Using 660 elementary school students from six school districts as subjects, a study examined student performance in reading aloud from four different reading curricula (Houghton Mifflin; Ginn 720; Holt, Rinehart & Winston; and Scott, Foresman), and the sensitivity of scores within each curriculum to growth across grade levels. Three different reading measurement tasks were developed. Two of the tasks were curriculum based; the other, which served as a common measure, was not curriculum bound. The two basal reading measurement tasks consisted of a reading passage and a vocabulary list; the common measure was a word list. The testing of the students was conducted within the first month of school. Results indicated that in all grades there appeared to be differences among the reading series in at least 3/6 of the measurement tasks—reading word lists or reading passages. The most consistent finding was that, when differences existed, Scott, Foresman and Houghton Mifflin appeared to be in the low student performance set, while Ginn 720 and Holt, Rinehart & Winston appeared in the high student performance set. This was especially true when the task utilized a word list rather than a reading passage. Results further indicated that the tasks for all four curricula adequately reflected student growth across grade levels. (HOD)
Research Report No. 93

CURRICULUM DIFFERENCES IN DIRECT REPEATED MEASURES OF READING

Gerald Tindal, Doug Marston, and Stanley L. Deno
Institute for Research on Learning Disabilities
University of Minnesota
Gary Germann
Pine County Special Education Cooperative

October, 1982

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
J. Ysseldyke

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."
Abstract

The performances of 660 elementary students in six school districts on two curriculum-based reading aloud tasks and one non-curriculum-based (common) measure were examined. Four different reading curricula were compared. In addition, the sensitivity of the reading measures to growth across grade levels was examined. Results indicated that differences existed in students' performance on the curriculum-based reading tasks at all grades. These tasks for all four curricula adequately reflected student growth across grades. The need for a common, non-curriculum-based, measure for comparisons among students using different reading curricula is emphasized.
The Institute for Research on Learning Disabilities is supported by a contract (300-80-0622) with the Office of Special Education, Department of Education, through Title VI-G of Public Law 91-230. Institute investigators are conducting research on the assessment/decision-making/intervention process as it relates to learning disabled students.

During 1980-1983, Institute research focuses on four major areas:

- Referral
- Identification/Classification
- Intervention Planning and Progress Evaluation
- Outcome Evaluation

Additional information on the Institute's research objectives and activities may be obtained by writing to the Editor at the Institute (see Publications list for address).

The research reported herein was conducted under government sponsorship. Contractors are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent the official position of the Office of Special Education.
Curriculum Differences in Direct Repeated Measures of Reading

Concern about the usefulness of standardized achievement tests frequently is expressed by educators. Specifically, the notion that standardized achievement tests aid school personnel in making accurate placement, instructional planning, and progress evaluation decisions has been challenged (Salvia & Ysseldyke, 1981). In the area of reading, Chall (1970) noted that many standardized tests distort our view of the student's real level of reading. Too much emphasis is placed upon the norm-referenced scores derived from such tests, scores that provide little or no information about actual reading skills. Similarly, Farr (1969) indicated that the most serious deficiency of standardized reading tests is their lack of diagnostic validity.

These warnings have been substantiated by research findings. Eaton and Lovitt (1972) found that student achievement in reading is frequently a function of the test used for assessment. Tallmadge (1977) demonstrated that low-achieving students, although appearing to make adequate progress on standardized reading tests, actually fall farther behind their peers. Hively and Reynolds (1975) stated that tests are not adequately designed to measure growth. A review of factor analytic studies focusing on standardized group reading tests convinced Spache (1976) that the tests measured few specific reading skills. He stated that "diagnostic profiles based on comprehension subskill performance often are relatively meaningless" (p. 252).

