DOCUMENT RESUME

ED 261 074 TM 850 474

AUTHOR Manning, Maryann; And Others
TITLE A Comparison among Measures of Reading Achievement with Low Income Black Third Grade Students.
PUB DATE Mar 85
NOTE 26p.; Paper presented at the Annual Meeting of the American Educational Research Association (69th, Chicago, IL, March 31-April 4, 1985).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Ability Grouping; *Achievement Rating; Black Students; Cloze Procedure; *Comparative Testing; Correlation; Criterion Referenced Tests; Grade 3; *Informal Reading Inventories; Low Income Groups; Norm Referenced Tests; Primary Education; *Reading Achievement; *Reading Tests; Scores; *Student Evaluation; Teacher Attitudes; Test Validity; Urban Schools
IDENTIFIERS Alabama Basic Competency Tests; California Achievement Tests; Classroom Reading Inventory; Houghton Mifflin Informal Reading Inventory; Teacher Expectations

ABSTRACT

The results of different types of reading achievement measures were compared for 58 low-income urban black third graders. Two formal tests were administered: the norm-referenced California Achievement Tests (CAT), and the criterion-referenced Alabama Basic Competency Test (ABCT). Informal measures included the Houghton-Mifflin Informal Reading Inventory (HMIRI), the Classroom Reading Inventory (CRI), a cloze procedure, and teacher judgment (as indicated by the basal reader assignment for each student). Results indicated that correlations among all of the measures were moderate to high. The formal tests, particularly the CAT, tended to produce lower scores than the informal measures. In spite of high correlations, the CAT and ABCT results revealed very different distributions of student ability. With the ABCT, more students showed average and above average performance. Examination of teachers' judgments regarding reading book placement, as compared to test results, indicated that teachers underestimated students' reading ability and placements did not reflect test results. HMIRI results also suggested that a number of students could have been assigned to a higher-level reading book. It was suggested that informal measures be used for book placement and that multiple measures of reading achievement be used in decision making. (GDC)

A Comparison Among Measures of Reading Achievement with Low Income Black Third Grade Students

Maryann and Gary Manning
School of Education
University of Alabama at Birmingham
Birmingham, AL 35294

Caroline B. Cody
National Association of State
Boards of Education
Alexandria, Virginia

Paper presented at the Annual Meeting of the American Educational Research Association in Chicago, Illinois, March 31-April 4, 1985

A COMPARISON AMONG MEASURES OF READING ACHIEVEMENT WITH LOW INCOME BLACK THIRD GRADE STUDENTS

This study compares the results of informal and formal reading achievement measures, norm and criterion referenced tests, and teacher opinion for low income black third grade students in a Southern city. Findings include higher informal test scores, moderate correlations between norm and criterion referenced test results, and teacher opinions reflective of standardized test scores with resulting lower book assignments than would be indicated by informal reading inventories designed to place students.

A Comparison Among Measures of Reading Achievement with Low Income Black Third Grade Students

Introduction

At the present time in the United States, reading achievement measures are highly important. In urban school districts, in particular, reading achievement has become a community issue and reading scores for districts, schools, and grade levels appear in the local newspaper. High test scores are used as incentives by real estate agents and chambers of commerce.

Achievement scores are also used for a variety of instructional and administrative purposes within school districts. For example, individual students are grouped for instruction using test data. In some cases students are promoted or not promoted on the basis of their scores on various tests. In some systems the success or failure of teachers is determined by achievement test gains made by their classes. Individual schools may receive special programs or unique allocations as a result of either low or high average scores.

In 1985, most school districts are able to show reading achievement gains; many urban school superintendents claim dramatic growth in reading achievement. Public opinion has responded to such announcements of improved test scores with new

confidence and pride, as is suggested by recent Gallup polls. Parents and business people again are beginning to have faith that graduates of the public schools will be able to read the books and materials that are vital to life outside the school.

It is important, therefore, that information about reading achievement be valid, and that schools provide some assurance that the measures being used are the best that are available.

