

Resource Consultant Training Program
Research Report No. 4

RCTP

Combining Informal Teacher
Judgment and Objective Test
Scores to Make Cross-Classroom
Reading Group Placements

Richard Parker
Jan Hasbrouck
Gerald Tindal

Published by
Resource Consultant Training Program
Division of Teacher Education
College of Education
University of Oregon

Copyright © 1989 University of Oregon. All rights reserved. This publication, or parts thereof, must not be used or reproduced in any manner without written permission. For information address Resource Consultant Training Program, 275 Education, University of Oregon, Eugene, OR 97403-1215.

Parker, Richard
Hasbrouck, Jan
Tindal, Gerald
Combining Informal Teacher Judgment and Objective Test Scores to Make Cross-Classroom Reading Group Placements
Research Report No. 4

Staff

Gerald Tindal, Program Director
Katie Essick, Editor
Clarice Skeen
Denise Styer
Mark Baldwin

Acknowledgements

Preparation of this document was supported in part by the U.S. Department of Education, grant number G008715106. Opinions expressed herein do not necessarily reflect the position or policy of the U.S. Department of Education, and no official endorsement by the Department should be inferred.

Cover design: George Beltran

COMBINING INFORMAL TEACHER JUDGMENT AND OBJECTIVE TEST SCORES TO MAKE CROSS-CLASSROOM READING GROUP PLACEMENTS

Richard Parker
Jan Hasbrouck
Gerald Tindal

Abstract

The two purposes of this study are to (a) demonstrate a method of summarizing and displaying test data to support teachers' decisions on group placement, and (b) apply this method to cross-classroom reading group placements, using data from three reading assessment procedures (Analytical Reading Inventory, Maze, and Oral Reading Fluency). The study emphasizes the presentation of objective test data in a manner congruent with typical elementary classroom decision-making. The three reading tests were administered to Grade 4-6 students who had been placed in existing reading groups by teacher consensus on the basis of classroom observations and a variety of informal criteria. Discriminant function analysis (DFA) was used to estimate the classification accuracy of each of the three measures. The DFA procedure also yielded stacked area graphs, which allow a teacher to judge the relative certainty with which a given student can be assigned to each of the five reading groups. The teacher also can judge when to rely on alternate, informal criteria.

Recently, researchers have begun to promote an active decision-making role by teachers (Borko & Shavelson, 1982; Orlich, et al., 1985; Peterson, 1988) and consideration of the classroom context or ecology (Doyle, 1986; Gump, 1987; Leinhardt, Weidman, & Hammond, 1987) in the context of effective instruction. Their views offer a new legitimacy to informal teacher decisions, largely in recognition that such decisions are made from a wealth of relevant, immediate cues. Researchers themselves are being challenged to develop conceptual and observational tools to measure the decision-making process and relevant contextual cues (Wicker, 1987).

One type of teacher decision with long-term effects on what and how elementary school stu-

dents learn is reading group placement. In a survey by Salmon-Cox (1981), elementary teachers ($N = 68$) from urban and suburban schools said they distrusted standardized test results for making grouping decisions. Most (66%) said they relied instead on informal observation over the first few weeks or months of school. Burry, Catterall, Choppin, and Dorr-Bremme (1982) conducted on-site interviews with 44 elementary teachers from a broader national survey ($N=256$). For placing students in instructional groups and/or changing group placements, "...nearly every respondent reported that 'my own observations and students' classwork' was a crucial or important source of information" (p. 24). Their finding of disregard for test results in favor of direct classroom observations had been

noted earlier by Resnick (1981), Sproull and Zubrow (1981), Stetz and Beck (1979), and Goslin (1967). In forming instructional groups within their classrooms, teachers are able to observe students over a wide range of reading-related activities throughout the day, and form an impression of functional skills, habits, and tolerances over several days or weeks. Group placement decisions therefore tend to be subjective judgments, based upon information gathered from multiple sources.

Strike (1983) found that teachers' placement decisions were based largely on the likelihood that a student would profit from the instructional strategy used in a particular group. Teachers considered work habits, classroom behavior, and personal/social abilities, as well as measurable reading skills, in estimating the fit between child and group. Within a single classroom, daily informal observations of trial placements often are used to gauge the fit between a child and his/her reading group; alternative group placements can be made with little disruption.

Haller and Waterman (1985) collected interview data on 60 intermediate-grade teachers to identify criteria for placing students ($N = 1,031$) in reading groups the following year. Four prime placement criteria were identified: estimated reading ability (45% of cases), general academic competence (24%), work habits (16%), and behavior/personality (14%). In no case, however, were the prime criteria the *only* criteria considered. Teachers were likely to consider other, non-skill factors, particularly in placing students whose reading abilities fell close to the margin between two groups. Those with marginal reading abilities—who could be placed reasonably in two groups—are a special concern of this study. Reading assessment data that do not indicate placements with reasonable certainty have low predictive power or predictive error. These marginal score ranges, uncertain placements, and tests with low predictive power indicate the necessity for other, non-test criteria, including informal teacher judgment.

Salmon-Cox (1981) confirmed that test results are of secondary importance to most teachers in making reading group placements, and are considered relevant only when they are "precise, closely matched to curricula and instruction, and timely" (p. 634). Burry et al. (1982) found that the tests teachers most often used were developed by the teachers themselves or were curriculum-related skill inventories. Teachers' most common criticism of standardized reading tests is the discrepancy between test scores and both informal measures and teacher judgment (Coleman & Harmer, 1982).

When test results are used to help make group placements, Informal Reading Inventories (IRIs), whether teacher-constructed, basal series-based, or commercial, are the most popular assessment tools (Harris & Lalik, 1987). IRIs are flexible, allow timely feedback, and may be directly related to the curriculum.

The popularity of IRIs might have another basis: Their format and content are consistent with the type of teacher judgments and decision-making that currently are being documented in elementary classrooms (Burry et al., 1982). IRIs are unique among reading tests in that they (a) encourage subjective judgment in administration, scoring, and interpretation of results; (b) are multi-faceted, often including scores for oral reading fluency and accuracy, types of miscues, and reading comprehension; (c) encourage holistic judgments of a student's concentration, effort, and frustration; and (d) involve passage-reading performance that is relatively "natural" to a reading group. These IRI attributes are congruent with common group placement preferences and practices documented by Salmon-Cox (1981), Haller and Waterman (1985), Strike (1983), and Burry et al. (1982).

Although IRIs are popular and appear consistent with prevalent decision-making practices, they have been criticized in the professional literature. For example, little agreement exists among practitioners on the construction, administration, and scoring of IRIs (Powell, 1968; Ekwall, 1974; Page, 1976; Pflaum, 1980; Brecht, 1977; Kender & Rubenstein, 1977; Schwartz, 1984), and on the designation of independent, instructional and frustration reading levels (Pikulski, 1974). Furthermore, IRIs have demonstrated low accuracy in estimating instructional levels (Pikulski & Shanahan, 1982; Bristow, Pikulski, & Pelosi, 1983).

Additional criticism has been levied against commercial IRIs for lacking the content validity of inventories derived from classroom materials (Harris & Niles, 1982). And some commercial IRIs have demonstrated the same problem with passage-independent questions commonly found in standardized tests (Marr & Lyon, 1980). Basal series-integrated IRIs offer greater content validity, but they seldom are validated against external criteria, and often have unknown reliability. Finally, there are no standard methods used across basal programs for constructing these IRIs to ensure representativeness of text material.

Other informal classroom materials-based methods for placing students in reading materials or instructional groups include the cloze procedure, its maze variation (Guthrie, Seifert,

Burnham, & Caplan, 1974), and oral reading fluency. Like the studies on IRIs, research on the reliability and validity of the Cloze and Maze generally is supportive but shows the measures to be most useful (a) in making gross judgments, (b) over a broad grade/skill range, and (c) with the support of other test data (Rankin & Culhane, 1969; Bormouth, 1968; Alexander, 1968; Bradley, Ackerson, & Ames, 1978; Bradley & Meredith, 1978; Pikulski & Tobin, 1982).

Oral Reading Fluency is a popular assessment tool within the curriculum-based measurement (CBM) movement. Studies on the reliability and validity of oral reading fluency generally have produced strong results (Deno, Marston, Mirkin, Lowry, Sindelar, & Jenkins, 1982; Deno, Mirkin, & Chiang, 1982; Marston & Magnusson, 1985; Deno, Marston, Shinn, & Tindal, 1983; Fuchs, Fuchs, & Maxwell, 1988). The usefulness of oral reading fluency, however, appears to be limited by its lack of face validity because teachers generally do not regard it as an acceptable measure of reading comprehension (Ysseldyke, 1979).

While they have their limitations, informal assessment measures such as IRIs, the maze procedure, and Oral Reading Fluency can improve decision-making for reading group placements. They are especially valuable when instructional groups are formed *across* classrooms within one grade level or across several grade levels. In these cases, a teacher's reliance upon informal observation and judgment for group placement poses a special problem for consensual group decision-making: Informal observations from all homeroom teachers probably are not comparable because they lack the common frame of reference that objective measures can offer.

A major challenge is to improve teachers' use of objective measures. Objective test data should be presented in such a way as to highlight the error or limitations of a measure and simultaneously to support subjective teacher judgments based on ongoing classroom observations and a variety of other non-test score criteria.