Finally, Jenkins and Pany (1978) examined the overlap between the content of several standardized achievement tests of reading and commercial reading series. Upon finding extreme variability in the content of the tests, the authors concluded that because of the
presence of curriculum bias, reading achievement tests yield unsatisfactory assessment information:

What educators need is an instrument to measure learning that is sensitive to curricular differences. Some form of criterion referenced or curriculum based assessment may provide the solution. Frequent and direct measures of a child's performance in a specific curriculum should reveal what skills within the curriculum have or have not been mastered, as well as provide some index of progress that would be sensitive to what was being taught. (p. 453)

The problematic nature of standardized achievement tests has indeed prompted educators to search for assessment procedures sensitive to both curriculum differences and student improvement. Frequent and direct measurement in a child's curriculum possesses these desired characteristics (Becker & Englemann, 1976; Deno & Mirkin, 1977; White & Haring, 1981).

In the measurement of reading achievement, advocates of frequent and direct measurement suggest that the skill area assessed should be fluency in reading aloud from either word lists or passages (Jenkins, Deno, & Mirkin, 1979). There is substantial evidence at the elementary grade levels to support this recommendation. Guthrie and Tyler (1978) found a substantial correlation between reading rate and comprehension. Deno, Mirkin, and Chiang (1982) corroborated these results and, in addition, found performance in reading aloud from both text and word lists to be sensitive to student improvement across six elementary grade levels.

Despite these useful claims of technical adequacy for reading rate measures, there is a need for further research before such practices are implemented. An unanswered question is the issue of whether reading aloud scores vary with the curriculum as do
standardized achievement test scores, thus revealing important differences in the curriculum. The issue becomes significant in those situations where students performing in different curricula are compared to one another. If student performance varies as a function of the curriculum, what measures can be used for cross-curricula comparisons?

A second unresolved issue is sensitivity to growth. Deno, Mirkin, and Chiang (1982) demonstrated the discriminant validity of correct performance in reading aloud between elementary grade levels. The research, however, was conducted on only a few selected text passages. No effort was made to determine whether the same results could be obtained consistently with a variety of commonly used curricula.

The purpose of this research was to undertake an analysis of both issues. In this investigation we examined the comparability of student performance in reading aloud from four different reading curricula and the sensitivity of scores within each curriculum to growth across grade levels.

**Method**

**Subjects**

A random sample of 660 elementary students attending six different school districts in a rural midwestern education cooperative served as subjects for this study. Class rosters were obtained for each grade (1-6) in the six districts and students' names numbered. Using a table of random numbers, equal numbers of students were selected from all classes for each grade. In four districts, 20
students from each grade were selected; 15 students were selected from each grade in the remaining two districts. No attempt was made to obtain equal representation of males and females.

Materials

Three different reading measurement tasks were developed. Two of the tasks were curriculum-based; the other, which served as a common measure, was not curriculum bound. The two basal reading measurement tasks consisted of a reading passage and a vocabulary word list; the common measure was a word list.

Four different basal reading series were in use in the six school districts, with two districts using the same reading series as two other districts. The four different series included: Houghton-Mifflin (1971 edition; Durr, Lapere, & Brown, 1971), Ginn 720 (1979 edition, Clymer & Fenn, 1979), Holt-Rinehart-Winston (1980 edition; Weiss, Everetts, Cruickshank, & Steuer, 1980), and Scott-Foresman (1981 edition; Aaron, Jackson, Riggs, Smith, & Tierney, 1981). Using the publishers' criteria for dividing the various books and magazines into appropriate grade levels, material was selected that corresponded to the beginning of each grade level.

Reading passages. The reading passage measure was developed by randomly selecting enough pages from this material (magazine, book, level) to present a minimum of 50 words in grade 1, 100 words in grade 2, and 200 words in grades 3-6. The only criterion for screening the pages selected was that no special dialogue or poetry be present in the passage. Two copies of each passage were made, one from which the student read and one on which the tester followed along. The tester's
copy included a cumulative count of the number of words in each successive line.

