

Test Review

Beaver, J. M., & Carter, M. A. (2006). *The Developmental Reading Assessment—Second Edition (DRA2)*. Upper Saddle River, NJ: Pearson.

Reviewed by: Allison M. McCarthy and Theodore J. Christ, *University of Minnesota, Minneapolis, MN*
DOI: 10.1177/1534508410363127

The *Developmental Reading Assessment—Second Edition (DRA2)* was published in 2006 as a teacher-administered assessment to identify students' instructional level, along with their strengths and weaknesses in reading. Once teachers calculate and interpret scores, the data can purportedly be used to inform, and possibly individualize, instruction. The second edition includes components not contained in the original tool such as measures of oral reading rate, a revised stopping point for test administration, and instructional reading level placement for each student.

Purpose

The DRA2 is a criterion-referenced assessment, and the data can be used to (a) assess reading engagement, oral reading fluency, and comprehension, (b) identify reading strengths and weaknesses, (c) determine students' reading levels, (d) inform reading instruction, (e) monitor progress in reading, and (f) aid in planning reading interventions. The developers contend that the DRA2 is a convenient assessment because classroom teachers can administer it to students in a one-on-one setting in kindergarten through eighth grade. The developers also posit that the DRA2 can be administered repeatedly both within and across academic years, which is necessary for progress monitoring.

Description

The DRA2 assessment materials include benchmark assessment books, a teacher's guide, blackline masters (book and CD), student assessment folders, an assessment procedures overview card, a training DVD, and an online management tool. The DRA2 has three sections that assess reading engagement, oral reading fluency, and comprehension. Directions and procedures are scripted, and it is necessary to administer all three sections to derive estimates of instructional level, reading strengths, and weaknesses. In general, the procedures are as follows: (a) students fill out a reading engagement survey, (b) oral reading is scored on accuracy and rate, (c) students retell a story or answer six types of comprehension questions, and (d) teachers analyze students' responses to determine level of performance and strengths

and weaknesses. Estimates of likely administration durations are not in the manual. The reviewers estimate that administration of all three sections is likely to take at least 30 to 45 minutes.

The Reading Engagement section is a survey of the student's engagement in reading. Reading engagement is defined as how often a student reads, the student's knowledge of books, and reading goals defined by the student. When surveying a student, the teacher may ask with whom the student reads, what stories are preferred, and what or how many books have been read within the past 2 months. Reading engagement level is determined by summing two scales: the *Wide Reading Scale* and the *Self-Assessment/Goal Setting Scale*. The result classifies the student's reading engagement into one of four ordinal levels, which are intervention, instructional, independent, and advanced.

Administration of the Oral Reading Fluency section follows the Reading Engagement section. The administrator preselects the text from two to three books that are part of the benchmark assessment books. The administration and scoring procedures for this section differ by the age of the examinee. For the youngest students, the teacher models appropriate reading so that the student can hear the pattern of the text. Students in the middle grades predict the outcome of the book based on illustrations and are then required to read the entire text. Upper grades need only to read a prespecified passage in the chosen text. The evaluation of performance for younger students depends on phrasing, monitoring and self-corrections, problem-solving unknown words, and accuracy. For older students, it depends on expression, phrasing, rate, and accuracy. The combined scores of each of those respective components compose the Oral Reading Fluency score. Assessment of the student occurs at successively lower levels if his or her performance is below that of text- or age-level benchmarks.

The Comprehension/Printed Language Concepts is the final section administered. Younger students are administered the Printed Language Concepts section to assess one-to-one word matching and directionality. The Comprehension section assesses story retell and story understanding. Procedures for this section also vary according to the age of the student. Depending on age, the initial task is to read either

orally or silently from a text. The student then provides an oral retell of the story, orally responds to comprehension questions, writes a summary, and/or writes responses to comprehension questions. Actual procedures are more extensive than described here and are specific to each age level. Responses are judged against the main ideas, key facts, characters, and events.

Scoring

The definition of oral reading rate is words read correctly per minute using guidelines from the Teacher Observation Guide. The Teacher Observation Guides are specific to age and achievement levels, which correspond with the respective texts in the benchmark assessment books. The guide includes directions and scripts for administration and is also used to record notes from the administration and regarding student performance and scores. Benchmarks for rate depend on the level of the text used during administration and whether the text was fiction or nonfiction. Benchmarks place students at the intervention, instructional, independent, or advanced level. Performance levels achieved based on the oral reading rate benchmarks also serve as overall scores for the Oral Reading Fluency measure.