For the present study, the discriminative validity of three informal reading measures (IRI, Maze, Oral Reading Fluency) was investigated for placing Grade 4, 5, and 6 students in across-classroom reading instruction groups. The study was designed to be congruent with current teacher decision-making practices in five ways: (a) Informal, consensual judgments by teachers serve as the group placement criterion; (b) two of the reading measures are easily produced, administered, and

scored by teachers, and are drawn from local classroom materials; (c) one of the reading measures is an IRI, selected for its documented popularity among teachers; (d) the results are presented in terms of the relative certainty of each alternative placement; and (e) the results are presented to help illuminate the areas of greatest need for informal teacher judgments based on alternative, non-test score criteria.

Informal teacher judgment based upon multiple criteria is especially valuable when more objective measures show the least discriminative power (greatest error). This study uses data summary and display techniques that point out test score ranges where objective measures have the least power and informal teacher judgment is needed most.

METHOD

Subjects

Participating in this study were 120 students sampled randomly from class rosters of 14, Grade 4, 5, and 6 classrooms in a 280-pupil elementary school located in a middle SES, rural western community. Random selection was stratified within existing reading groups: Ten students were randomly selected from each of four reading groups at each grade, resulting in 40 subjects per grade level.

Reading groups were formed *across* classrooms (approximately four classrooms per grade level), but *within* grade levels. The groups had been established 6 months earlier (in September) by consensus among the four or five teachers at each grade level. Placement decisions were based upon a variety of information, including cumulative folder reports, recommendations of the previous year's teacher, informal classroom observations, and IRIs. School-wide standardized test results also were available but carried little weight in the group placement decisions. After initial group placements had been made, changes were allowed if warranted by informal teacher observations.

Ten students were selected for each of four reading groups: High, High-Medium, Low-Medium, and Low. The Low group was divided further into Low and SPED/Chapter I groups, with group sizes of approximately six and three, respectively. Students in the SPED/Chapter I group were pulled for intensive reading instruction within either the school's Special Education Resource Room program or its Chapter I remedial program. Exact numbers of students within each group at each grade level are contained in Table 1.

Instrumentation

All students in the study were tested with three instruments: a modified version of the Analytical Reading Inventory (ARI, Woods & Moe, 1985) and two measures derived from students' basal readers: Maze tests and Oral Reading Fluency procedures.

The ARI

A modification of the ARI was administered individually, yielding an instructional reading level (grade equivalency score) for each student. The ARI could not be administered exactly as prescribed in the manual because: "some of the decision-making is based upon subjective evaluations," (e.g., "A careful look at the severity of the miscues and the overt demeanor of the student would be necessary before a final judgment could be made") (Woods & Moe, 1985, p. 18). The emphasis on subjective judgments did not allow reasonable interrater reliability among the five research aides. Similar difficulties would be posed if teachers administered the ARI in separate classrooms and attempted to compare data for consensual group placements.

Reasonable interrater reliability ($r=.90$) for instructional reading levels across the data collectors was obtained in the training session with four modifications: (a) "Miscue analysis" and "story retell" were deleted, as neither was required to calculate reading levels; (b) objective decision rules were established for determining how many test passages to use and in what order; (c) objective procedures were established for computing instructional levels from reading rate and comprehension scores; and (d) the preliminary word list reading task was omitted as unnecessary; current group placement guided the ARI entry level instead.

Basal Passages

Passages for the two curriculum-related measures were selected from Grade 4, 5, and 6 reading texts, *Running Free*, *Three Cheers*, and *Distant Views* (Scott, Foresman Reading, 1987). Each student read three passages, with calculated readabilities at grade placement level, 1 year above grade placement, and 1 year below grade placement. Based on the recommendations of local reading specialists, readability was defined as the average of Spache, Fry, and RAYGOR formulas. Three passages used

Table 1: Performance of Language Arts Instructional Groups in Grades 4, 5, and 6 on Two, Basal Reader-Based Measures (Maze and Oral Reading Fluency) and the Analytical Reading Inventory

	Maze			Oral Reading Fluency			ARI		
	N	M	SD	N	M	SD	N	M	SD
Grade 4									
All groups	37	74.76	19.74	39	103.67	31.99	39	5.53	1.58
SPED/Chapt. I	3	25.67	4.04	2	39.5	16.26	3	2.83	1.04
Low	5	63.2	12.83	7	71.43	19.12	7	4.29	1.29
Low-Medium	9	70.22	12.16	10	102.	10.9	9	5.33	.9
High-Medium	10	85.4	8.26	10	109.3	26.56	10	5.55	.89
High	10	88.7	6.53	10	135.1	19.04	10	7.35	.58
Grade 5									
All groups	34	82.85	11.26	40	115.4	31.54	40	6.18	1.58
SPED/Chapt. I	2	59.	2.83	3	60.	21.93	3	3.83	1.04
Low	6	69.33	6.98	7	93.	15.68	7	4.36	1.18
Low-Medium	10	83.	4.78	10	112.6	23.76	10	6.25	1.06
High-Medium	9	88.11	5.82	10	125.7	26.29	10	6.45	.64
High	7	94.29	2.98	10	140.2	23.69	10	7.8	.79
Grade 6									
All groups	32	81.94	11.32	38	118.71	29.31	38	6.89	1.66
SPED/Chapt. I	1	53.	-	1	75.	-	1	6.	-
Low	6	74.67	11.02	9	91.11	22.87	8	4.81	1.44
Low-Medium	8	77.13	8.15	9	113.89	18.29	9	6.67	1.35
High-Medium	9	86.22	7.76	10	125.1	23.99	10	7.2	0.72
High	8	91.	6.07	9	148.89	16.45	10	8.55	0.64

for each grade level had the following readabilities: 3.5, 4.6, 5.6 (Grade 4); 4.6, 5.6, 6.3 (Grade 5); 5.6, 6.3, 7.5 (Grade 6). Because some passages were used for multiple grade levels, a total of only four different passages were used. These four passages are included in Appendix A.

The Maze

One of the two basal-derived measures was a multiple-choice cloze or "Maze" (Guthrie et al., 1974). Excluding the first and last sentence of a 250-word passage, every sixth word was deleted, and students were offered five alternatives in a "key-word search" task (Roid & Haladyna, 1982). The four distractors were random selections from the pool of deleted words. Each Maze contained approximately 34 items. The dependent measure was "percent correct," averaged over three passages. The maze tests used on the four passages are included in Appendix B.

Oral Reading Fluency

The second basal-derived reading measure, Oral Reading Fluency, was obtained on the same three (intact) 250-word passages. Errors counted were additions, substitutions, omissions, and hesitations of at least 3 seconds, after which students were told the word. Students were corrected immediately if they skipped to the wrong line of text (counted as only one error). The dependent measure was the percent of words read correctly per minute, averaged over the three passages.

Procedures

In March 1988, 12 teachers consecutively administered three Maze tests to their reading groups from scripted directions in a 20-minute session. The presentation order of the three tests was counterbalanced within each group. Over the next 2 weeks, subjects were withdrawn from class individually to read aloud the same three passages and to complete the ARI. The Oral Reading/ARI administration order was counterbalanced within groups, as was the order in which the three passages were read. The Maze/Oral Reading order could not be counterbalanced because Oral Reading would reveal answers to the Maze.

ARI and Oral Reading Fluency assessments required approximately 27 minutes per student (15-50 minute range). The ARI alone required approximately 18 minutes for most students (8-35 minute range), while the Oral Reading Fluency assessment was somewhat faster, averaging 9 minutes (6-15 minute range).

Complete data from all students could not be collected. Usable Maze results were available for only 103 of the original 120 students. Because individual testing sessions could be rescheduled, results were available for 117 of the 120 students.

RESULTS

Predictive Validity of Measures

Table 1 presents descriptive summaries for each of the three measures by reading group, as well as by grade level across reading groups. All measures showed an increase in mean scores across consecutive grades, and from lower to higher instructional groups. The lone exception to this trend was a higher-than-expected ARI score from the single student in the Grade 6 SPED/Chapter I group. This trend of greater mean scores across grade levels is evident for the Maze and Oral Reading Fluency, despite the fact that students in the higher grades were presented with more difficult text (see Table 1).

A major objective of this study was to assess the relative accuracy of the three objective reading measures in matching cross-classroom reading group placements based on subjective teacher judgments. Following a preliminary Bartlett's test for equal variances, ANOVAs were conducted on differences among reading groups at each grade level for each of the three tests (Maze, Fluency, and ARI). The *F* ratios in Table 2 reflect highly significant differences at all grade levels.

The generalized correlation ratio or "eta" (Kerlinger, 1986, pp. 216-217) computed for each *F* test indicate that between 54% and 81% of the variance in Maze scores could be accounted for by reading group membership. For Oral Reading Fluency scores, the proportion of variance explained by groups ranged from 52% to 66%, and for the ARI, the proportion ranged from 64% to 69%. All of the measures appeared to be moderately strong predictors of instructional group membership, with the ARI the most stable predictor (across the three grades) and the Maze somewhat stronger on the average than the other two.