Word lists. Two types of word lists were developed. The basal word list measure consisted of a random selection of 150 words from the pool of all vocabulary words appearing within each of the grade levels from which text passages were drawn. The lists were cumulative and stratified, with 60% of the words sampled from grade level and 40% consisting of words sampled equally from previous grades. For example, a fifth grade list of 150 words would include 90 words from the fifth grade level and 15 words from each of the first through fourth grade levels. Again, two copies of each word list were made, one for the student to read, and a follow-along list for the tester.

The common word list consisted of a random selection of 150 words appearing in the core list of words in the Basic Reading Vocabulary (Harris & Jacobson, 1972), a computer based compilation of over 5000 words appearing in commonly used basal readers. The domain of words from which the list was constructed was limited to words from preprimer through grade four. Two copies of each list were made, one for the student to read, and a follow-along list for the tester.

Procedures
The testing of the 660 students was conducted within the first month of school. Ten educational aides were trained in the measurement procedures during a two-hour session and given all the necessary materials, including word lists, reading passages, recording forms, and stop watches. Each aide was assigned to a school, given a list of students, and a time schedule. All testing was done on an
individual basis and was completed within one week. Each session involved the administration of two reading passages, one basal word list, and one common measure. Within each testing session, the order of administration of the three measures was counterbalanced. Each of the reading measures was a one-minute sample. Directions given to the student were as follows:

- When I say start, begin reading at the top of the page. If you wait on a word for too long, I'll tell you the word. If you come to a word that you cannot read, just say pass and go on to the next word. Do not attempt to read as fast as you can. This is not a "speed reading" test. Read at a comfortable rate, and at the end of one minute I'll say "stop."

The testers followed along on their copy and marked errors of omission, substitution, and mispronunciation. If the student failed to continue within approximately 5 seconds, the tester prompted the student to skip that word and continue. Immediately following each one-minute measure, the student's performance was scored and recorded, with no feedback given to the student.

Results

Comparison of Districts Using the Same Curriculum

Prior to comparing the basal reading series with one another, the two pairs of districts using the same series were compared to determine whether any population differences existed. The two school districts using Ginn 720 and the two districts using Houghton-Mifflin were compared at each grade level on the two basal reading measures—word lists and passages. Findings from these analyses indicated that in three grades there was a significant difference in the performance of students in the two districts using the Houghton-Mifflin reading
series. In grade 2, differences between districts appeared on both passage and basal word list tasks in the Houghton-Mifflin reading series and on the Ginn 720 basal word list task. In grade 4, differences between districts appeared on Ginn 720 basal passage and word list and on the Houghton-Mifflin basal passage. Finally, in grade 5, there was a difference between the two districts on the Ginn 720 basal passage. When the two districts using Ginn 720 were compared on the common word list measurement task, significant differences appeared for grades 1 and 2.

Comparisons Among Curricula

Means and standard deviations for students on the three reading measurement tasks for each basal reading series appear in Table 1. One-way analyses of variance were conducted on these data to ascertain the extent to which any differences existed between the reading series. Significant results were tested using the Student-Newman-Keuls a posteriori test.

For the reading passage data, there was a significant difference between the various readers in four of seven grade level comparisons (see Table 2). Only in grades 2 and 3 and averaging across all six grades was there no significant difference. The a posteriori tests revealed that in the four grades showing a difference, Scott-Foresman consistently appeared in the low set (characterized by lower student performance) while Ginn 720 and Holt-Rinehart appeared primarily in
the high set (characterized by higher student performance). Analysis of performance, on the basal vocabulary word list produced many of the same findings (see Table 3), with a total of three grades and the average across grades showing a significant different among the reading series. In grades 1-3, no significant difference among the four series was indicated. The \textit{a posteriori} tests again showed Scott-Foresman and Houghton-Mifflin in the low set, while Ginn 720 and Holt-Rinehart quite consistently appeared in the high set. Student performance on the common word list provides a direct measure of comparability of students' performance unconfounded by the effect of the basal reader. No significant differences occurred in any of the grade levels (see Table 4).