Research Questions

This study was designed to compare the results obtained from a variety of reading achievement measures used in an urban school district with children from low socioeconomic backgrounds. The questions addressed include the following:

Do informal and formal tests provide comparable results among low income Black third graders?

Do norm-referenced tests and criterion-referenced tests provide comparable information in this same population?

Do teacher judgments as reflected in classroom decisions concerning reading placement agree with other measurement data in the target classrooms?

Review of the Literature

A review of related literature provides an extensive discussion of the comparable results of formal and informal measures. For a period of years, studies have compared the results of standardized achievement tests and informal reading inventories--both teacher made and commercial. These studies are

almost unanimous in their findings that students' norm-referenced standardized achievement test scores are higher than their informal reading inventory (IRI) results. Study after study has found that standardized tests place students one to two years higher than do IRI's (Harris and Sipay, 1980; Jones and Pikulski, 1974; McCracken, 1962; Rupley and Blair, 1979). All of the above studies have been done on hetereogeneous populations; no study focused on low-income minority children.

Reading experts for many years have recommended using informal tests for placing students in appropriate reading books (Dechant, 1981; Johns, 1977; Zintz, 1981). Although many studies have pointed out the limitations of this informal procedure for student placement, no other measure has been widely advocated. There have been contradictory findings from efforts to validate standardized achievement tests using correlations with informal inventories (Farr and Beck, 1984; Smith and Beck, 1980). At issue as well is the use of teacher-made IRI's, as opposed to those developed and distributed commercially. In general, there is some support for the use of commercially produced informal instruments as shown in the work of Jongsma and Jongsma (1981). Researchers who have studied IRI's have suggested the desirability of more frequent use of such informal measures (McKenna, 1983; Peterson, Greenlaw, and Tierney, 1978; Schell, 1982).

The use of criterion-referenced tests (CRT) has increased

with the adoption by many states and localities of minimum competency tests. In addition, many basal reading programs now provide skill management systems which include criterion-referenced tests. In many cases, however, these tests have not met the standards of validation that would justify the ways in which they are used (Lyons, 1984). There have also been efforts to analyze the usefulness of criterion-referenced tests in predicting reading performance. Horodezky and Labercane (1983) found that one basal criterion-referenced test battery was as effective as any other measure in predicting performance at grade one; as students gained proficiency in reading, however, the criterion-referenced tests appeared to be "less capable of tracking this divergence in reading skill development." This study also concluded that the most consistent predictor of reading performance was the Classroom Reading Inventory (Silvaroli, 1984).

Teacher judgment, as reflected in rating or ranking of students, or as demonstrated in classroom groupings, has been studied for reliability and for correlation with other reading measures. Although studies report reliability problems in teachers' judgments of reading competence and or disability (McKenna, 1983; Schell, 1982), most studies suggest that teacher judgments provide vital information for analyzing students' reading. Arnold and Sherry (1975) found that disabled readers were most often assigned to texts at their IRI frustration level,

that is at a level at which they were not able to read successfully. Other studies have found that teachers' judgments of instructional levels were similar to results of standardized tests (Kermoian, 1962). However, Brown (1963) found that correlations between teacher judgment and IRI results were higher than between teacher judgment and standardized test results. In general, teacher judgment tends to overestimate students' abilities to read materials and consequently, teachers tend to assign students to reading materials which are too difficult for them. Nevertheless, those judgments are at least as accurate as standardized test scores for selecting appropriate texts (Oliver and Arnold, 1978).

In summary, the literature which addresses the comparison of a range of reading measures suggests that formal standardized tests tend to place students in books which are too difficult; commercial informal measures are the best predictors of students' general reading ability. It should be remembered, however, that some IRI procedures have more limitations than others.

The Study

This study was designed to compare the results of a variety of reading achievement measures with Black, low socioeconomic status, third-grade students. The test results examined include the reading section of the California Achievement Tests (1977), the reading subtest of Alabama Basic Competency Test (1980), the

Houghton-Mifflin Informal Reading Inventory (1983), and the Classroom Reading Inventory (1984). In addition, a cloze procedure, developed from material from the third grade social studies book, which had a readability level of 3.0 using the Fry (1977) readability graph, was administered in the selected classrooms. The three classroom teachers provided information as to basal reader assignments for each child in their classes.