Multiple discriminant function analysis (DFA), an analytical tool for assessing accuracy in predicted group placements from single or multiple test scores (Klecka, 1981), was used for this study. Predicted placements produced from DFA from Maze, Oral Reading Fluency, and ARI scores were cross-tabulated with actual placements, and the resulting table was analyzed with Kendall's Tau and Cohen's Kappa, two indices of relationship that are most sensitive to the number of clas-

Table 2: Tests of Homogeneity of Variance and Mean Differences Among Classroom Groups in Grades 4, 5, and 6 on Maze, Oral Reading Fluency, and Analytical Reading Inventory, with Effect Size for Each Measure

	Bartlet's Variance Test			Mean Differences			Effect Size	
	F	p	df	SS _b	SS _w	F	p	eta
<u>Grade 4</u>								
Maze	1.41	.23	4, 32	11159.3	2873.5	31.1	<.000	.89
ORF	1.45	.22	4, 34	25735.5	13139.2	16.6	<.0001	.81
ARI	1.08	.36	4, 34	66.1	28.8	19.5	<.0001	.83
<u>Grade 5</u>								
Maze	1.0	.40	4, 29	3398.6	781.7	31.5	<.0001	.90
ORF	.41	.80	4, 35	20009.5	18788.1	9.3	<.0001	.72
ARI	.83	.51	4, 35	66.8	29.9	19.5	<.0001	.83
<u>Grade 6*</u>								
Maze	.68	.57	3, 27	1297.7	1811.8	6.4	.002	.74
ORF	.50	.68	3, 33	15618.0	14195.6	12.1	<.0001	.74
ARI	2.6	.05	3, 33	63.5	37.3	18.7	<.0001	.80

*Grade 6 SPED/Chapter I group has only one member, so ANOVA was conducted on the other four groups only.

sification errors (Klecka, 1980; Fleiss, 1981).

Table 3 contains the "percent of correct classifications" from the DFA procedure by each of the three discriminant measures separately, as well as by Maze and Oral Reading Fluency together. The relatively low percents (28-62%) of direct Hits improved considerably (none below 82%), when a Hit is redefined as "within an error band of one category."

Table 3 also presents the canonical correlation coefficients for the discriminant functions produced from the DFA procedure. For the Maze and Oral Reading Fluency combined, coefficients are presented for only the first function, as the second function produced non-significant correlations. Standardized weights for the Maze and Oral Reading Fluency, respectively, were: .76 and .42 for Grade 4, .95 and .16 for Grade 5, and .57 and .59 for Grade 6.

DFA also yields canonical loadings, similar to factor loadings in factor analysis. For the Maze plus Oral Reading Fluency, the respective loadings for the first discriminant function were: .92 and .71 for Grade 4; .99 and .41 for Grade 5; and .85 and .87 for Grade 6. The similarity between standardized weights and canonical loadings is considered supportive of the DFA results (Klecka, 1980).

Table 3 includes Kendall's Tau (Klecka, 1980, p. 51) and Cohen's Kappa (Fleiss, 1981, pp. 212-

236), two indices of relationship between the discriminant measures and group membership that were computed from analysis of the actual versus predicted cross-tabulations. Hits/Misses, Kendall Tau, and Cohen's Kappa statistics computed on the combined crosstabs rank the measures similarly: Maze plus Oral Reading Fluency first, followed by ARI, Maze, and Oral Reading Fluency.

Across the three grades, the Maze plus Oral Reading Fluency classified most accurately, followed by the ARI. Oral Reading Fluency was reasonably strong for all grades but 5, while the Maze was very strong in Grade 5 and weak in Grade 6. The ARI, while not the strongest, was reasonably consistent across the three grade levels. Tau and Kappa, neither of which is sensitive to "degree of error," confirmed this rating of the tests on Hits/Misses. When the degree of classification error ("within one group") was considered, however, the Maze alone classified most accurately.

In addition to yielding *categorical* predicted placements, DFA calculates, as an intermediate step, a set of probabilities that each case belongs to each one of the groups (Klecka, 1980, pp. 44-45). These probabilities (based on Mahalanobis distances from group centroids) offer a more detailed picture of test results.

The graphs included in Appendix C depict these multiple probabilities within computer-gen-

erated (Wilkinson, 1987) stacked area graphs for ARI, Maze, and Oral Reading Fluency scores at the Grade 4 level only. The Maze plus Oral Reading Fluency graph is not included. Probabilities of belonging to each group are portrayed as patterned areas under a curve. The ARI graph also has been labeled with representative probability values. Generally, the steeper curves and bands of less similar widths indicate more accurate discrimination between groups. The stacked area graphs for Grades 5 and 6 are included in Appendix D.

At the top of the stacked area graphs (marked by solid vertical lines) are the predicted placements in deterministic, categorical terms. A comparison of the categorical-versus-probabilistic placements gives a sense of the degree of error in making any particular categorical placement.

DISCUSSION

One purpose of this study was to assess the accuracy of three objective reading measures (IRI, Maze, Oral Reading Fluency) in matching teachers'

cross-classroom reading group placement decisions based on informal observation and other non-test score criteria. Approximately 35 students were selected from five reading groups at Grades 4, 5, and 6. The ARI was chosen because of the popularity of Informal Reading Inventories among teachers. The other two measures, Oral Reading Fluency and the Maze, are curriculum based and easily produced and administered by the classroom teacher.

When placement decisions must be made consensually and when student information is drawn from different teachers and classrooms, objective reading measures allow results to be compared across decision-makers.

This study also sought to demonstrate a method of summarizing and displaying the discriminative accuracy of these three measures in a manner that highlights those placement decisions where informal teacher judgment and other, non-test criteria are needed most. Research by Strike (1983), Haller and Waterman (1985), and Salmon-

Table 3: Classification Accuracy of Two Basal Reader-Based Measures (Maze and Oral Reading Fluency) and the Analytical Reading Inventory at Three Grade Levels

	N	Percent Correct		First Discrim Function	Tau B	Kappa
		Hits	Within One			
<u>Grade 4</u>						
Maze	37	46	97	.89**	.69**	.41**
ORF	39	51	89	.81**	.65**	.37*
Maze + ORF	36	56	94	.88**	.68**	.41**
ARI	39	46	87	.83**	.67**	.31*
<u>Grade 5</u>						
Maze	34	62	100	.90**	.80**	.51**
ORF	40	40	83	.72**	.60**	.23
Maze + ORF	34	59	100	.90**	.79**	.47**
ARI	40	55	98	.83**	.74**	.42**
<u>Grade 6</u>						
Maze	32	28	91	.74**	.50**	.06
ORF	38	50	87	.74**	.64**	.36*
Maze + ORF	32	48	84	.77**	.57**	.33*
ARI	38	58	82	.80**	.63**	.47**
<u>Grades 4, 5, & 6 combined</u>						
Maze	103	49	95		.66**	.34**
ORF	117	47	86		.62**	.32**
Maze + ORF	103	54	94		.68**	.41**
ARI	117	53	89		.67**	.40**

*p < .01

**p < .001

Cox (1981) emphasizes that teachers neither trust nor use standardized tests in making reading-group placements. They instead rely upon a wide variety of informal criteria, including judgments based upon daily observations. For objective test data to be used better by teachers, they should be assisted in understanding the relationship between the data and prevailing decision-making practices. In addition, the limitations or inaccuracies of the data should be made clearer to emphasize when other judgment criteria are most needed.

Classification Accuracy

Results of discriminant function analyses (DFA) indicated that all three measures were moderately good predictors of existing teacher groupings, accounting for roughly 50-80% of the variability among groups. When defined as Hits and Misses, the accuracy of the measures was less impressive, however, with Hit rates of only 47-53% across the three grades. The Hit rates increased to 86-95% when the definition of a Hit was broadened to include "within one category of actual placement." The higher figures are less reliable, however, as they capitalize more upon chance.

The classification accuracy of the three measures can be summarized in two ways: across the full range of scores or (using the stacked area graphs) by limited score ranges (high, medium, low). Across the full range of scores, the Maze was the strongest predictor at Grades 4 and 5, and the weakest at Grade 6; the ARI was the strongest Grade 6 measure. The Maze/Oral Reading Fluency combination did not substantially improve the Maze's predictive efficiency at Grades 4 and 5. On the basis of these results, the Maze alone could be recommended at Grades 4 and 5, but at Grade 6 neither of the curriculum-based measures compared favorably with the ARI.

When the definition of a Hit was broadened to include "within one category of actual placement," the Maze was the strongest predictor in Grades 4, 5, and 6. This broader definition may be adequate for the placement needs of some schools.

By referring to the three Grade 4 stacked area graphs (see Appendix C), one also can assess the classification accuracy of the three measures for limited ranges of scores. The graphs show that there is not "one best test" for predicting reading group placements made on the basis of informal teacher judgment. The test(s) of choice will depend largely on the approximate skill level of a student being placed. For Grade 4, the Maze was strongest in the Low score range, followed by Oral Reading

Fluency. In the Low-Medium range, none of the measures were particularly strong, but the ARI and Oral Reading Fluency displayed reasonable power for higher scores.

At Grade 5 (graphs not shown), the Maze appeared best for all scores except for those in the High-Medium range. Oral Reading Fluency and the ARI performed similarly and could be recommended only for the highest and lowest scores.

At Grade 6, the ARI was strong among lower scores and reasonably strong for all other score ranges. The Maze was best among lower score ranges and Oral Reading Fluency among higher scores; both lacked power for scores in the middle range.

The differential effectiveness of the three tests at low, medium, and high skill levels may be due in part to an attenuated distribution of scores on the Maze. The 100% correct ceiling on the Maze was relatively attainable, while the ceiling on the ARI (Grade 9 level) was more difficult to reach, and Oral Reading Fluency scores had no imposed ceiling. Considering the predictive strength of the Maze, modifications in that instrument that raise the ceiling, yet do not sacrifice replicability in construction or scoring, might be warranted.

