Resource Consultant Training Program
Research Report No. 5

RCTP

The Utility of Pflaum's Oral Reading Miscue Categories

Richard Parker
Jan Hasbrouck
Gerald Tindal

University of Oregon, Division of Teacher Education, Special Education Area,
Eugene, Oregon, 97403-1215

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

Parker, Richard
Hasbrouck, Jan
Tindal, Gerald
The Utility of Pflaum's Oral Reading Miscue Categories
Research Report No. 5

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

THE UTILITY OF PFLAUM'S ORAL READING MISCUE CATEGORIES

Richard Parker
Jan Hasbrouck
Gerald Tindal

Abstract

Miscue analysis as advocated by Goodman (Goodman, 1969) is a clinical tool that permits inferences about a student's reading processes. Miscues also can be approached from a utilitarian point of view by teachers who must make practical decisions about, for example, group placements and functional reading levels. These teachers may need to know: (a) Will two assessment specialists using miscue analysis arrive at the same conclusions about a given student's oral reading errors? (b) do miscue analysis results correspond with results from other reading comprehension measures? and (c) can miscue analysis results be used in a non-clinical manner for instructional decisions such as group placement? In this study, a derivation of Goodman and Burke's (1972) miscue analysis developed by Pflaum (1980a, 1980b) was investigated. It was applied to the audio-taped reading of three, 250-word passages by 77 students in Grades 4, 5, and 6. First, four types of reliability estimates were obtained on data from two scorers. Next, miscue types were used to predict placement in existing four reading groups at each grade level. Finally, miscue counts were correlated with three other comprehension measures. Results generally were not supportive of using miscue types for making decisions about individual students. In general, a simple count of the total number of oral reading errors was more reliable and had more predictive efficiency than any miscue type. A large variation in results from one grade level to the next indicates danger in generalizing about miscues over the elementary grades. There was some indication that the non-clinical utility of miscue analysis may be studied more fruitfully with lower achieving students.

MISCUES IN THE CLASSROOM

Informal Reading Inventories (IRIs), whether teacher-constructed, basal series-based, or commercially developed, have survived a barrage of criticism for their lack of reliability and validity (Ekwall, 1976; Harris & Niles, 1982; Marr & Lyon, 1980; Pikulski & Shanahan, 1982; Powell, 1970; Schwartz, 1984) and remain the most popular reading assessment tools in the classroom (Harris & Lalik, 1987). For many teachers, criticisms of low test reliability and validity are largely irrelevant: They regard an IRI not as a test but as a clinical tool for studying a student's reading processes (Powell, 1970). They contend that "an informal inventory is no better than the person using it" (Pikulski, 1974, pp. 146-147). Others take a utilitarian view of IRIs, using them to make practical classroom decisions such as locating instructional reading levels and placing students in reading groups—with a reasonable degree of consistency and precision (Harris & Lalik, 1987).

For those with the "clinical investigation" point of view, Goodman's (1969) system of miscue analysis provides a useful framework for exploring the reading process. Although commercial IRIs vary in how oral reading errors are counted and classified (Ekwall, 1976; Schwartz, 1984), newer versions commonly recommend some variation of Goodman's miscue analysis (Harris & Niles, 1982). "Simple quantitative assessment . . . should be discouraged . . . in favor of a combined quantitative/qualitative assessment that helps the teacher examine patterns of miscues that reveal the strategies used by the child in finding meaning" (Harris & Niles, 1982, p. 160).

Miscue analysis follows traditional error identification (mispronunciations, substitutions, omissions, insertions, hesitations) and is used to infer how and to what extent the reader gains meaning from context clues. The full set of procedures, represented by the Reading Miscue Inventory (RMI, Goodman & Burke, 1972), is complex, requiring five comparisons between each error and its correct alternative: dialect, intonation, visual configuration, phonetic configuration, and grammatical function. The examiner also must determine whether each error (a) is corrected by the reader, (b) is grammatically correct, (c) is semantically correct, and (d) results in a change of meaning. Interrelationships among various error patterns then are examined to gain insight into the reader's comprehension strategies.

