES/r1](http://www.pearsoned.com/RESRPTS_FOR_POSTING/READING_RESEARCH_STUDIES/r1)

[1.ResearchPaper\\_DRA.pdf](http://1.ResearchPaper_DRA.pdf)