Teacher Use of the Stacked Area Graphs

The stacked area graphs have three main uses. First, they improve placement precision by suggesting a test that best matches a student's "guesstimated" reading ability. The teacher should make an initial "ballpark" judgment and select only a test that can significantly improve upon that judgment. Generally, the Maze can be recommended for low-performing students. For middle scores, the three measures are similar, with the Maze only slightly stronger. For high-performing students, either Oral Reading Fluency or the ARI appear suitable.

The second use of the area graphs is to indicate how much group discrimination power is available for a particular scores and score ranges. For example, for a relatively high Oral Reading Fluency score of 118 at the Grade 4 level, the student could, with almost equal probability, be placed in the High, High-Medium, or the Low-Medium groups. If the teacher's concern is in which of these groups to place the student, some other criteria must be used.

The third use of the area graphs is to identify particular placements or placement options with a known probability of being wrong. For example, for a low ARI score of 3 at Grade 4, one can be 77%

certain that the SPED/Chapter I placement is best. The alternative Low group placement, with a meager 20% chance of being "correct," would be seriously considered only if other criteria were given considerable weight.

Efficiency of Use

Because of its group administration, the Maze clearly was the most time-efficient reading measure. This efficiency must be weighed against the lack of opportunity it affords to observe students in the process of reading/an aspect of IRIs rated highly by teachers. The anonymity of group testing allowed a handful of students to merely guess on the Maze, an unlikely phenomenon in individual test administration. Oral Reading Fluency is relatively time-efficient and also allows observation of the reading process.

It may be tempting to group-administer the Maze as an efficient screener, followed by the more time-consuming ARI or Oral Reading Fluency. However, Oral Reading Fluency appeared to add significantly to the predictive power of the Maze only in Grade 6. At that level, the Maze would be a weak screening device.

The modified ARI was included in the study because of its popularity and as a standard against which other measures could be judged. Even in a shortened and streamlined form, the ARI was considerably more demanding in terms of the training, administration, and scoring time requirements than were the other two measures. This additional time appeared warranted by the DFA results for Grade 6 but not for Grades 4 and 5.

CONCLUSION

Graphing multiple probabilities provides information to teachers beyond that afforded by recommended categorical placements, qualified by confidence intervals. With the stacked area graphs, a teacher can select the best test for a student's approximate skill level. From test results, the teacher also can gauge the relative certainty or accuracy with which *each one* of the alternative group placements can be made. In those cases of relatively equal predicted accuracy (equal heights of the patterned areas), the teacher must rely upon other criteria to make the decision. The graphs thus clearly depict where data from classroom observation, student work habits, attendance, family support, and social skills might be needed to make placement decisions.

REFERENCES

- Alexander, H. (1968). *An investigation of the cloze procedure as a measuring device designed to identify the independent, instructional, and frustration reading levels of pupils in the intermediate grades*. Unpublished doctoral dissertation, University of Illinois, Champaign, IL.
- Borko, H., & Shavelson, R. J. (1982). Speculations on teacher education: Recommendations from research on teachers' cognitions. *Journal of Education for Teaching, 8*, 279-283.
- Bormouth, J. R. (1968). Close test readability: Criterion reference scores. *Journal of Educational Measurement, 5*, 291-299.
- Bradley, J., Ackerson, G., & Ames, W. (1978). The reliability of the maze procedure for classroom assessment. In P. D. Pearson & J. Hansen (Eds.), *Reading: Disciplined inquiry in process and practice*, Twenty-Seventh Yearbook of the National Reading Conference. Clemson, SC: National Reading Conference.
- Bradley J., & Meredith, K. (1978). The reliability of the maze procedure for intermediate and junior high school students. In P. D. Pearson & J. Hansen (Eds.), *Reading: Disciplined inquiry in process and practice*, Twenty-Seventh Yearbook of the National Reading Conference. Clemson, SC: National Reading Conference.
- Brecht, R.D. (1977). Testing format and instructional level with the informal reading inventory. *Reading Teacher, 31*(1), 57-59.
- Bristow, P.S., Pikulski, J.J., & Pelosi, P.L. (1983). A comparison of five estimates of reading instructional level. *The Reading Teacher, 37*(3), 276-279.
- Burry, J., Catterall, J., Choppin, B., & Dorr-Bremme, D. (1982). *Testing in the nation's schools and districts: How much? What kinds? To what ends? At what costs?* CSE Report No. 194, Center for the Study of Evaluation, Graduate School of Education, UCLA.
- Coleman, M., & Harmer, W. R. (1982). A comparison of standardized tests and informal placement procedures. *Journal of Learning Disabilities, 9*, 481-487.
- Deno, S., Marston, D., Mirkin, P., Lowry, L., Sindelar, P., & Jenkins, J. (1982). *The use of standard tasks to measure achievement in reading, spelling and written expression: A normative and developmental study* (Research Report No. 87). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

- Deno, S., Marston, D., Shinn, M., & Tindal, G. (1983). Oral reading fluency: A simple datum for scaling disability. *Topics in Learning and Learning Disabilities*, 2(4), 53-59.
- Deno, S. L., Mirkin, P., & Chiang, G. (1982). Identifying valid measures of reading. *Exceptional Children*, 49(1), 36-45.
- Doyle, W. (1986). Classroom organization and management. In M. C. Wittrock (Ed.), *Handbook of research on teaching*. New York: MacMillan.
- Ekwall, E. E. (1974). Should repetitions be counted as errors? *Reading Teacher*, 27, 365-367.
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York: John Wiley & Sons.
- Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education*, 9(2), 20-28.
- Fry, E. A. (1968). A readability formula that saves time. *Journal of Reading*, 11, 513-516.
- Goslin, D. A. (1967). *Teachers and testing*. New York: Russell Sage Foundation.
- Gump, P. (1987). School and classroom environments. In D. Stokols & I. Altman (Eds.), *Levels of environmental analysis: Situations, settings, and places*. New York: John Wiley & Sons.
- Guthrie, J. T. (1973). Models of reading and reading disability. *Journal of Educational Psychology*, 65, 9-18.
- Guthrie, J.T., Seifert, M., Burnham, N.A., & Caplan, R.I. (1974). The maze technique to assess, monitor reading comprehension. *The Reading Teacher*, 28(2), 161-68.
- Haller, E. J., & Waterman, M. (1985). The criteria of reading group assignments. *The Reading Teacher*, 38, 772-781.
- Harris, L.A., & Lalik, R.M., (1987). Teachers' use of informal reading inventories: An example of school constraints. *The Reading Teacher*, 40(7), 624-630.
- Harris, L.A., & Niles, J.A. (1982). An analysis of published informal reading inventories. *Reading Horizons*, 22(3), 159-174.
- Kender, J.P. & Rubenstein, H. (1977). Recall versus reinspection in IRI comprehension tests. *The Reading Teacher*, 30, 776-779.
- Kerlinger, F. N. (1986). *Foundations of behavioral research* (3rd ed). New York: Holt, Rinehart, & Winston.
- Klecka, W. R. (1980). *Discriminant analysis*. Beverly Hills, CA: Sage Publications Ltd.
- Leinhardt, G., Weidman, C., & Hammond, K. M. (1987). Introduction and integration of classroom routines by expert teachers. *Curriculum Inquiry*, 17(2), 135-176.
- Marr, M.B., & Lyon, K.R. (1980). Passage independence: An examination of three informal reading inventories. In M. L. Kamil (Ed.), *Perspectives on reading research and instruction: 29th yearbook of the National Reading Conference* (pp. 283-287). Washington, DC: National Reading Conference.
- Marston, D. & Deno, S. (1982). *Implementation of direct and repeated measurement in the school setting* (Research Report No. 106). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities. (ERIC Document Reproduction Service No. ED 226 048).
- Marston, D., & Magnusson, D. (1985). Implementing curriculum-based measurement of academic skills: An alternative to traditional screening, referral, and identification of learning disabled students. *Journal of Special Education*, 18, 109-118.
- Orlich, D. C., Hardner, R. J., Callahan, R. C., Kravas, C. H., Kauchak, D. P., Pendergrass, R. A., & Keogh, A. J. (1985). *Teaching strategies: A guide to better instruction*. Massachusetts: D.C. Heath.
- Page, W.D., (1976). Miscue research and diagnosis. In P.D. Allen & D.J. Watson (Eds.), *Findings of research in miscue analysis: Classroom implications*. Urbana, IL: National Council of Teachers of English.
- Peterson, P. L. (1988). Teachers' and students' cognitional knowledge for classroom teaching and learning. *Educational Researcher*, 17(5), 5-14.
- Pflaum, S. W. (1980). Diagnosis of oral reading. *Reading Teacher*, 33(3).
- Pikulski, J. J. (1974). A critical review: Informal inventories. *The Reading Teacher*, 28, 141-151.
- Pikulski, J.J., & Shanahan, T. (1982). Informal reading inventories: A critical analysis. In J.J. Pikulski & T. Shanahan (Eds.), *Approaches to the informal evaluation of reading*. Newark, DE: International Reading Association.
- Pikulski, J. J., & Tobin, A. W. (1982). The cloze procedure as an informal assessment technique. In J. J. Pikulski & T. Shanahan (Eds.), *Approaches to the informal evaluation of reading*. Newark, DE: International Reading Association.
- Powell, W.R. (1968). Reappraising the criteria for interpreting Informal Inventories. In D. L. DeBoer (Ed.), *Reading diagnosis and evaluation: Proceedings of the 13th annual Convention of the*