The questions addressed include the following: (1) Do formal tests (the California Achievement Tests and the Alabama Basic Competency Test), and informal tests (Houghton-Mifflin Informal Reading Inventory, the Classroom Reading Inventory, the cloze procedure, and teacher judgment) provide comparable results among third graders in this inner city school? Do norm-referenced tests (California Achievement Tests) and criterion-referenced tests (Alabama Basic Competency Test) provide comparable information? Do teacher opinions and classroom decisions about book placement reflect agreement with informal and formal measures?

Sample

The population for this study included all (58) third-grade students in an elementary school in a low-income neighborhood of a large Southern city. All of the students were Black: 31 boys, 27 girls. All were considered to be of low socioeconomic status as measured by the fact that they qualified for participation in the free lunch program at their school. Three teachers participated in the study.

Instrumentation for the Study

Several of the tests used as data in the study were available due to the statewide testing in Alabama which include the California Achievement Tests (CAT) and the state-developed Alabama Basic Competency Test (ABCT) which is a criterion-referenced test. The CAT reading battery provides a total score and subtest scores in the following areas: phonics, structural analysis, vocabulary, comprehension. The ABCT reading section consists of vocabulary and comprehension tests. The data from those testing programs were used in this study.

Informal testing was conducted using the Houghton-Mifflin Informal Reading Inventory (HMIRI) and the Classroom Reading Inventory (CRI). Both contain comprehension questions and are widely used for finding the instructional reading level for individual students. The HMIRI is a silent reading inventory; the CRI is an oral reading inventory.

Each teacher was asked to provide basal reader assignments for each child in her class. A form was provided by the investigators for this purpose.

Data Collection

All testing was done during the month of April, 1984. The Cat and ABCT tests were administered by the school and results were made available to the investigators. The investigators administered the two informal inventories (HMIRI and CRI) to

individual students and the cloze procedure to each class during that month.

Data Analysis

In order to compare data from the various instruments, below average, average and above average ranges were established for each set of data. Frequencies and percentages were established as a basis for comparison. The Pearson Product-Moment Correlation was used to examine the correlations among the various test scores. Correlations at both the moderate (.40 - .69) and high (.70 and higher) were considered in the analysis of the data. Histograms were prepared to show relaionships among selected data sets.

Results of the Study

Do informal and formal tests provide comparable results with this population? The results of this study indicate that correlations among all measures are in the moderate to high range (see Table 1). The highest correlations were between the reading portion of the CAT and the cloze (.70), and between the reading portion of the CAT and the CRI (.75).

Insert Table 1 about here

Table 2 illustrates the difference in distribution of scores in all measures. This table should be used to compare the CAT

scores with each other set. The per cent of students whose scores fell in the below average range on the CAT reading portion (5th stanine) exceeds 60%; on the other measures the scores distribute more evenly among below average, average, and above average ranges. Only on the cloze procedure does the per cent of students at the below average range exceed the per cent of students in the below average range on CAT. The per cent of students demonstrating above average achievement on the cloze is smaller than on any other measure. Unlike most other studies, this study finds that the formal tests, in particular the CAT reading portion, are producing lower scores for most students than other measures.

Insert Table 2 about here

Do norm-referented and criterion-referenced measures provide comparable results? A comparison of the results of the reading portion of the CAT and the ABCT reading subtest reveals a very different distribution in spite of correlations in the high moderate range (.5798 and .6419) as seen in Table 2. Only 3.5% of the students were below average on the word-attack test and about one-third of the students on the comprehension test. By contrast on the criterion-referenced ABCT reading subtest most students in the classes (95% on word attack; 67% on comprehension) could be said to have average and above average performance, as compared with the norm-referenced CAT reading subtest which found that only

40% of the students could be said to be performing satisfactorily. One would have to conclude that the results of the two tests are not comparable in their applicability as a basis for teacher decisions. The use of the CAT reading subtest as a measurement of reading achievement in these classrooms would provide a very different picture of students' achievement than the ABCT reading subtest.