\begin{center}
Insert Tables 2-4 about here
\end{center}

An analysis of covariance was conducted to adjust for any differences between populations and thus provide a more valid comparison of the different reading series. The results of this analysis appear in Table 5, along with the rank order of each reading series. The rank order was obtained following adjustment for performance on the common measure. As can be seen from the F-ratios and F-probabilities in Table 5, there were significant differences among the four basal readers, both when the dependent measure was reading from a passage and when it was reading a vocabulary word list. In all six grade levels and in the average across grades, there was a significant difference among the reading series for at least one of
the measures. In grades 2 and 5, using the basal reading passage, and in grade 1, on the basal word list, no significant differences appeared.

Insert Table 5 about here

Table 6 is a summary of the rank order data from Table 5. As is evident in Table 6, the four reading series tended to cluster quite consistently in a dichotomous fashion. For the basal reading passage, Scott-Foresman received mostly ranks of 4, while Ginn 720 and Holt-Rinehart primarily received ranks of 1 and 2. Houghton-Mifflin fell in the middle. On the basal word list, Scott-Foresman and Houghton-Mifflin consistently received lower ranks (3 and 4) while Ginn 720 and Holt-Rinehart received ranks of 1 and 2.

Insert Table 6 about here

The data presented in Tables 7 and 8 help clarify whether reading rate measures are sensitive to growth within all reading curricula. Our major premise was that reading rate should increase with grade level if the measure is sensitive to a child's progress. The grade level means for the four reading curricula appeared to increase with grade. To test this hypothesis, the grade level means for each measure within the curriculum were first subjected to an analysis of variance. All F-values reported in Table 7 were significant. A more crucial measure of sensitivity, however, is a test of linearity. If
reading rate is sensitive to progress, it should increase with grade. Reading rates on basal passages and basal word lists from Ginn 720, Scott-Foresman, Houghton-Mifflin, and Holt-Rinehart all displayed linear trends, significant at the .05 level (see Table 8). Reading rate within any of the four curricula appears to be sensitive to educational progress.

-----------------------------
Insert Tables 7 and 8 about here

-----------------------------

Discussion

The data from this study consistently indicated that there are differences in the four basal reading series when using reading rate as the dependent measure. In each of the various analyses, the results were similar. Using either a one-way analysis of variance or an analysis of covariance, differences were apparent among the reading series on both the reading passage and basal word list measurement tasks. For most grades, this difference was found on both basal reading measures.

In a few grades, a difference was found in one measure but not the other. There is no obvious reason why this incongruity appears. One possibility is that for those grades, the two basal reading tasks were differentially difficult as a result of the procedures used in generating the measurement materials. That is, by randomly selecting words and passages, provision is made for stimulus equivalence only over the long run. Any particular selection may be more or less difficult. Given the differences within any reading series in the
type of vocabulary (Carnine & Silbert, 1979) and the readability of successive passages (Fuchs, 1981), it is possible that measurement stimuli were generated that were not representative of the material in general. The result was that the passage selected was unusually easy or difficult and failed to differentiate students or corroborate previously found differences between the various series. For the vocabulary list, the frequency of regular or irregular words may have resulted in a task atypical of that series and level in general. Again, with the task changed in difficulty, there is less likelihood of differentiating between students and basal series in a manner similar to the general pattern of findings. For the reading passages, the great differences in the readability found by Fuchs (1981) may have been operating in this study also.

In general, this research revealed real differences in the reading series with no strong population sampling effect indicated. There did not appear to be any differences between students of various districts, using both the common measure and both basal reading measures. Because two of the districts in this study were using the same reading series as two other districts, it was possible to compare directly these districts on all three measures. If differences appeared on the basal reading measures, the evidence supporting any differences in reading series would be seriously weakened.

