
April 1981

Reading Assessment: Formal and Informal

Priscilla A. Drum

University of California at Santa Barbara

Follow this and additional works at: <https://scholarworks.gvsu.edu/mrj>

Recommended Citation

Drum, Priscilla A. (1981) "Reading Assessment: Formal and Informal," *Michigan Reading Journal*: Vol. 15 : Iss. 1 , Article 5.

Available at: <https://scholarworks.gvsu.edu/mrj/vol15/iss1/5>

This Other is brought to you for free and open access by ScholarWorks@GVSU. It has been accepted for inclusion in Michigan Reading Journal by an authorized editor of ScholarWorks@GVSU. For more information, please contact scholarworks@gvsu.edu.

inferential comprehension. Just what might constitute stronger and richer previews and whether or not they facilitate inferential comprehension are matters for further research.

The results with respect to the effects of *grade* and *ability* require only one comment, and the comment is tangential to the main thrust of the study. Differences within grades are frequently very large; here the ability differences were quite large, while there were no differences between grades. As teachers, we must frequently take this fact into account in teaching and making assignments.

In conclusion, we would simply say that the present study provides reasonable evidence for the effectiveness of previewing. We hope that future studies will support this finding, go on to further explore the effects of previewing, and go on to validate—or invalidate—other common teaching practices.

REFERENCES

1. Anderson, R.C. "The Notion of Schemata and the Educational Enterprise." In R.C. Anderson, R.J. Spiro, & W.E. Montague (Eds.) *SCHOOLING AND THE ACQUISITION OF KNOWLEDGE*, Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.
2. Aulls, M.W. *DEVELOPMENTAL AND REMEDIAL READING IN THE MIDDLE GRADES*. Boston, Mass.: Allyn and Bacon, 1978
3. Bransford, J.D., & Johnson, M.K. "Contextual Prerequisites for Understanding: Some Investigations of Comprehension and Recall." *JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR*, 1972, 11, 717-726.
4. Clark, H. H., & Haviland, S. E. "Comprehension and the Given-new Contract." In R.O. Freedle (Ed.), *DISCOURSE PRODUCTION AND COMPREHENSION*. Norwood, N. J.: Ablex Publishing, 1977.
5. Graves, M.F. "Validating Teaching Procedures Designed to Facilitate Secondary Students' Comprehension of Content Area Texts." *RESEARCH IN READING IN SECONDARY SCHOOLS*, 1979, 2(2), 1-15.
6. Graves, M.F., & Bender, S.D. "Preteaching Vocabulary to Secondary Students: A Classroom Experiment." *MINNESOTA ENGLISH JOURNAL*, 1980, 10(2), 27-34.
7. Graves, M. F., & Clark, D. L. The Effect of Adjunct Questions on High School Low Achievers' Reading Comprehension." *READING IMPROVEMENT*, in press.
8. Graves, M. F., Palmer, R.J., & Furniss, D.W. *STRUCTURING READING ACTIVITIES FOR ENGLISH CLASSES*. Urbana, Illinois: National Council of Teachers of English, 1976.
9. Harris, A. J., & Sipay, E. R. *HOW TO INCREASE READING ABILITY*. New York: David McKay, 1975.
10. Rumelhart, D. E., & Ortony, A. "The Representation of Knowledge in Memory." In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *SCHOOLING AND THE ACQUISITION OF KNOWLEDGE*. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1977.
11. Smith, F. *UNDERSTANDING READING*. New York: Holt, Rinehart and Winston, 1978.
12. Spache, G. D., & Spache, E. B. *READING IN THE ELEMENTARY SCHOOL*. Boston: Allyn and Bacon, 1977.
13. Thorndyke, P. W. "Cognitive Structures in Comprehension and Memory of Narrative Discourse." *COGNITIVE PSYCHOLOGY*, 1977, 9, 77-110.

Reading Assessment: Formal and Informal

Priscilla A. Drum

Priscilla A. Drum is an assistant professor and head of the Language Development and Reading Program within the Graduate School of Education at the University of California at Santa Barbara.

Teachers spend hours assessing the reading achievement of their students. Are the hours spent in testing useful? Are the tests helpful in guiding instruction? Improving performance? What uses are made of testing information? This paper will examine these questions.

FORMAL ASSESSMENT

Formal assessment refers to the use of published testing instruments, usually administered at scheduled times during the academic year. The manuals that accompany these tests provide interpretations of the scores such as

expected grade-level performance.

Two types of formal reading assessment instruments commonly used are norm-referenced tests and criterion-referenced tests. The difference between the two is mainly in intent or purpose for testing (6). A norm-referenced test interprets a score in reference to other test-takers to determine how well individuals or groups are progressing as compared with other pupils. A criterion-referenced test is constructed so that a score is interpreted as indicating what skills or knowledge of the content are known.