Miscues, which are uncorrected mistakes, are also observed and counted to determine accuracy. Examples of mistakes considered to be miscues include substitutions, omissions, reversals, insertions, and repetitions.

To score the Comprehension section, the teacher must choose the “best description of the student’s performance on each indicator” (p. 24) and then sums these scores to calculate a comprehension score that ranges from 1 to 4. Again, students can be placed in one of four overall reading levels: a 1 corresponds to the intervention level, a 2 is the instructional level, a 3 is the independent level, and a 4 equates to the advanced level.

To score at the instructional level for the overall placement, the total Oral Reading Fluency or Comprehension must be at the instructional level on the continuum (a chart that indicates the student’s level of performance based on the three sections of the assessment). To score at the independent level, Oral Reading Fluency and Comprehension must be at the independent level. To score at the advanced level, both Oral Reading Fluency and Comprehension must be at the advanced level. Reading Engagement does not appear to have an impact on the overall score for a student’s established reading level.

Technical Adequacy

Reliability

The technical manual includes estimates of internal consistency, parallel equivalency reliability, test–retest reliability,

and interrater reliability. The sample used to determine the technical adequacy of the DRA2 consisted of 1,676 students in kindergarten through eighth grade. Analyses were conducted in spring 2006. No other details about the group were provided in the technical manual.

The developers disaggregated reliability coefficients by reading level. Reliabilities between Oral Reading Fluency and Comprehension at all levels were moderate to high, ranging from .50 to .80. Passage equivalency within reading levels was examined with a MANOVA. The developers determined that there was equivalency across passages because the greatest total variability was found within passages, not between passages. There were no significant differences in difficulty between passages across age levels. Correlation coefficients for test–retest reliability were all above .90, indicating consistent evaluations across administrations. To investigate interrater reliability, 26 independent raters administered the DRA2 to 30 students. Raters were individuals identified as experienced in administering the DRA2. Interrater reliability was examined for the Oral Reading Fluency and Comprehension sections of the test. Raters agreed 66% and 72% of the time on Oral Reading Fluency and Comprehension, respectively. To correct for chance agreement, Gwet’s kappa was calculated, yielding coefficients of .57 and .65. The developers reported that these coefficients demonstrate moderate and substantial interrater reliability, respectively. Comparing raters to expert raters, Gwet’s kappa coefficients were .58 and .72 for Oral Reading Fluency and Comprehension, respectively. Interrater reliability was not established for the Reading Engagement section.

Validity

The technical manual includes details on face validity, criterion-related validity, including concurrent and predictive validity, and construct validity. Face validity was reported by teachers and reading specialists. A total of 66 individuals were asked to rate how well the DRA2 measured the construct of reading. Specifically, they rated the instrument on accuracy in measuring students’ reading levels, sensitivity to students’ growth in reading performance, incorporation of appropriate components of fluency and comprehension, and informing instruction for individual students. Professionals reported high face validity, indicating that the DRA2 appropriately and fully covers the domain of reading from their perspective. Individuals also reported high usefulness of the assessment.

Criterion-related validity can be broken into two categories: concurrent and predictive. Concurrent validity was determined by comparing students’ scores on the DRA2 and on other standardized reading tests taken within 1 week of each other. No significant differences between scores on

the DRA2 and the comparison tests were found. Correlations between test scores were in the .60 to .70 range. Predictive validity was calculated by comparing teacher ratings to students' scores on the DRA2, yielding coefficients of .63 and .60 for Oral Reading Fluency and Comprehension, respectively.

Construct validity was calculated to determine the extent to which the DRA2 is measuring fluency and comprehension. Interitem and subtest correlations were calculated. Items in the Oral Reading Fluency section and the overall score were compared and resulted in a high correlation of .78. Items in the Comprehension section and the overall score were compared and also resulted in a high correlation at .89. Across the subtests, results for construct validity were low to moderate at .41. The developers then carried out a factor analysis, in which fluency and comprehension emerged as distinct factors at all levels. Validity was not reported for the Reading Engagement section of the test.