Do teacher judgments and classroom decisions regarding reading book placement reflect agreement with other measurement data in these classrooms? The evidence from this study suggests that teacher judgment, as reflected in students' placement in books, is not in agreement with the other test data. Table 2 demonstrates the degree to which students have been placed in books below their level of competence as measured by any of these reading measures, including the CAT reading portion which provided the lowest scores for students. Tables 3 and 4 illustrate the underplacement of students. On Table 3, assuming that placement within one level of the CAT reading portion score could be considered appropriate, there are no students placed in books which are one level or more above their CAT reading subtest scores. However, 12 students were assigned to the 2.0 books whose CAT reading subtest scores would justify books one or two levels higher. At the 2.5 level, 16 students could be predicted by their CAT reading scores to read books more difficult. At the 3.0 level, 10 students could be predicted by the same test to read more difficult books. No students were assigned to books above

the first reader of third grade in spite of the fact that 11.9% of the students scored above the 5th stanine on the third grade form.

Insert Table 3 about here

Table 4 shows the differences between the results of the HMIRI and book placement. The HMIRI is designed to place students at appropriate levels in the basal reading program used in these three classrooms. At each book level, no readers seem to be reading books that their HMIRI scores would indicate are too difficult. However, at each level, about half the students in the group, on the basis of their HMIRI scores, are assigned to readers which are too easy for them: 12 students assigned to 2.0 level books, 11 students assigned to 2.5 level books, and 7 students assigned to 3.0 level books. In the highest group, 3 students have scores which justify their placement in books two grade levels higher and 4 students have scores which would place them in books three grade levels higher.

Insert Table 4 about here

Summary and Discussion

Unlike other studies (see p. 4), this study finds that teacher judgment about book placement underestimates students' reading ability and is not reflective of the data from reading measures. The specific findings from this study include the following:

1. With this population, there is considerable variablility in the scores earned on different reading measures. This finding suggests using multiple measures of reading achievement, particularly in light of the way reading achievement scores are used to make decisions about students' work and school success.
2. In this study, the use of informal measures, particularly for book placement, is desirable since those measures appear to demonstrate students' highest levels of competence.
3. Lack of transfer of reading competence to subject matter books, as seen by the results of the cloze procedures, needs further investigation.
4. Student placement in basal readers by teachers is lower than any of the measures used would justify. This finding raises questions about teacher judgment and level of expectation for students in this population.
5. Additional research with larger populations of this demography are needed.

Implications

The findings of this study indicate that this population of students can read better than their teachers think they can read.

Since the original Rosenthal and Jacobson (1968) study in the 60's, few educators have disagreed with the concept that teacher expectancy affects achievement (Good and Brophy, 1977; Shavelson, 1983). Teacher expectancy has been shown to influence many teacher behaviors in the classroom, including seating patterns, attention to low achievers, etc. (Good, 1981; Good and Brophy, 1977).

It is in the area of reading instruction where the development of higher order thinking skills takes place that expectancy studies take on great significance. Studies have shown that perceived high achievers are questioned in different ways, given longer wait times to answer, and given clues and repeated or rephrased questions (Good and Brophy, 1977). High achievers are criticized more and helped to draw relationships about what is being read. Perceived poor readers are praised more, and are interrupted more often. Teachers are more likely to supply a correct word or a graphophonic clue for poor readers whereas good readers are more likely to be given context or meaning-related clues (Allington, 1980; Gumperz and Hernandex-Chavez, 1972; Pflaum, 1972; Weinstein, 1976).

If, indeed, the teachers in this study were convinced by more comprehensive data that their students were better readers, it is conceivable that reading instruction would be conducted differently in these classrooms and that the cycle of improved comprehension scores could be established. The use of a variety of measures, including an IRI, at the beginning of the school year before teachers have developed their initial expectations, might

influence their organization of reading groups and the books to which those groups would be assigned. More accurate initial placement is important because there is evidence that although flexibility in group placement is advocated by reading educators, there is in actuality little mobility among reading groups after the initial grouping is assigned (Resnick, 1981; Weinstein, 1976). Change in basal reader or mobility of grouping is discouraged by the large amount of anciliary material which students must often complete in order to proceed into another level reader.