- International Reading Association*, Newark, DE: International Reading Association.
- Rankin E. F., & Culhane, J. (1969). Comparable cloze and multiple-choice comprehension scores. *Journal of Reading*, 13, 193-198.
- Resnick, D. P. (1981). Testing in America: A supportive environment. *Phi Delta Kappan*, 631-634.
- Roid, G., & Haladyna, T. (1982). Items for prose writing. In G. Roid & T. Haladyna, (Eds.), *A technology for test-item writing*. New York: Academic Press.
- Salmon-Cox, L. (1981). Teachers and standardized achievement: What's really happening? *Phi Delta Kappan*, 62, 631-634.
- Schwartz, S. (1984). *Measuring reading competence: A theoretical-prescriptive approach*. New York: Plenum Press.
- Scott, Foresman Reading Series (1987). *An American Tradition*. Chicago, IL: Scott, Foresman & Co.
- Spache, G. (1976). *Diagnosing and correcting reading disabilities*. Boston: Allyn & Bacon.
- Sproull, L., & Zubrow, D. (1981). Standardized testing from the administrative perspective. *Phi Delta Kappan*, 628-631.
- Stetz, F., & Beck, M. (1979). *Teachers' opinions of standardized test use and usefulness*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Strike, K. A. (1983). Fairness and ability grouping. *Educational Theory*, 33, 125-134.
- Wicker, A. W. (1987). Behavior settings reconsidered: Temporal stages, resources, internal dynamics, context. In D. Stokols & I. Altman (Eds.), *Levels of environmental analysis: Situations, settings, and places*. New York: John Wiley & Sons.
- Wilkinson, L. (1987). *SYSTAT: The system for statistics*. [computer program]. Evanston, IL: SYSTAT, Inc.
- Woods, M. L., & Moe, A.J., (1985). *Analytical Reading Inventory* (3rd ed.). Columbus, OH: Charles E. Merrill.
- Ysseldyke, J. (1979). Psychoeducational decision-making. In J. E. Ysseldyke & P. K. Mirkin (Eds.), *Proceedings of the Minnesota Roundtable Conference on Assessment of Learning Disabled Children* (Monograph No. 8). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

APPENDIX A
BASAL READING PASSAGES

Jim Gary's Giants

What do you think of when you ride past a junkyard? Do you see just old cars and auto parts? Jim Gary sees animals and a great place to visit. Gary is a sculptor with a strange specialty. He makes giant animals—from junk!

Gary didn't start out enjoying junk. As a boy he was crazy about dinosaurs. "I had my own collection of model dinosaurs," he recalls proudly. Then, as he grew older, he fell in love with cars.

"I didn't have enough money to buy a car," Gary says. "So I just fooled around with them. Then I started building my own cars.

"I went to the junkyards near my home for most of the parts I needed. Then one day I was just sitting in the middle of the junk, looking around. All of a sudden, the car parts started looking like animal parts to me. An oil pan looked like part of a skull. An axle became an arm. Nuts and bolts looked like eyes and ears.

"I was only fourteen," states Gary, "but I knew that I wanted to make animals out of these old auto parts. I only owned a few tools and some welding gear, but I had plenty of ideas."

The Turtle Street Trading Co.

"What kind of business could four kids go into?" Morgan asked.

"I don't know, Fran replied.

"There's got to be some business we could go into. Come on, Morgan. Think!" P.J. demanded.

Morgan told the others about the most successful businessperson he knew—his dad. His dad was always talking business. Business, business, business, that's all he ever thought about. He owned not one but three stores already, and he was thinking about opening a fourth. They all had the same name: Peter's Party Supply Store. And they all made money.

Morgan's dad has started from the bottom and worked his way up from there. He said that if you wanted to be a success in your own business, you had to keep in mind the basic principle behind all American big business—supply and demand. That meant you had to dream up an idea for a product or service the public really needed—even if it didn't know it yet—make it available, then convince the public that it couldn't live without it.

What product or service could four inexperienced kids make available to the public that somebody else hadn't already made available?

That was a toughie!

Life on a Space Shuttle

The chances that you may one day take a trip into space are getting better all the time. As it becomes easier to live in a space capsule, more and more people who are not trained astronauts will be able to travel in space. At first, the National Aeronautics and Space Administration (NASA) would only choose as astronauts trained pilots to carry out the early space programs. These were the Mercury and Gemini programs. For the Apollo and Skylab flights, doctors and scientists were added to the list of those able to take part in space flights.

A space shuttle crew may number up to seven people. The commander, the pilot, and the mission specialist are NASA astronauts. Other people in the crew are called payload specialists. They may or may not be NASA astronauts. Payload specialists are people who are already trained in special scientific or medical skills. They are taken along on spaceflights to carry out the scientific and medical experiments of a mission. What special skills of your chosen career might be useful on a space shuttle mission? This is where your chance to travel in space might be.

Today, choosing candidates for space travel is still hard to do.

Beauty and the Beast

Once upon a time there was a merchant whose wife died, leaving him with three daughters.

The two older daughters were good-looking but very disagreeable. They cared only for themselves and for their appearance; they spent most of the time admiring their reflections in a looking-glass.

The third and youngest daughter was quite different from the other two. She was beautiful—so beautiful that she was known as Beauty. She was also good and kind. Everyone loved Beauty, except for her sisters, who were jealous of her. They hated her.

One day the merchant heard that a ship was expected in a far-away port with a valuable cargo for him. He prepared at once to set off on the long journey to the port, for he hoped to make a great deal of money by selling the cargo there. Knowing of this, his two elder daughters hung about him, pestering him to bring expensive presents back to them. They wanted dresses of silk and satin and jewelry of gold and precious stones. Only Beauty remained silent: all that she wanted was that her father should bring himself home again safe and sound. She dared not say so, for fear of the jeering sisters.

APPENDIX B
MAZE TESTS FOR THE FOUR PASSAGES

Date _____	Name _____	Grade _____
Teacher _____		School _____

Life on a Space Shuttle

The chances that you may one day take a trip into space are getting better all the time. As it becomes easier to ___(1)___ in a space capsule, more ___(2)___ more people who are not ___(3)___ astronauts will be able to ___(4)___ in space. At first, the ___(5)___ Aeronautics and Space Administration (NASA) ___(6)___ only choose as astronauts trained ___(7)___ to carry out the early ___(8)___ programs. These were the Mercury ___(9)___ Gemini programs. For the Apollo ___(10)___ Skylab flights, doctors and scientists ___(11)___ added to the list of ___(12)___ able to take part in ___(13)___ flights.

A space shuttle crew ___(14)___ number up to seven people. ___(15)___ commander, the pilot, and the ___(16)___ specialist are NASA astronauts. Other ___(17)___ in the crew are called ___(18)___ specialists. They may or may ___(19)___ be NASA astronauts. Payload specialists ___(20)___ people who are already ___(21)___ in special scientific or medical ___(22)___ . They are taken along on ___(23)___ to carry out the scientific ___(24)___ medical experiments of a mission. ___(25)___ special skills of your chosen ___(26)___ might be useful on a ___(27)___ shuttle mission? This is where ___(28)___ chance to travel in space ___(29)___ be.

Today, choosing candidates for space travel is still hard to do.

DIRECTIONS:

For each blank in the story, circle the word below that fits the best.

- (1) trained, space, live, were, space
- (2) may, and, National, mission, might
- (3) are, and, skills, trained, the
- (4) your, spaceflights, trained, travel, not
- (5) would, skills, and, National, your
- (6) were, mission, are, and, would
- (7) career, might, what, spaceflights, pilots
- (8) those, your, space, skills, were
- (9) travel, live, payload, skills, and
- (10) and, what, would, career, the
- (11) spaceflights, National, space, were, career
- (12) mission, what, might, space, those
- (13) are, what, space, travel, people
- (14) live, may, not, career, travel
- (15) What, Space, The, Your, Travel
- (16) skills, and, pilots, mission, would
- (17) might, people, National, might, were
- (18) payload, your, career, and, those
- (19) the, may, space, live, not
- (20) travel, would, mission, are, career
- (21) live, trained, are, what, skills
- (22) might, pilots, and, skills, your
- (23) space, mission, spaceflights, may, and
- (24) and, people, not, the, those
- (25) Pilots, What, Not, Mission, May
- (26) trained, space, payload, those, career
- (27) people, pilots, space, the, may
- (28) mission, your, payload, and, live
- (29) are, career, might, what, those

Date _____ Name _____ Grade _____
 Teacher _____ School _____

Beauty and the Beast

Once upon a time there was a merchant whose wife died, leaving him with three daughters.

The two older daughters were ___(1)___ but very disagreeable. They cared ___(2)___ for themselves and for their ___(3)___; they spent most of the ___(4)___ admiring their reflections in a ___(5)___.

The third and youngest daughter ___(6)___ quite different from the other ___(7)___ . She was beautiful—so beautiful ___(8)___ she was known as Beauty. ___(9)___ was also good and kind. ___(10)___ loved Beauty, except for her ___(11)___, who were jealous of her. ___(12)___ hated her.

One day the ___(13)___ heard that a ship was ___(14)___ in a far-away port with ___(15)___ valuable cargo for him. He ___(16)___ at once to set off ___(17)___ the long journey to the ___(18)___, for he hoped to make ___(19)___ great deal of money by ___(20)___ the cargo there. Knowing of ___(21)___ his two elder daughters hung ___(22)___ him, pestering him to bring ___(23)___ presents back to them. They ___(24)___ dresses of silk and satin ___(25)___ jewelry of gold and precious ___(26)___ . Only Beauty remained silent: all ___(27)___ she wanted was that her ___(28)___ should bring himself home again ___(29)___ and sound. She dared not say so, for fear of the jeering sisters.