Assessment for Effective Intervention

The DRA2 offers a reading assessment that appears to be favored by school personnel, as evidenced by face validity findings. The test is designed for teacher administration, which reduces the need for specialized personnel. It also allows the teacher to observe student performance within standardized conditions, which can lead to an improved understanding of individual student needs and achievement. Another positive element is that the DRA2 was developed for use across nine grades (K–8). The DRA2 measures two important components of reading and adds the unique measure of what the developers describe as “reading engagement.” Finally, the developers report convincing evidence for face validity and construct validity to identify reading difficulties; however, there is no evidence to support some of the recommended applications, such as progress monitoring or identification of specific strengths or weaknesses.

Although the DRA2 offers a number of desirable elements, there are areas in which the assessment can be less positively critiqued. Overall, the DRA2 can be utilized to determine reading level within the texts of the DRA2 materials. However, the procedures are detailed and specific to subtests. Moreover, those detailed procedures are varied and specific to the age and development range of the student. It seems that this test requires substantial practice and training to ensure the correct standardized procedures are used and that results are accurate. It is clear that multiple administrations within subtest are required for some students to ensure they are assessed at the correct level. Although administration of the entire battery is likely to take at least 30 to 45 minutes, there are no estimates of likely administration time within the manual, which is a limitation. If the estimates of the reviewers are accurate,

then the inefficiency of assessment may limit the feasibility for use as either a large-group screening or a progress-monitoring tool.

In terms of technical adequacy, reliability coefficients were not sufficient for this assessment to be used to make high-stakes educational decisions. It is recommended that reliability coefficients be .90 or higher to use an instrument to inform decisions (Salvia, Ysseldyke, & Bolt, 2007), which is a stated purpose of the DRA2. The assessment might still be used for low-stakes decisions and to guide instruction.

The Reading Engagement section represents an area of reading that is not commonly measured. Understandably, there are limited standards to which to compare the procedures, but they seemed subjective for a standardized assessment. Estimates of validity and reliability were not reported in the technical manual. Similarly, open-ended comprehension questions that must be graded by the teacher are often graded with subjectivity, and the procedures for scoring seem potentially time-consuming.

The developers of the DRA2 have also created their own benchmarks without providing explanations for their development. Those benchmarks are based on rate as well as the type of text, either fiction or nonfiction, which is not a common practice. The developers propose that DRA2 outcomes inform instruction; however, results categorize students into the levels of intervention, instructional, independent, and advanced. It is not clear how such categorizations are used to guide instruction or how they align with the curriculum, which pose challenges for applications of formative evaluation (Salvia et al., 2007). Scores that place students at these levels seem subjective. Teachers must decide if a student has demonstrated the described skill, but the target skills are sometimes ill defined and less than comprehensive. For instance, scoring the Comprehension section is particularly difficult because scoring depends on the level at which students are reading and which of the six types of comprehension questions are asked (e.g., literal, inferential, evaluative). Finally, the purported purpose of identifying strengths and weaknesses seems subjective. Teachers determine strengths and weaknesses based on their own interpretation of the scores obtained and are supposed to decide on three to five focuses for future instruction without specific instructions for teachers to make this decision.

Because of the difficulty of administration that is evident from the technical manual, along with the numerous scales and forms, this assessment is likely to require substantial training and practice to establish proficiency at administration. Moreover, the tasks are likely to require a fairly substantial amount of time to administer. An assessment that requires extensive training, practice, and time might have limited application in schools. In addition,

it does not meet the intended purposes of identifying strengths and weaknesses with objectivity. Neither the development, administration procedures, nor available evidence seem to support large-group screening or progress-monitoring applications. There appears to be little evidence that the information gleaned from this assessment could be used in the intervention design process. Similarly, without application as a progress-monitoring tool, determining the effectiveness of an intervention is not possible.

This assessment allows for added teacher attention for individual students across nine grades. The DRA2 also incorporates the unique and intriguing measurement of reading engagement. However, low reliability scores and some lower validity scores suggest that the DRA2 is not appropriate for its intended purposes.

Reference

Salvia, J., Ysseldyke, J. E., & Bolt, S. (Eds.). (2007). *Assessment in special and inclusive education*. Boston: Houghton Mifflin.