Much further study is needed, not just about how various measures used to assess reading correlate with each other and children's book placement, but about the consequences of modifying teachers' expectations and book assignments (based on IRI's) on children's reading and thinking progress.

TABLE 1

Pearson Correlations for All Variables

	CAT Phon.	CAT Struc.	CAT Com.	CAT Tot.	ABCT Voc.	ABCT Com.	Cloze	CRI Voc.	CRI Com.	HMIRI
CAT Voc.	.4609	.5566	.5754	.7969	.4857	.4587	.6166	.6208	.4774	.4179
CAT Phon.	-	.4119	.6114	.8105	.4142	.5380	.5266	.5960	.4118	.4072
CAT Struc.	-	-	.4971	.6865	.4801	.5485	.5233	.4541	.3911	.4289
CAT Com.	-	-	-	.8631	.5350	.5823	.5767	.6544	.5941	.5603
CAT Tot.	-	-	-	-	.5798	.6419	.7083	.7532	.5787	.5688
ABCT Voc.	-	-	-	-	-	.5531	.5654	.5353	.5716	.4633
ABCT Com.	-	-	-	-	-	-	.4207	.4260	.5065	.3322
Cloze	-	-	-	-	-	-	-	.6325	.6022	.5751
CRI Vbc.	-	-	-	-	-	-	-	-	.5860	.6006
CRI Com.	-	-	-	-	-	-	-	-	-	.6605