In two districts, all elementary grades were using Ginn 720, while two other districts were using Houghton-Mifflin. A comparison of performance in these four districts revealed very few differences. In some grades (4 and 5), a difference appeared in one of the measures
but not the others. In the use of any random selection procedures, equivalence is obtained only over repeated samplings and any given sample may be biased in one direction or the other. It is entirely possible that the two groups of students from these districts were not representative of each of their general populations or that the measurement stimuli generated were not equivalent in difficulty. As a result, more (or less) proficient students were selected and compared with each other on more or less comparable tasks. It is interesting to note that such an occurrence is actually more than likely given the number of districts and grades sampled. In all, 36 different populations (of students and grade level materials) were sampled. The likelihood of obtaining no population differences in all of these is low. However, the word lists and passages in these districts were not exactly the same, but rather sampled from within the same grade level book. Therefore, population differences were confounded with curriculum differences, which may explain the population differences that were found.

The findings obtained on the one-way analyses of variance were essentially corroborated when analyses of covariance were used. Using student performance on the common measure as a covariate, a more exact comparison of the reading series could be made without any influence of differential entry skills. In all grades, there appeared to be differences among the reading series, in at least one of the measurement tasks—reading word lists or reading passages.

This research attempted not only to determine whether differences existed between the reading series, but the ways in which the
differences were exhibited. The two analyses which examined this aspect included the Student-Newman-Keuls test and the rank ordering of basal series on the reading rates obtained. In the former (Tables 2-4), reading series are put into homogeneous subsets when there are no significant differences in the mean performance. With four basal readers being compared, the total number of possible subsets is four. However, this analysis revealed that when differences appeared among the series, the resulting grouping into sets was dichotomous, with one set characterized by lower student performance and the other characterized by higher student performance. The most consistent finding was that, when differences existed, Scott-Foresman and Houghton-Mifflin appeared in the low set, while Ginn 720 and Holt-Rinehart appeared in the high set. This was especially true when the task utilized a word list rather than a reading passage. Ginn 720 clearly fell within the high set on the word list task.

A similar grouping of series was done by rank ordering them in difficulty using the deviation of performance from the grade mean, after adjusting for differences on the covariate (Table 6). Again, the data showed a very consistent dichotomy. The four reading series fell into two groups—one group in which ranks of 1 and 2 consistently were obtained and one in which ranks of 3 and 4 consistently were obtained. Again, Scott-Foresman and Houghton-Mifflin showed the lowest performance and corresponding ranks, while Ginn 720 and Holt-Rinehart revealed higher performance and ranks of 1 and 2. On the basal passage, the majority of ranks were consistent with this dichotomy, while on the basal word list, all ranks dichotomize in like
manner. This finding coincides very well with the analysis by Carnine and Silbert (1979). They found that the number of irregular words introduced in the first 50 lessons of Ginn 720, Houghton-Mifflin (1974), and Scott-Foresman (1981) was 29, 38, and 125, respectively. It is possible that reading rates are highly related to the frequency of irregular words.

In conclusion, it appears that differences do exist between the four basal reading series; these differences indicate that Scott-Foresman (1981 edition) and Houghton-Mifflin (1974 edition) generate lower reading rates than those obtained from the use of Holt-Rinehart (1981 edition) or Ginn 720 (1979 edition). The implications of this finding are particularly noteworthy.

According to Jenkins and Pany (1978), there is differential overlap between various achievement tests and basal reading series. This finding, however, was a result of a content analysis of the tests and reading series and was predicated on certain assumptions concerning student learning and performance. To corroborate this analysis, empirical data on actual student performance were needed. The findings from the present research partially provide such evidence. Given that students perform differentially according to the reading series used, and that achievement tests sample differentially from these series, it is likely that such biases will be reflected in students' performance on the achievement measures. This is particularly true for tests emphasizing decoding and vocabulary words. In essence, achievement tests no longer can be considered as impartial, common measures on which comparisons can be made among
students working in different readers.