Miscue analysis has much to commend it. First, while traditional error categories are defined by within-word referents only, miscue analysis errors are defined more broadly, in terms of relationships at the sentence and passage levels, reflecting more recent psycholinguistic conceptions of reading comprehension. Second, oral reading errors are viewed as varying in their severity or importance to the overall understanding of the passage. Third, miscue analysis allows students to be classified qualitatively, according to the strategies they most commonly use to get meaning from text. Identification of learner strategies can be potentially useful in instructional planning. Fourth, a considerable body of literature has been generated on miscue analysis. The main research base is composed of 18 doctoral dissertations completed between 1970 and 1973, 13 of them at Wayne State University, where Kenneth Goodman directed a U. S. Office of Education project on reading miscue research. The dissertations are in-depth, qualitative, clinical investigations, conducted with reading samples from relatively few students; Ns of 3 to 15 are cited (Allen, 1976). A bibliography of miscue research publications is included in Appendix A.

Miscue analysis also has been criticized, however, for being too complex and time-consuming (Hood, 1978; McGinnis & Smith, 1982; Pflaum, 1980b; Pikulski, 1974), and for using overlapping error categories (Pikulski, 1974) that result in unreliable error classification (Page, 1976; Pflaum, 1980a, 1980b). Simpler versions of Goodman's original procedures have been developed (Pflaum, 1980a; Weaver, 1980), utilizing less complex derivations of the full RMIC (McGinnis & Smith, 1982; Smith & Weaver, 1978). Furthermore, reliability seldom is a concern for practitioners tending toward qualitative, clinical exploration, such as the original Wayne State researchers.

Reliability is of concern mainly to those practitioners with utilitarian needs. Since they use these measures for making practical decisions, precision in measurement and consistent results are important. A representative of this latter group is Pflaum (1980a, 1980b), who developed alternative miscue analysis procedures that are purported to be less time consuming, more reliable, and of greater utility than Goodman and Burke's (1973). Pflaum's simplified miscue analysis procedures require only three to five decisions for each oral reading error: whether the error (a) is self-corrected, (b) changes the meaning of the sentence, (c) renders the sentence incomprehensible, and (d) involves the use of phonic cues. Finally, degree of phonic use is rated as high, partial, or slight (pp. 280-281). This coding system "was tested for coder reliability with four experimenters whose average levels of agreement ranged from 83 to 100%" (p. 280), which is substantially higher than that reported by Page (1976) on simpler oral reading error ratings. Yet the specifics of Pflaum's (1980a) scorer training and reliability estimation are sketchy and ill-defined.

The purpose of this study was to investigate Pflaum's (1980a, 1980b) miscue analysis procedures by applying them to three, 250-word passages read by 77 students in Grades 4, 5, and 6. The study addresses the following three utilitarian concerns:

1. Can Pflaum's miscue categories be reliably identified and coded by different raters after a reasonable training period?

2. Can the number and type of identified miscues differentiate existing within-grade reading group placements?

3. Do the number and types of identified miscues correlate significantly with other estimates of

reading comprehension (teacher ratings, Analytical Reading Inventory [Woods & Moe, 1985], or the Maze procedure [Guthrie et al., 1974]) ?

METHOD

Students

This study was conducted with 120 students in a 280-pupil elementary school located in a middle-SES, rural West Coast community. The district generally performs around the 60th to 70th percentiles on national achievement tests. Very few minority students attended the school; none were sampled in the study. Students were sampled randomly from fourth-, fifth-, and sixth-grade reading groups that had been formed 6 months before the study (in September) on the basis of cumulative folder reports, recommendations of the previous year's teachers, informal classroom observations, and IRIs. Four reading groups (High, High-Medium, Low-Medium, and Low) had been established across classrooms (four classrooms per grade level) at each of the three grade levels. The Low group was composed of students who were withdrawn from the classroom for either Chapter I or Special Education/Resource Room instruction. The other three groups all represent single instructional units. Through stratified random selection, 10 students were chosen from each of the four reading groups, totaling 40 subjects at each of the three grade levels.