The question discussed here is how useful are these two types of tests for teacher decisions in the classroom.

NORM-REFERENCED TESTS

Norm-referenced reading tests provide a stable measure of performance by which comparisons can be made with the norming population (8). For instance, School District XYZ had an average reading achievement score two percentage points above that of the norming group, or John Doe performed at the 4.5 grade level though he is actually in the seventh grade. The results are

likely to be reliable for the district where one is comparing large groups. Students who perform better than they usually would are balanced by students who perform worse than usual. John, however, had a bad headache the day the test was administered; his teacher knows that he is usually one of the best readers in the class. Any individual score that is above or below the average is likely to move toward the average score if re-tested, which is why some tests report a range. Extreme deviations from a reader's true score, as the hypothetical John Doe example, are not probable but are possible, so interpretations of a score for a particular child require caution.

There are several major limitations in using norm-referenced instruments in assessing pupil performance for instructional decisions within a classroom. The first of these is that tests are built to monitor large group performance, for the grade level scores, the percentile rankings, and the stanine designations are all determined by the norming group's performance. The items retained in the final form of a test are those items where most of the "good" readers, as determined by classroom grades, I.Q. results, or performance on other tests, obtain the correct response, and "poor" readers select an inappropriate choice. Thus, item selection reflects the ability characteristics of the norm group. Any item that is missed by a high proportion of those children who do well on most of the test is deleted from the final form. As a result, all a score means is that the individual tested is more like the "good" or "bad" norm group readers. A student's score reflects general reading abilities or test-taking abilities rather than areas of strength or weakness for instructional decisions (1).

A second limitation is found in using subtest scores for instructional decisions. Many of the subtests—vocabulary, comprehension, auditory discrimination, syllabication, etc.—are included so that particular abilities or deficits can be noted. However, children who do well on vocabulary are also likely to do well on comprehension. This result is not just determined by a general, pervasive ability factor, but because the item in subtests are selected by general ability perfor-

mance in the same manner as in the test as a whole. Percent correct on any item usually varies from .25 to .75 (4). Items that could indicate individual differences within subtests are not retained. Most subtest scores are highly correlated with one another. A high correlation means that the score on one subtest is likely to be similar to the score on another subtest. Even moderate correlations mean that one must hesitate before deciding that a child knows one skill but not another. (9).

Another limitation is that the items selected for tests and for subtests may or may not reflect a particular school's curriculum (10). If test items are not similar to what has been taught, then success or failure on the test does not evaluate the quality of instruction. Teaching to the test may result. Test items, of necessity, represent a small set of possible items and are deliberately chosen to represent an average range of difficulty. Basing instruction on test items will thus limit the curriculum without any guarantee that the test is representative of the skills needed to develop competent reading skills.

The criterion for success on normed tests is determined by the performance of the norm group, which may or may not be similar to the school or district group tested. Generally, norm comparisons as to demographic variables are made; but comparisons based on instructional history and prior achievement are rarely considered. A class where the average reading level is third grade should not be tested using an instrument where the norm sample has an average reading level of ninth grade even if the mean chronological age of both groups is fourteen years. This testing instrument is intended for ninth grade readers. Even with extrapolation the results will not provide accurate reading levels for the lower-achieving class.

These tests cost money and some extensive batteries can take many hours to administer. If there are alternative forms for pre-and post-testing, the cost in both money and time will double. The information obtained should be weighed against the time lost in instruction and in practice reading.

Appropriate norm-referenced tests are useful at the district level to establish general levels of perfor-

mance as compared with the nation and to monitor changes in performance by school and district. They are not useful for instructional decisions, nor for aiding individual students.

CRITERION-REFERENCED TESTS

Criterion-referenced reading tests do establish performance levels for specific instructional tasks (5). Instead of measuring comparative performances, a mastery level of percentage correct on each skill tested is established by external standards (11). The external standards are determined by content experts in the domain tested, who analyze the abilities needed to perform a complex task such as reading. Then tests are constructed in which all items in a subtest are intended to pertain to one ability or one objective (3).

Basically, the purpose for criterion-referenced tests is a step in the direction of providing useful information for instruction. Test results indicating the objectives a student has mastered can be inserted in "cum" folders to aid next year's teacher in planning for instruction. It is possible to match general school and class objectives to test items (7). Differences between test content and school curricula can be minimized.

However, criterion-referenced tests also cost money and take time to administer and most published criterion-referenced tests cover a number of instructional objectives. Thus, they look quite similar to norm-referenced tests with subtests, though both easy and hard items are more likely to be found in criterion tests. However, the purpose of these tests is quite different. Criterion tests are intended to match students to different instruction by patterns of performance on subtests. The performance levels on any subtest can be used for instructional decisions, but there are generally moderate to high correlations for the different performance levels. The objectives are not independent of one another and have only very limited value for diagnostic purposes. A child who does poorly on one objective is also likely to do poorly on other objectives.