TABLE 2

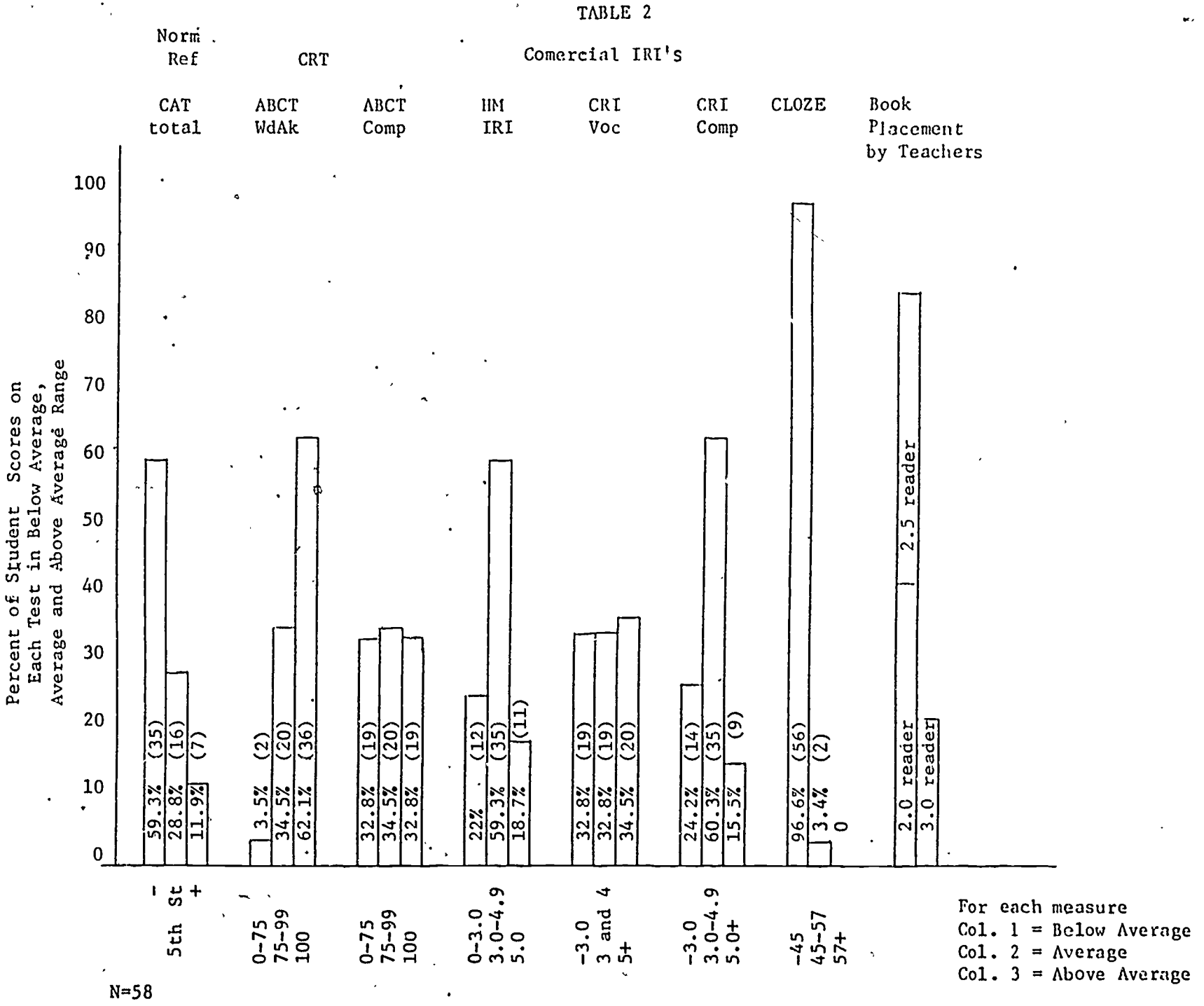
Norm Ref
CRT
Comercial IRI's
CAT total
ABCT WdAk
ABCT Comp
HM IRI
CRI Voc
CRI Comp
CLOZE
Book Placement by Teachers
Percent of Student Scores on Each Test in Below Average, Average and Above Average Range
100
90
80
70
60
50
40
30
20
10
0
59.3% (35)
28.8% (16)
11.9% (7)
3.5% (2)
34.5% (20)
62.1% (36)
32.8% (19)
34.5% (20)
32.8% (19)
22% (12)
59.3% (35)
18.7% (11)
32.8% (19)
32.8% (19)
34.5% (20)
24.2% (14)
60.3% (35)
15.5% (9)
96.6% (56)
3.4% (2)
0
2.0 reader
3.0 reader
2.5 reader
- 5th St +
0-75 75-99 100
0-75 75-99 100
0-3.0 3.0-4.9 5.0
-3.0 3 and 4 5+
-3.0 3.0-4.9 5.0+
-45 45-57 57+
N=58
For each measure
Col. 1 = Below Average
Col. 2 = Average
Col. 3 = Above Average

Table 3

Distribution of Students Using Book Placement and Differences Between CAT Reading Total and Book Placement

CAT Differences

Book Placement By Grade Level	-.6 -1.0	-.01 -.5	Equal	+.1 +.5	+.6 +1.0	+1.1 +1.5	+1.6 +2.0	+2.1 +2.5	+2.6 +3.0	+3.1 +3.5
2.0	1 1.7%	1 1.7%	0 0%	9 15.5%	4 6.9%	5 8.6%	3 5.2%			
2.5				7 12.1%	11 19.0%	4 6.9%	1 1.7%			
3.0				2 3.4%	5 8.6%	4 6.9%	1 1.7%			

N=58

TABLE 4

Distribution of Students by Book Placement and Differences Between Book Placement and IRI Results

Book Placement By Grade Level	Grade Level Book Placement ± IRI Score									
	-1.0	-.5	Equal	+.5	+1.0	+1.5	+2.0	+2.5	+3.0	+3.5
2.0		5 8.6%	4 6.9%	2 3.4%	7 12.1%	5 8.6%				
2.5			2 3.4%	10 17.2%	7 12.1%			2 3.4%		2 3.4%
3.0				5 8.6%			3 5.2%		4 6.9%	

N=58

References Cited

Alabama Basic Competency Tests. (1980) Alabama: Montgomery, Alabama: Alabama State Department of Education.

Allington, Richard L. (1980). Teacher interruption behaviors during primary-grade oral reading. Journal of Educational Psychology, 72, 371- 377.