DIRECTIONS:

For each blank in the story, circle the word below that fits the best.

- (1) safe, a, good-looking, and, looking-glass
- (2) only, good-looking, expensive, about, father
- (3) prepared, sisters, appearance, only, she
- (4) that, she, time, they, appearance
- (5) on, this, looking-glass, sisters, stones
- (6) wanted, two, was, time, only
- (7) was, selling, two, merchants, safe
- (8) that, and, appearance, expensive, they
- (9) prepared, she, father, that, a
- (10) Expensive, Two, Everyone, Sisters, Port
- (11) sisters, this, that, stones, about
- (12) Wanted, They, Time, Only, Good-looking
- (13) that, a, about, safe, merchant
- (14) time, wanted, expected, they, on
- (15) she, a, was, selling, and
- (16) stones, she, prepared, looking-glass, two
- (17) appearance, on, prepared, this, sisters
- (18) selling, everyone, port, they, two
- (19) port, safe, father, a, everyone
- (20) that, selling, this, time, father
- (21) she, expensive, this, stones, was
- (22) sisters, that, good-looking, only, about
- (23) expensive, about, expected, on, wanted
- (24) merchant, this, wanted, two, that
- (25) port, and, expected, sisters, father
- (26) a, selling, appearance, stones, she
- (27) that, stones, safe, only, everyone
- (28) they, merchant, about, father, prepared
- (29) time, safe, wanted, was, on

Date _____ Name _____ Grade _____
 Teacher _____ School _____

The Turtle Street Trading Co.

"What kind of business could four kids go into?"

Morgan asked.

"I don't know," Fran replied.

"___(1)___ got to be some business ___(2)___ could go into. Come on, ___(3)___ Think!" P.J. demanded.

Morgan told ___(4)___ others about the most successful ___(5)___ he knew—his dad. His ___(6)___ was always talking business. Business, ___(7)___ business, that's all he ever ___(8)___ about. He owned not one ___(9)___ three stores already, and he ___(10)___ thinking about opening a fourth ___(11)___ all had the same name: ___(12)___ Party Supply Store. And they ___(13)___ made money.

Morgan's dad had ___(14)___ from the bottom and worked ___(15)___ way up from there. He ___(16)___ that if you wanted to ___(17)___ a success in your own ___(18)___, you had to keep in ___(19)___ the basic principle behind all ___(20)___ big business—supply and demand. ___(21)___ meant you had to dream ___(22)___ an idea for a product ___(23)___ service the public really needed— ___(24)___ if it didn't know it ___(25)___—make it available, then convince ___(26)___ public that it couldn't live ___(27)___ it.

What product or service could ___(28)___ inexperienced kids make available to ___(29)___ public that somebody else hadn't ___(30)___ made available? That was a toughie!

DIRECTIONS:

For each blank in the story, circle the word below that fits the best.

- (1) We, There's, Thought, His, Be
- (2) even, we, started, without, that
- (3) Morgan, business, they, the, that
- (4) American, even, already, that, the
- (5) mind, four, already, the, businessperson
- (6) even, dad, thought, we, without
- (7) business, four, even, the, was
- (8) up, thought, or, the, Peter's
- (9) the, mind, already, but, said
- (10) even, already, yet, was, up
- (11) American, Even, Up, Or, They
- (12) Without, Peter's, four, There's, yet
- (13) all, business, the, Morgan, or
- (14) already, the, all, American, started
- (15) be, without, even, his, dad
- (16) that, Peter's, but, said, four
- (17) business, without, up, be
- (18) business, businessperson, yet, dad, up
- (19) the, mind, Peter's, but, there's
- (20) four, said, American, even, thought
- (21) Already, All, That, Business, The
- (22) American, up, but, businessperson, the
- (23) all, dad, mind, or, Morgan
- (24) without, even, mind, Morgan, yet
- (25) yet, dad, be, businessperson, or
- (26) four, was, said, there's, the
- (27) without, thought, started, his, they
- (28) Morgan, business, already, four, was
- (29) was, but, up, the, mind
- (30) said, mind, they, without, already

Date _____ Name _____ Grade _____
 Teacher _____ School _____

Jim Gary's Giants

What do you think of when you ride past a junkyard? Do you see just old ___(1)___ and auto parts? Jim Gary ___(2)___ animals and a great place ___(3)___ visit. Gary is a sculptor ___(4)___ a strange specialty. He makes ___(5)___ animals—from junk

Gary didn't ___(6)___ out enjoying junk. As a ___(7)___ he was crazy about dinosaurs. "___(8)___ had my own collection of ___(9)___ dinosaurs," he recalls proudly. Then, ___(10)___ he grew older, he fell ___(11)___ love with cars.

"I didn't ___(12)___ enough money to buy a ___(13)___," Gary says. "So I just ___(14)___ around with them. Then I ___(15)___ building my own cars.

"I ___(16)___ to the junkyards near my ___(17)___ for most of the parts ___(18)___ needed. Then one day I ___(19)___ just sitting in the middle ___(20)___ the junk, looking around. All ___(21)___ a sudden, the car parts ___(22)___ looking like animal parts to ___(23)___ . An oil pan looked like ___(24)___ of a skull. An axle ___(25)___ an arm. Nuts and bolts ___(26)___ like eyes and ears.

"I ___(27)___ only fourteen," states Gary, "but ___(28)___ knew that I wanted to ___(29)___ animals out of these old ___(30)___ parts. I only owned a few tools and some welding gear, but I had plenty of ideas."

DIRECTIONS:

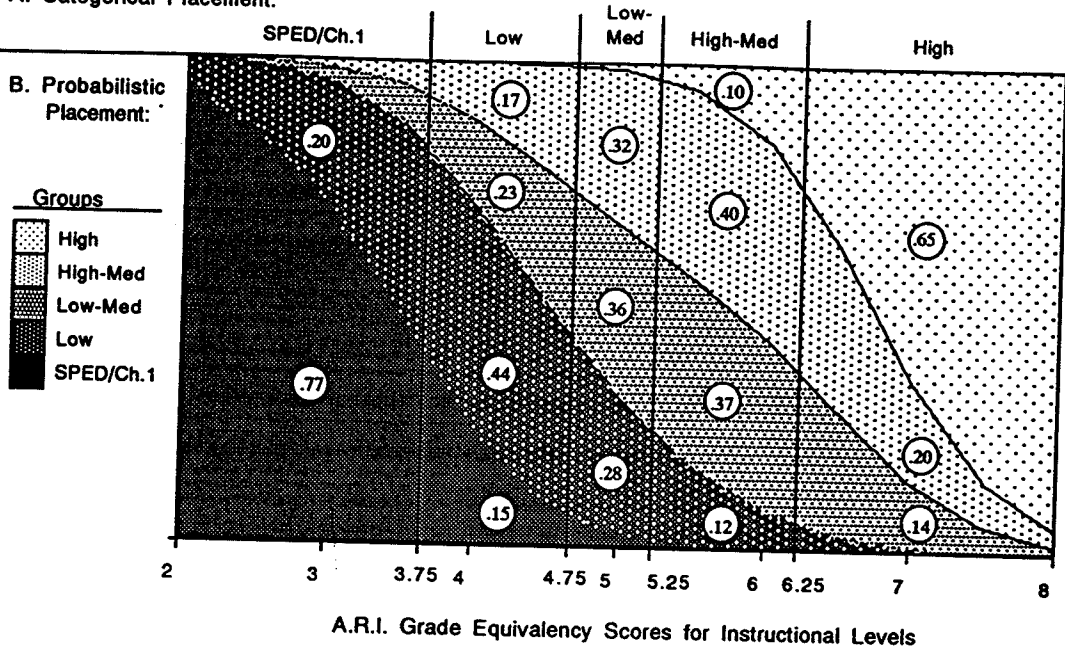
For each blank in the story, circle the word below that fits the best.

- (1) sees, cars, home, started, of
- (2) sees, car, became, to, have
- (3) auto, giant, was, in, to
- (4) I, with, went, model, boy
- (5) I, cars, of, giant, looked
- (6) I, start, started, part, fooled
- (7) auto, have, cars, boy, sees
- (8) home, I, was, with, started
- (9) I, was, make, me, model
- (10) model, I, became, as, looked
- (11) fooled, in, boy, giant, to
- (12) boy, as, have, in, of
- (13) of, have, boy, started, car
- (14) fooled, of, auto, was, giant
- (15) make, with, started, I, to
- (16) start, liked, went, me, model
- (17) home, started, in, went, sees
- (18) became, part, I, fooled, home
- (19) car, of, auto, boy, was
- (20) was, of, giant, I, fooled
- (21) started, of, model, became, to
- (22) of, looked, make, sees, started
- (23) as, me, have, car, started
- (24) I, home, part, start, with
- (25) became, part, boy, start, looked
- (26) cars, with, me, looked, have
- (27) of, became, giant, in, was
- (28) as, part, I, car, was
- (29) auto, make, of, fooled, sees
- (30) to, started, auto, was, model