An alternative to this type of measurement system would be the use of curriculum-based assessment as advocated by Jenkins, Deno, and Mirkin (1979). As long as measurement and analysis of student performance remains within a reading series, there is no bias in comparisons among students. However, with the use of many different basal readers, some common measure is needed that does not differentially sample from any one basal reading series. In many educational cooperatives, in which the delivery of special education is in concert with several different districts, the use of this common measure is imperative for valid between-district comparisons. Indeed, one of the basic tenets of curriculum-based assessment is the use of local norms. The use of such a system will result in generation of valid data having direct instructional relevance. Not only should the norms be based on a comparable population, but also utilize a comparable set of stimulus materials for conducting assessment.
References


## Table 1
Students' Performance on Various Reading Measurement Tasks Across Four Different Basal Reading Series

<table>
<thead>
<tr>
<th>Measurement Task</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reading Passage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Houghton-Mifflin</td>
<td>13.9</td>
<td>14.8</td>
<td>49.9</td>
<td>37.7</td>
<td>72.4</td>
<td>43.1</td>
<td>99.1</td>
</tr>
<tr>
<td>Ginn 720</td>
<td>8.3</td>
<td>9.4</td>
<td>39.9</td>
<td>29.6</td>
<td>69.9</td>
<td>41.2</td>
<td>111.8</td>
</tr>
<tr>
<td>Holt-Rinehart</td>
<td>2.8</td>
<td>8.6</td>
<td>45.2</td>
<td>22.1</td>
<td>92.0</td>
<td>43.2</td>
<td>105.4</td>
</tr>
<tr>
<td>Scott-Foresman</td>
<td>5.7</td>
<td>5.3</td>
<td>49.4</td>
<td>38.2</td>
<td>81.6</td>
<td>41.3</td>
<td>83.1</td>
</tr>
<tr>
<td><strong>Basal Word List</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Houghton-Mifflin</td>
<td>1.1</td>
<td>2.5</td>
<td>18.5</td>
<td>30.0</td>
<td>26.8</td>
<td>19.3</td>
<td>39.4</td>
</tr>
<tr>
<td>Ginn 720</td>
<td>1.4</td>
<td>2.4</td>
<td>19.3</td>
<td>19.6</td>
<td>32.0</td>
<td>20.7</td>
<td>44.8</td>
</tr>
<tr>
<td>Holt-Rinehart</td>
<td>1.2</td>
<td>2.5</td>
<td>21.3</td>
<td>18.8</td>
<td>35.4</td>
<td>16.5</td>
<td>45.9</td>
</tr>
<tr>
<td>Scott-Foresman</td>
<td>1.6</td>
<td>1.6</td>
<td>21.9</td>
<td>20.5</td>
<td>24.1</td>
<td>19.8</td>
<td>25.3</td>
</tr>
<tr>
<td><strong>Common Word List</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Houghton-Mifflin</td>
<td>.6</td>
<td>3.0</td>
<td>13.1</td>
<td>18.0</td>
<td>31.4</td>
<td>21.9</td>
<td>53.2</td>
</tr>
<tr>
<td>Ginn 720</td>
<td>.6</td>
<td>1.1</td>
<td>10.7</td>
<td>14.3</td>
<td>29.0</td>
<td>21.6</td>
<td>45.7</td>
</tr>
<tr>
<td>Holt-Rinehart</td>
<td>.3</td>
<td>1.0</td>
<td>10.3</td>
<td>.86</td>
<td>35.1</td>
<td>19.5</td>
<td>50.0</td>
</tr>
<tr>
<td>Scott-Foresman</td>
<td>.1</td>
<td>.3</td>
<td>17.8</td>
<td>18.2</td>
<td>37.7</td>
<td>21.9</td>
<td>51.8</td>
</tr>
</tbody>
</table>
Table 2