Arnold, R.D., & Sherry, N. (1975). A comparison of the reading levels of disabled readers with assigned textbooks. Reading Improvement, 12, 207-211.

Brown, A.A. (1963). A comparison of five widely used standardized reading tests and an informal reading inventory for a selected group of elementary children. Unpublished doctoral dissertation, University of Georgia, Athens.

California Achievement Tests. (1977). Monterey, California: McGraw-Hill.

Dechant, Emerald. (1981). Diagnosis and remediation of reading disabilities. Englewood Cliffs, New Jersey: Prentice-Hall Co.

Farr, Roger, & Beck, Michael. Validating the instructional reading level score of the Metropolitan Achievement Tests. Journal of Research and Development in Education, 17, 55-64.

Fry, Edgar. (1977). Fry's readability graph: Clarifications, validity, and extension to level 17. Journal of Reading, 21, 242-252.

Good, Thomas & Brophy, J. (1977). Educational Psychology: A realistic approach. New York: Holt, Rinehart, & Winston.

Good, Thomas. (1981). Teacher expectations and student perceptions: A decade of research. Educational Leadership, 38, 415-422.

Gumperz, J.J., & Hernandez-Chavez, E. (1972). Bilingualism, bidialectism, and classroom interaction. In C. B. Cazden, V.P. John, and D. Hymes (Eds.), Functions of Language in the Classroom. New York: Teachers College Press.

Harris, Albert J. & Sipay, Edward R. (1980). How to increase reading ability. New York: Longman.

Horodezky, Betty, & Labercane, George. (1983). Criterion-referenced tests as predictors of reading performance. Educational and Psychological Measurement, 43, 657-662.

Houghton-Mifflin Informal Reading Inventory. (1983). Boston: Houghton-Mifflin Co.

Johns, Jerry., Garton, Sharon, Schoenfelder, Paula, & Skriba, Patricia. (1977). Assessing reading behavior: Informal reading inventories. Newark, Delaware: International Reading Association.

Jones, Margaret B., & Pikulski, Edna C. (1974). Cloze for the classroom. The Reading Teacher, 17, 432-38.

Jongsma, Kathleen, & Jongsma, Eugene A. (1981). Test review: Commercial informal reading inventories. The Reading Teacher, 34, 697-705.

Kermoian, S.B. (1962). Teacher appraisal of first grade readiness. Elementary English, 39, 196-201.

Lyons, Kevin. (1984). Criterion-referenced reading comprehension tests: New forms with old ghosts. Journal of Reading, 27, 293-298.

McCracken, Robert. (1962). Standardized reading tests and informal reading inventories. Education, 82, 366-369.

McKenna, Michael C. (1983). Informal reading inventories: A review of the issues. The Reading Teacher, 36, 670-679.

Oliver, Jo Ellen, & Arnold, Richard D. (1978). Comparing a standardized test, an informal inventory and teacher judgment on third grade reading. Reading Improvement, 15, 56-59.

Peterson, Joe M., Greenlaw, Jean, & Tierney, Robert J. (1978). Assessing instructional placement with an IRI; the effectiveness of comprehension questions. Journal of Educational Research, 71, 247-250.

Pflaum, S., Pascarella, Ernest T., Boswick, Mary, & Auer, Carol. (1980). The influence of pupil behavior and pupil status factors on teacher behavior during oral reading lessons. Journal of Educational Research, 74, 99-105.

Resnick, Lauren B. (1981). Introduction: Research to inform a debate. Phi Delta Kappan, 62, 623-635.

Rosenthal, R. & Jacobson, L. (1968). Pygmalion in the classroom: Teacher expectation and pupil's intellectual development. New York: Holt, Rinehart, Winston.

Rupley, William H., & Blair, Timothy R. (1979). Reading diagnosis and remediation: A primer for classroom and clinic. Boston: Houghton-Mifflin Co.

Schell, Leo M. (1982). How accurate are oral reading tests? Reading World, 22, 91-97.

Shavelson, Richard J. (1983). Review of research on teachers' pedagogical judgments, plans, and decisions. The Elementary School Journal, 83, 393-413.