APPENDIX C
STACKED AREA GRAPHS
FOR GRADE 4

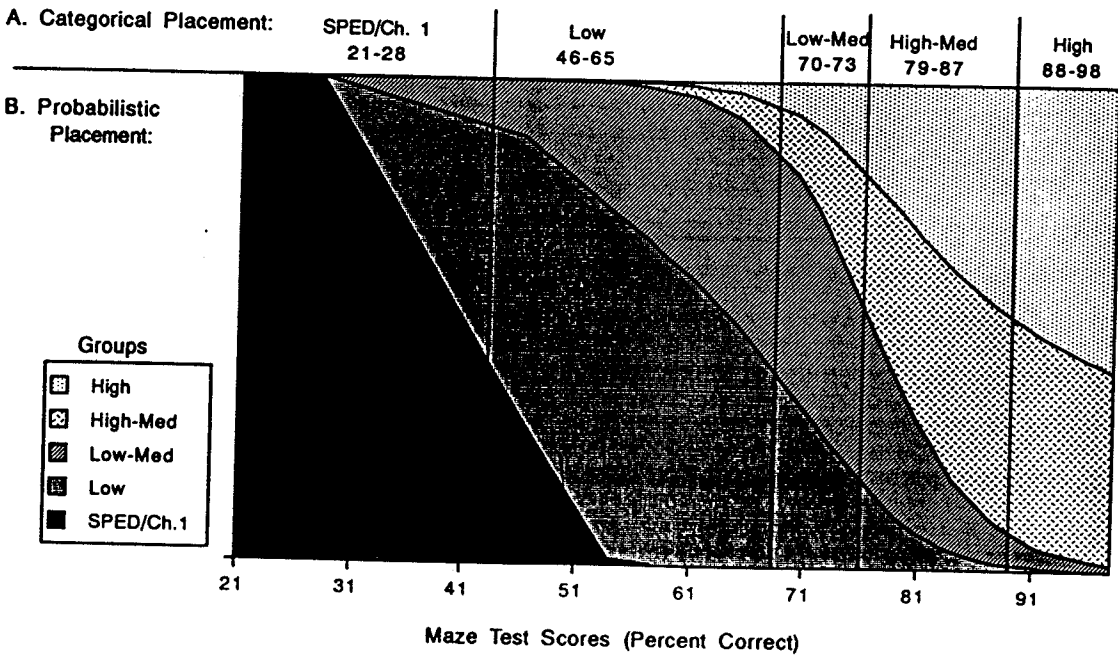
Predicting Grade 4 Language Arts Groups from A.R.I. Scores:
Categorical and Probabilistic Predictions

A. Categorical Placement:

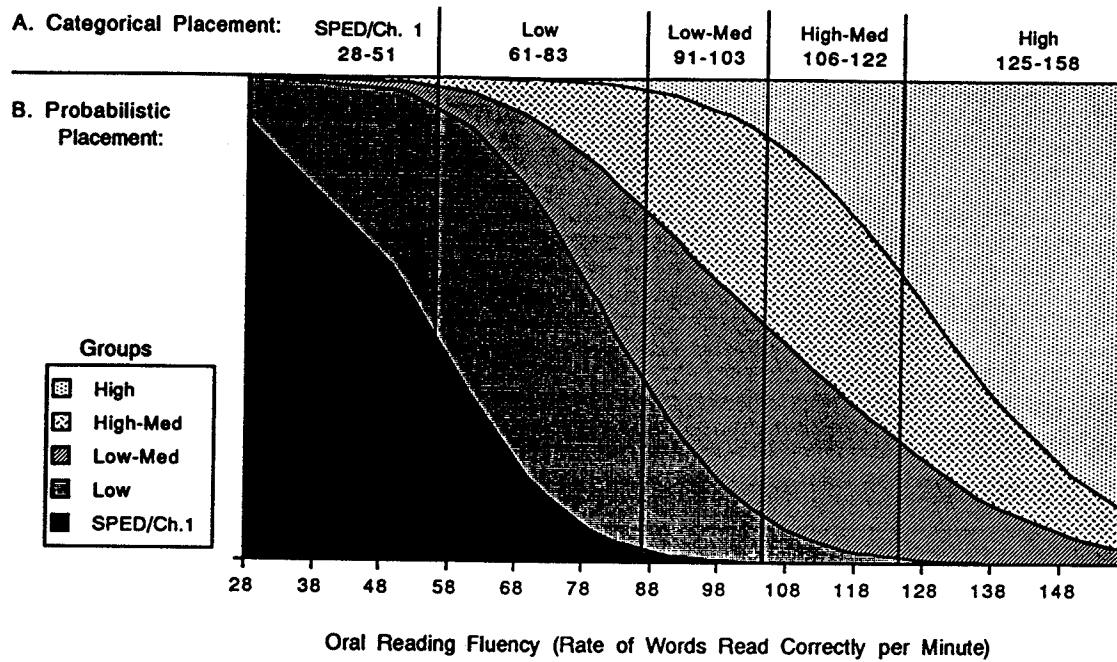


Predicting Grade 4 Language Arts Groups from Maze Test Scores:
Categorical and Probabilistic Predictions

A. Categorical Placement:

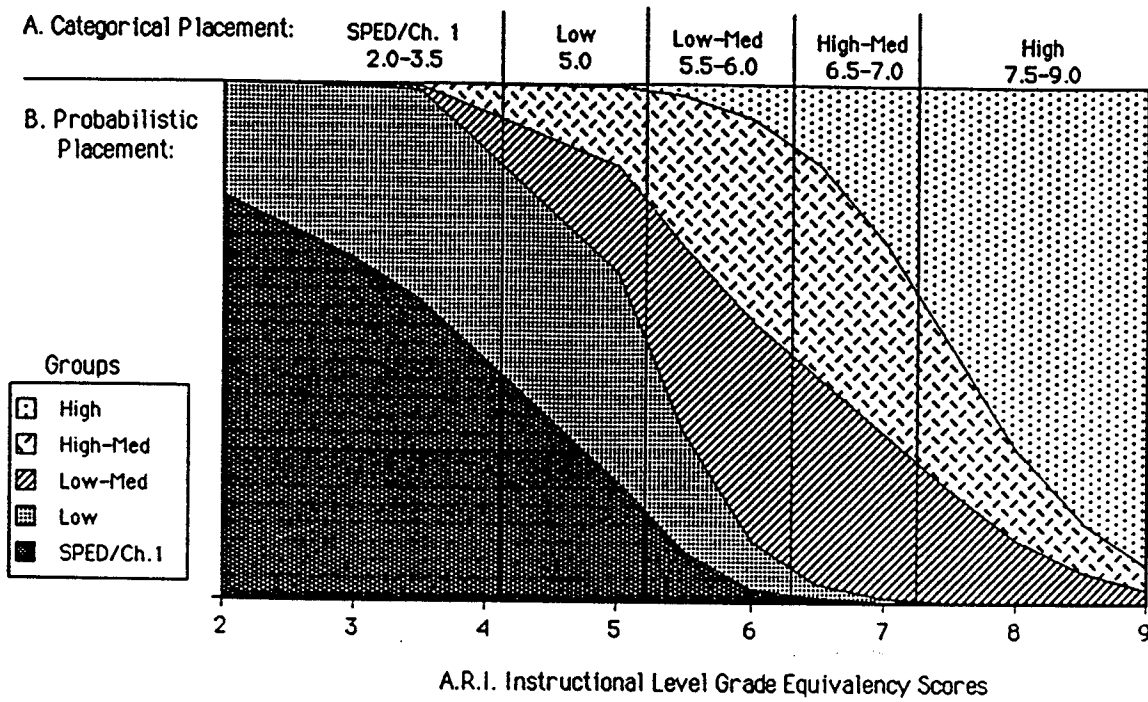


Predicting Grade 4 Language Arts Groups from Oral Reading Fluency Scores:
Categorical and Probabilistic Predictions