One Way Analysis of Variance Comparing Basal Readers Using Performance on Reading Passage Measurement Task: F-Ratio, F-Probability, and Grouping into Homogeneous Subsets Using Student-Newman-Keuls

<table>
<thead>
<tr>
<th>Grade Level</th>
<th>F-Ratio</th>
<th>F-Probability</th>
<th>Number of different sets</th>
<th>Name of Reading Series in Low Set</th>
<th>High Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.71</td>
<td>.00</td>
<td>2</td>
<td>SF, HR</td>
<td>HM</td>
</tr>
<tr>
<td>2</td>
<td>.68</td>
<td>.57</td>
<td>1</td>
<td>Ginn</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.21</td>
<td>.31</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3.02</td>
<td>.03</td>
<td>2</td>
<td>SF</td>
<td>Ginn</td>
</tr>
<tr>
<td>5</td>
<td>2.57</td>
<td>.05</td>
<td>2</td>
<td>SF</td>
<td>HM, Ginn</td>
</tr>
<tr>
<td>6</td>
<td>2.58</td>
<td>.05</td>
<td>2</td>
<td>SF, HM</td>
<td>Ginn, HR</td>
</tr>
<tr>
<td>Average</td>
<td>1.36</td>
<td>.245</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3


<table>
<thead>
<tr>
<th>Grade Level</th>
<th>F-Ratio</th>
<th>F-Probability</th>
<th>Number of different sets</th>
<th>Name of Reading Series in Low Set</th>
<th>High Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.29</td>
<td>.84</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.16</td>
<td>.93</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.40</td>
<td>.25</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>6.14</td>
<td>.00</td>
<td>2</td>
<td>SF HR Ginn HR</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6.65</td>
<td>.00</td>
<td>2</td>
<td>HM SF Ginn HR</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3.54</td>
<td>.02</td>
<td>1</td>
<td>HM SF Ginn HR</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>6.50</td>
<td>.00</td>
<td>2</td>
<td>HM SF Ginn HR</td>
<td></td>
</tr>
</tbody>
</table>
Table 4


<table>
<thead>
<tr>
<th>Grade Level</th>
<th>F-Ratio</th>
<th>F-Probability</th>
<th>Number of different sets</th>
<th>Name of Reading Series in Low Set</th>
<th>High Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.39</td>
<td>.76</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1.01</td>
<td>.39</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.84</td>
<td>.47</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.97</td>
<td>.41</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.62</td>
<td>.19</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1.38</td>
<td>.25</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>.16</td>
<td>.93</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5

Analysis of Covariance Comparing Basal Readers on Two Basal Reading Measurement Tasks Using Performance on a Common Reading Task as the Covariate

<table>
<thead>
<tr>
<th>Grade Level</th>
<th>Basal Reading Passage - Ranking of Series*</th>
<th>Basal Word List - Ranking of Series*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F-Ratio</td>
<td>F-Probability</td>
</tr>
<tr>
<td>1</td>
<td>4.58</td>
<td>.00</td>
</tr>
<tr>
<td>2</td>
<td>1.02</td>
<td>.39</td>
</tr>
<tr>
<td>3</td>
<td>2.88</td>
<td>.04</td>
</tr>
<tr>
<td>4</td>
<td>12.59</td>
<td>.00</td>
</tr>
<tr>
<td>5</td>
<td>1.07</td>
<td>.37</td>
</tr>
<tr>
<td>6</td>
<td>5.64</td>
<td>.00</td>
</tr>
<tr>
<td>Average</td>
<td>9.75</td>
<td>.00</td>
</tr>
</tbody>
</table>

*Ranks obtained following adjustment for performance on measure used as covariate.
Table 6
Comparison of Four Basal Reading Series on Number of Times Represented by each Possible Ranking on Two Measurement Tasks