The results for each objective are summed over items. It is only at the item level that a teacher can tell ex-

actly what it is the child can or cannot do. But identifying response patterns by individual items is time-consuming. If a child misses six of the ten items used for measuring knowledge of syllabication rules, the teacher cannot just accept the fact that this child has not yet reached mastery on syllabication. Instead, a comparison must be made of what is known (four items) and what needs to be taught.

The higher level objectives include many of the lower level skills. A child who is required to answer comprehension questions on passages must have some word recognition skills, vocabulary knowledge, and grammatical understanding. Children who can perform well on higher level objectives either do not need instruction on the more basic skills, or the more basic skills may not actually be prerequisites for the advanced skills. Perhaps the best approach would be to test children on the highest level of skills deemed reasonable at a particular grade. Those children who are successful would need no further testing other than their daily performance on more advanced instructional material. For those who fail at even a simple task, such as matching letters, further probing is necessary.

Any test requires that the child understand the directions for what is to be done and how to mark and answer. The test scores should reflect knowledge of the content tested, not faulty understanding of the instructions. One solution for this problem would be to provide sufficient sample items so that the children tested can establish perfect performances, thereby indicating they understand the task. Then increase the difficulty of the items until different performance levels are obtained. Of course, performance at successive levels could be related more to children's ability to persevere than to their skill. Children may drop out simply because they are tired of doing the task. If testing is carried out individually or in small groups with only those students who couldn't perform the higher level tasks, then the teacher can monitor the performance and ask questions about answer choices. Again the teacher must note the item level of performance. Formal tests usually provide one or two sample items to clarify in-

structions. These are often done in concert with the teacher, and then the children proceed on their own. The many subtests of a criterion-referenced battery can cause confusion as a child moves from one type of activity to another. Again, the teacher can ameliorate this situation by administering only one mastery test at a time.

Successful use of criterion-referenced tests to guide instructional decisions requires time and effort. The teacher must analyze each child's error pattern in order to decide what should be taught. Reassessment of the same skill is needed to insure that the pupils have truly mastered a particular ability. Also, the items available must be related to the class instruction and the materials used. The cost of having appropriate and sufficient item pools on hand to administer whenever an instructional decision must be made is high. For most teachers, the necessity of deciding how to help individual students is a daily task.

INFORMAL ASSESSMENT

Informal assessment refers to teacher-made or teacher-selected tasks used to evaluate pupil knowledge and/or ability on a specific instructional unit. These are administered on an ad-hoc schedule whenever a decision must be made about what to teach and to whom. The procedures for selecting tasks and administering them are similar to the criteria given for optimal use of criterion-referenced items. The difference is that the tasks are always available within the classroom materials used for instruction.

What is done day to day in class forms the basis for continuous assessment. For instance, one general objective might be to increase knowledge of word meanings. A list of words is presented to the class with a matching group frequently associated with each word. Children must identify the primary meaning of each word. Those children who can't do this task are provided the basic meaning for each word and given practice in using each word in the appropriate context.

The children who can do the original task are asked to explain the different meanings when the word is used in different contexts. Those who recognize the similarities

and differences of the various meanings are then expected to use the words in speech or in writing. Thus assessment and instruction proceed together. The approach systematically increases difficulty from matching, to recognizing similarities and differences, to production. There has been no interruption in instruction. Each piece of information is used for an immediate decision as to what happens next.

Any one or all of the vocabulary assessments can be crossed with conditions where the rest of the class is working silently on some task or where there are several discussions occurring throughout the class. The teacher can thus note attention to primary task and gain some idea of the students who are able to attend to a task despite noise and those who will need close monitoring. As a result of this information, the teacher will be better able to plan instructional conditions—grouping and supervision—that will fit the capacities of the various students.

Speed can be crossed with accuracy for any assignment. Students who finish first and also are accurate are likely ready for more advanced assignments. Slow but accurate students may be learning more or may still be struggling to apply principles learned. In any case, assignments for such students take longer. Knowing this the teacher can provide more practice until fluent mastery is obtained. Fast but inaccurate performance indicates inattention or lack of learning. Probe questioning can clarify the reasons for inadequate performance. By noting both rate and correctness, teachers can obtain diagnostic information and direction for instruction.

The value of on-going informal assessment is that it can be done with any instruction materials for every educational objective. Of course, mistakes will be made. Students may be mis-evaluated on a particular skill, but this lack of reliability can be corrected by reassessment in subsequent instructional units. No decision becomes part of a permanent record.