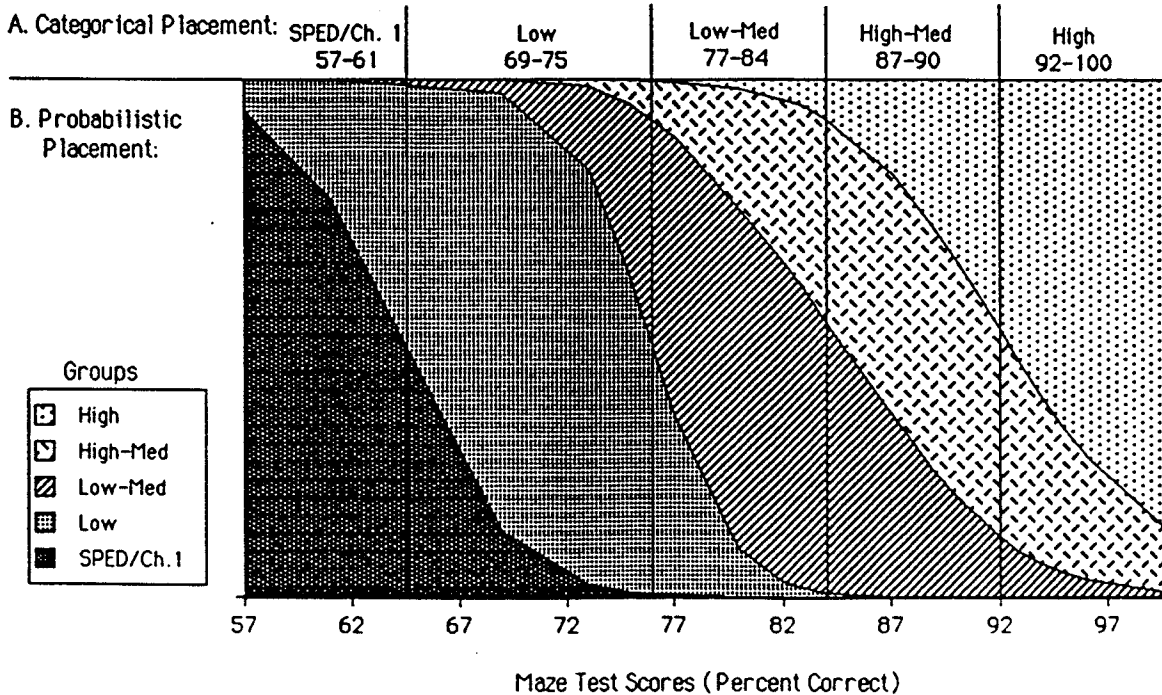


APPENDIX D
STACKED AREA GRAPHS
FOR GRADES 5 AND 6

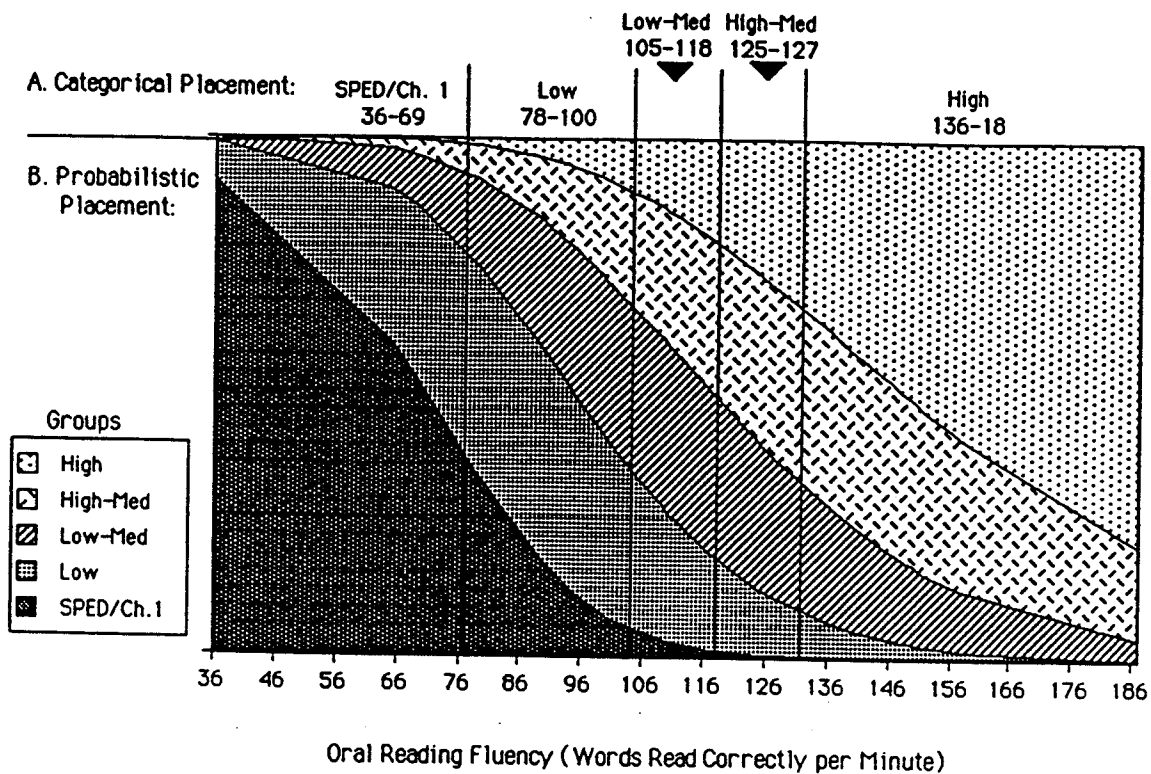
Predicting Grade 5 Language Arts Groups from Analytical Reading Inventory (A.R.I.) Scores: Categorical and Probabilistic Predictions



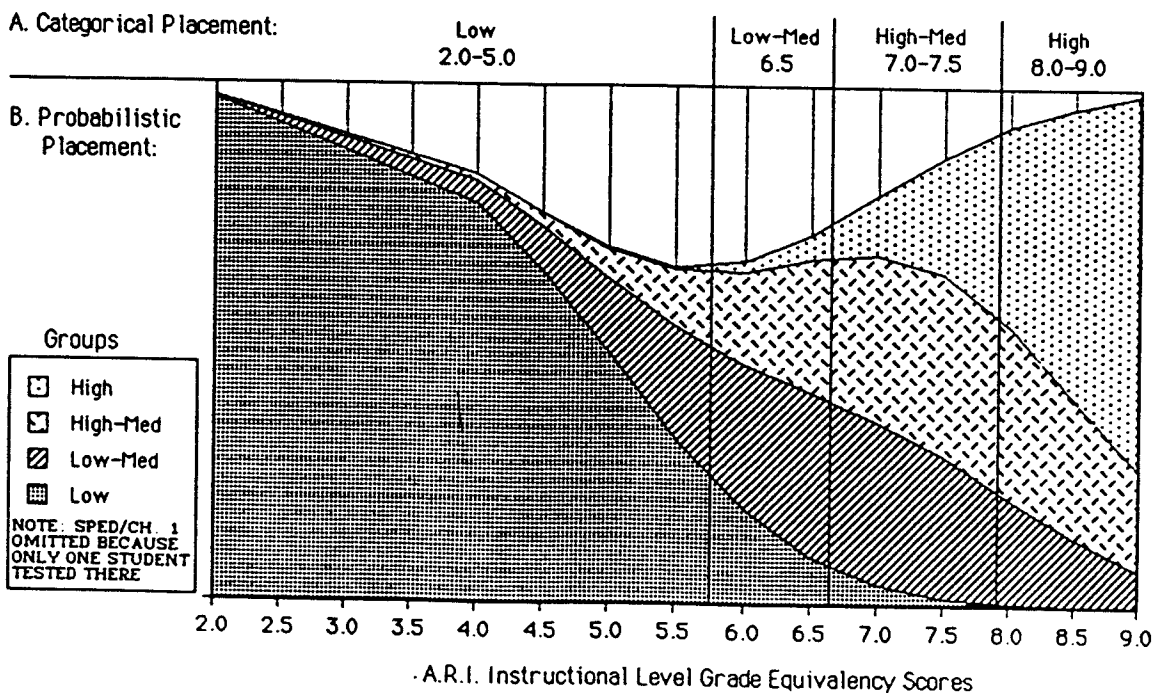
Predicting Grade 5 Language Arts Groups from Maze Test Scores: Categorical and Probabilistic Predictions



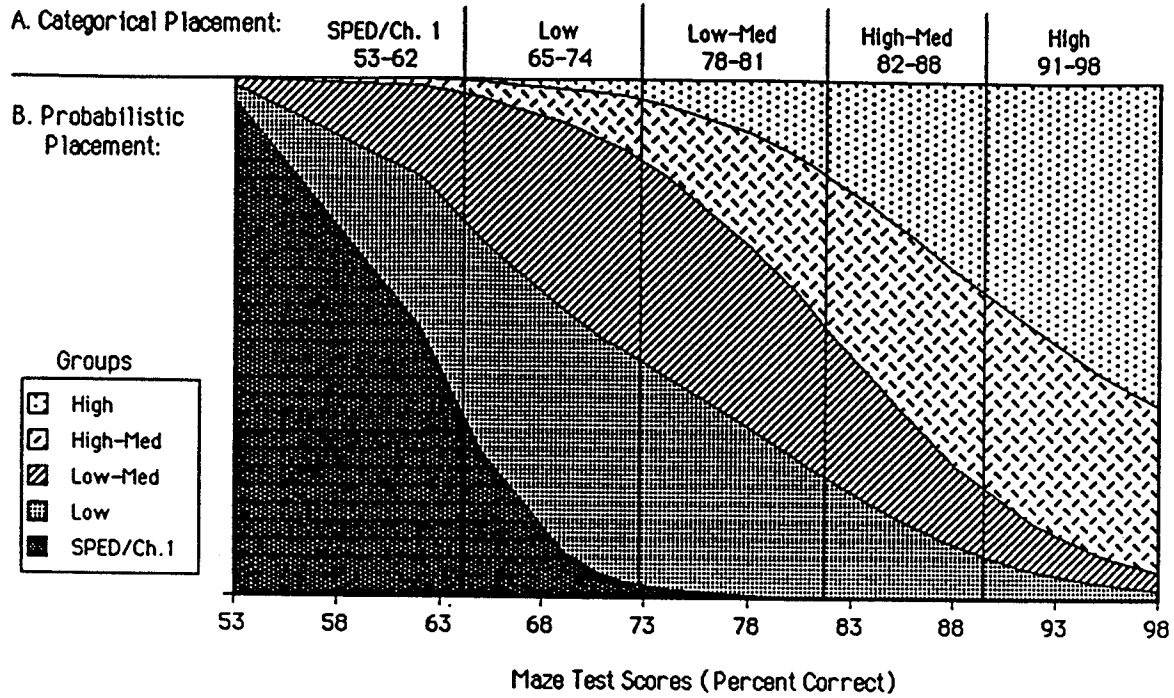
Predicting Grade 5 Language Arts Groups from Oral Reading Fluency Scores: Categorical and Probabilistic Predictions



Predicting Grade 6 Language Arts Groups from Analytical Reading Inventory (A.R.I.) Scores: Categorical and Probabilistic Predictions (SPED/Ch. 1 category omitted because only one student tested there.)



Predicting Grade 6 Language Arts Groups from Maze Test Scores: Categorical and Probabilistic Predictions



Predicting Grade 6 Language Arts Groups from Oral Reading Fluency Scores: Categorical and Probabilistic Predictions