<table>
<thead>
<tr>
<th>Reading Series</th>
<th>Basal Reading Passage Number of Times with Rank of*</th>
<th>Basal Word List Number of Times with Rank of*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>Houghton-Mifflin</td>
<td>3 1 3 0</td>
<td>0 0 5 2</td>
</tr>
<tr>
<td>Ginn</td>
<td>1 4 2 0</td>
<td>3 4 0 0</td>
</tr>
<tr>
<td>Holt-Rinehart</td>
<td>3 2 1 1</td>
<td>3 3 1 0</td>
</tr>
<tr>
<td>Scott-Foresman</td>
<td>0 0 1 6</td>
<td>1 0 1 5</td>
</tr>
</tbody>
</table>

*Ranks obtained following adjustment for performance on measure used as covariate.*
Table 7
Analysis of Variance Results for Grade Level Means Within Each Reading Series for Curriculum-Based Measures

<table>
<thead>
<tr>
<th>Reading Series</th>
<th>Basal Passage F-value</th>
<th>Basal Passage Significance</th>
<th>Basal Word List F-value</th>
<th>Basal Word List Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Houghton-Mifflin</td>
<td>56.47</td>
<td>.001</td>
<td>29.17</td>
<td>.001</td>
</tr>
<tr>
<td>Ginn 720</td>
<td>52.41</td>
<td>.001</td>
<td>29.42</td>
<td>.001</td>
</tr>
<tr>
<td>Holt-Rinehart</td>
<td>42.63</td>
<td>.001</td>
<td>27.46</td>
<td>.001</td>
</tr>
<tr>
<td>Scott-Foresman</td>
<td>24.76</td>
<td>.001</td>
<td>11.37</td>
<td>.001</td>
</tr>
</tbody>
</table>
Table 8
Tests of Linearity for Grade Level Means Within Each Reading Series for Curriculum-Based Measures

<table>
<thead>
<tr>
<th>Reading Series</th>
<th>Basal Passage</th>
<th>Basal Word List</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F-value</td>
<td>Significance</td>
</tr>
<tr>
<td>Houghton-Mifflin</td>
<td>3.68</td>
<td>.007</td>
</tr>
<tr>
<td>Ginn 720</td>
<td>2.78</td>
<td>.029</td>
</tr>
<tr>
<td>Holt-Rinehart</td>
<td>3.05</td>
<td>.021</td>
</tr>
<tr>
<td>Scott-Foresman</td>
<td>3.57</td>
<td>.008</td>
</tr>
</tbody>
</table>
The Institute is not funded for the distribution of its publications. Publications may be obtained for $3.00 per document, a fee designed to cover printing and postage costs. Only checks and money orders payable to the University of Minnesota can be accepted. All orders must be prepaid.

Requests should be directed to: Editor, IRLD, 350 Elliott Hall; 75 East River Road, University of Minnesota, Minneapolis, MN 55455.


Note: Monographs No. 1 - 6 and Research Report No. 2 are not available for distribution. These documents were part of the Institute's 1979-1980 continuation proposal, and/or are out of print.


Thurlow, M. L., & Greener, J. W. Preliminary evidence on information considered useful in instructional planning (Research Report No. 27). March, 1980.


Epps, S., McGue, M., & Ysseldyke, J. E. Inter-judge agreement in classifying students as learning disabled (Research Report No. 51). February, 1981.

Epps, S., Ysseldyke, J. E., & McGue, M. Differentiating LD and non-LD students: "I know one when I see one" (Research Report No. 52). March, 1981.


Graden, J., Thurlow, M., & Ysseldyke, J. Instructional ecology and academic responding time for students at three levels of teacher-perceived behavioral competence (Research Report No. 73). April, 1982.


Thurlow, M. L., Ysseldyke, J. E., Graden, J., Greener, J. W., & Mecklenberg, C. Academic responding time for LD students receiving different levels of special education services (Research Report No. 78). June, 1982.