Silvaroli, Nicholas J. (1984). Classroom reading inventory. Dubuque, Iowa: Wm. C. Brown.

Smith, Earl W., & Beck, Michael. (1980). Determining instructional reading level with the 1978 Metropolitan Achievement Tests. The Reading Teacher, 34, 313-319.

Weinstein, R. (1976). Reading group membership in first grade: Teacher behaviors and pupil experience over time. Journal of Educational Psychology, 68, 103-116.

Zintz, Miles V. (1981). Corrective reading. Dubuque, Iowa: Wm. C. Brown.

Bibliography

Alpert, Judith L. (1974). Teacher behavior across ability groups: A consideration of the mediation of pygmalion effects. Reading World, 66, 348-353.

Benninga, Jacques, Cuskey, Thomas, R. & Thornburg, Kathy, R. (1981). The relationship between teacher attitudes and student perceptions of classroom climate. The Elementary School Journal, 82, 67-75.

Blanchard, Jay S., Borthwick, Jr., Paul, & Hall, Ann. (1983). Determining instructional reading level: Standardized multiple choice versus IRI probed recall questions. Journal of Reading, 26, 684-689.

Bradley, John M. (1976). Evaluating reading achievement for placement in special education. Journal of Special Education, 10, 237-245.

Bristow, Page S., Pikluski, John J., & Pelosi, Peter L. (1983). A Comparison of five estimates of reading instructional level. The Reading Teacher, 37, 273-279.

Brophy, Jere E. (1983). Research on the self-fulfilling prophecy and teacher expectations. Journal of Educational Psychology, 75, 631-661.

Coleman, Margaret, & Harmer, William R. (1982). A comparison of standardized reading tests and informal placement procedures. Journal of Learning Disabilities, 15, 396-398.

Cooper, Harris M. (1979). Pygmalion grows up: A model for teacher expectation communication and performance influence. Review of Educational Research, 49, 389-410.

Cooper, Harris M., & Tom David. (1984). Teacher expectation research: A review with implications for classroom instruction. The Elementary School Journal, 85, 76-89.

Fuchs, Lynn, Fuchs, Douglas, & Deno, Stanley L. (1982). Reliability and validity of curriculum-based informal reading inventories. Reading Research Quarterly, 18, 6-26.

Good, Thomas L. (1982). How teachers' expectations affect results. American Education, 18, 25-32.

Guthrie, John T. (1980). The 1970's comprehension research. The Reading Teacher, 33, 880-882.

Hoffman, James V., & Baker, Christopher. (1981). Characterizing teacher feedback to student miscues during oral reading instruction. The Reading Teacher, 34, 907-912.

Kendall, Janet R., Mason, Jana M., & Hunter, William. (1980). Which comprehension? Artifacts in the measurement of reading comprehension. Journal of Educational Research, 73, 233-235.

Kerman, Sam. (1979). Teacher expectations and student achievement. Phi Delta Kappan, 60, 716-718.

Koenke, Karl. (1980). Recent developments in testing reading comprehension. The Reading Teacher, 33, 506-509.

Lass, Bonnie. (1984). Do teachers individualize their responses to reading miscues? A study of feedback during oral reading. Reading World, 23, 242-254.

Pikulski, John. (1974). A critical review: Informal reading inventories. The Reading Teacher, 28, 141-151.

Proctor, B. Patrick. (1984). Teacher expectations: A model for school improvement. The Elementary School Journal, 84, 469-481.

Rowe, M. (1974). Wait time and rewards as instructional variables, their influence on language, logic, and fate control. Journal of Research in Science, 11, 291-308.

Sipay, Edward R. (1964). A comparison of standardized reading scores and functional reading levels. The Reading Teacher, 17, 265-268.

Spiegel, Dixie Lee, & Rogers, Carol. (1980). Teacher responses to miscues during oral reading by second-grade students. Journal of Educational Research, 74, 8-12.

Stetson, Elton G. (1982). Reading tests don't cheat, do they? Journal of Reading, 25, 634-639.