The materials for the assessment come from the instructional program. It does take time to think through the objectives for a class, deciding which ones are appropriate goals. The tasks used to meet an objective need to be classified as easy or hard. The

guiding principle is what does an error mean given the particular materials, the responses required, and the classroom conditions. More details about these procedures are presented in *Theory and Practice of Early Reading* (2).

The assumption that students differ in prior knowledge and in skills is basic to this approach. Pupil performance in each curricular unit is evaluated, so permanent diversions into ability groups is much less likely. The students will be re-grouped by their responses. Some students may have extensive vocabulary knowledge but need extensive help on spelling. Their performance determines whether they can work independently or will need close supervision. Most help, including further assessment, will be supplied to those students who need it for a particular objective. Students are presented with harder tasks only when their responses are both accurate and fast on the present work. Thus, no student is pushed to do work where failure is likely.

The cost in money is slight; the cost in teacher in time is great. The teacher needs to analyze the tasks for every objective, to plan for different difficulty levels of performance for each objective, to note student responses at each level, and to record these responses. However, the rewards of knowing what each student can do and what instruction helped the student are also great. Instructional decision-making is

returned to the teacher, the one who knows most about the students and the one who is responsible for their instruction.

In summary, both formal and informal assessment procedures provide useful information but for different purposes. Norm-referenced reading achievement tests are the most reliable indicators of group progress over time. Teacher-selected assessment tasks obtain the information needed to guide instruction for individual pupils.

REFERENCES

1. Calfee, R.C., Drum, P.A., & Arnold, R.D. "What research can tell the reading teacher about assessment." In Samuel, S.J. (Ed.). *WHAT RESEARCH HAS TO SAY ABOUT READING INSTRUCTION*. Newark, Delaware: International Reading Association, 1978.
2. Calfee, R.C., & Drum, P.A. "How the researcher can help the reading teacher with classroom assessment." In Resnick, L.B. & Weaver, P.A. (Eds.), *THEORY AND PRACTICE OF EARLY READING*, V.2. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1979.
3. Davis, F.B. "Criterion-referenced tests: A critique." In Blanton, W.E., Farr, R., & Tuiman, J.J. (Eds.), *MEASURING READING PERFORMANCE*. Newark, Delaware: International Reading Association, 1974.
4. Drum, P.A., Calfee, R.C., & Cook, L.K. "Effect of surface structure variables on reading comprehension test performance." Manuscript submitted for publication, 1980.
5. Glaser, R. & Nitko, A.J. "Measurement in learning and instruction." In Thorndike, R.L. (Ed.), *EDUCATIONAL MEASUREMENT*. Washington, D.C.: American Council on Education, 1971.
6. Mehrens, W.A., & Lehmann, I.J. *STANDARDIZED TESTS IN EDUCATION*. New York: Holt, Rinehart, and Winston, 1980.
7. Popham, W.J. *EDUCATION EVALUATION*. Englewood Cliffs, New Jersey: Prentice Hall, 1975.
8. Slavia, J.A., & Ysseldyke, J.E. *ASSESSMENT IN SPECIAL AND REMEDIAL EDUCATION*. Boston: Houghton-Mifflin, 1978.
9. Thorndyke, R.L. "Dilemmas in diagnosis." In MacGinitie, W.H. (Ed.) *ASSESSMENT PROBLEMS IN READING*. Newark, Delaware: International Reading Association, 1973.
10. Venezky, R.L. *TESTING IN READING: ASSESSMENT AND INSTRUCTION DECISION-MAKING*. Urbana, Illinois: National Council of Teachers of English, 1974.
11. Womer, F.B. "What is criterion-referenced measurement?" In Blanton, W.E., Farr, R., & Tuiman, J.J. (Eds.), *MEASURING READING PERFORMANCE*. Newark, Delaware: International Reading Association, 1974.

Evaluating a Fledgling Reading Program

Susan R. Enke

*Susan R. Enke is a reading specialist
and English instructor for the
Roseville, Michigan Community Schools.*

A high school reading program, newly born and facing a new decade of financial cutbacks and declining enrollment, is a creature needing great nurturing to meet the increasing demands of students entering high school with inadequate reading skills to meet graduation requirements. In the Roseville, Michigan, high school, for example, the hard reality is a single reading teacher for a student population of 1094, over half of whom indicate a

need for reading skills development, and a single reading laboratory large enough to accommodate only twenty students at a time. One hardly dare call this fledgling a reading program. It would be more accurate to view this bare beginning as growth potential via an evaluation process that would pose straightforward questions: What is right (or wrong) with the reading curriculum that presently exists? What seems to be working

(or not working)? What needs change? What components are completely lacking?

The following guidelines, including needs assessment, goal-setting, and criteria development for a program and its evaluation, are helpful in promoting valid and workable answers to these questions.

NEEDS ASSESSMENT

A needs assessment clarifies four important elements: the problem,