Instrumentation

Eight trained examiners administered three reading measures: (a) a modified version of the Analytical Reading Inventory (ARI, Woods & Moe, 1985), (b) three Maze tests (Guthrie et al., 1974) derived from students' basal readers, and (c) timed, audio-taped Oral Reading of the same three basal passages.

The Analytical Reading Inventory

The ARI manual gave inadequate guidance for consistent scoring among the eight members of the research team: "Some of the decision-making is based upon subjective evaluations," for example, "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). Therefore, a modification of the ARI was administered individually, yielding an instructional reading level (grade-equivalency score) for each student.

Four modifications were made to the ARI: (a) "Miscue analysis" and "story retell" were deleted; (b) objective decision rules were established for determining how many test passages to administer, and in what order; (c) objective procedures were established for computing instructional levels from reading rate and comprehension scores only; and (d) ARI entry level was determined by current reading group placement, rather than from performance on the word list reading task. Reasonable average interrater reliability ($\underline{r}$ = .90) was obtained across the eight examiners after these modifications were made.

Basal Passages: Maze and Timed Oral Reading

Maze and timed Oral Reading measures sampled basal passages that had been selected from Grade 4, 5, and 6 reading texts in classroom use: *Running Free, Three Cheers,* and *Distant Views* (Scott, Foresman Reading Series, 1987). Readability scores were calculated for each passage, based on the average of Spache, Fry, and RAYGOR formulas. Each student read three passages, with readabilities calculated at grade level, 1 year above grade placement, and 1 year below grade placement.

Group-administered Maze (i.e., multiple-choice Cloze) (Guthrie et al., 1974) tests were produced for all selected passages, using the following procedures. Only the first and last sentences of a 250-word passage were left intact; otherwise, every sixth word was deleted, and students were offered five alternatives in a "keyword search" (Roid & Haladyna, 1982). The four distractors were random selections from the pool of deleted words. Students completed three Maze tests, each containing 34 items. For each student, "percent correct" scores for the three tests were averaged for analyses.

Audio-recordings were obtained on individually timed Oral Readings of the same three intact basal passages. Four types of errors were counted: mispronunciations/substitutions, additions, omissions, and hesitations of at least 4 seconds, after which students were told the word. The audio-tapes later were analyzed using Pflaum's miscue categories.

Teacher Ratings

Teachers were asked to estimate each student's ability to get meaning from text. Ratings were made on a grade equivalency scale from 1.0 to 12.0, in 6-month increments.

Procedures

In March, 12 teachers administered three Maze tests to their reading groups within a 20-minute

session, using scripted directions. The order of presentation of the three tests was counterbalanced within each classroom group. Over the next 2 weeks, students were withdrawn individually from class to read the same three passages aloud 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 orally read. The Maze/Oral Reading order could not be counterbalanced because a prior Oral Reading would reveal answers to the Maze.

ARI and Oral Reading assessments required approximately 27 minutes per student (15- to 50-minute range). The ARI alone required approximately 18 minutes for most students (8- to 35-minute range), while the Oral Reading averaged only 9 minutes (6- to 15-minute range).

The final sample of 77 students was composed of 26 in Grade 4, 22 in Grade 5, and 29 in Grade 6. Deleting unclear audio-tapes reduced the number of usable recordings from 309 (three passages read by 103 students) to only 231 (three passages, per 77 students), a 25% data loss. Because individual testing sessions could be rescheduled for absences, Oral Reading and ARI results were available for 117 of the 120 students. Group tests could not be rescheduled, so Maze results were available for only 103 of the original 120 students. All analyses are based on the reduced sample of 77 students for whom all data were available.

Miscue Analysis

Examiners were trained to analyze the 231 taped Oral Readings using some of Pflaum's (1980a) miscue categories: (a) first word and heading errors, (b) uncorrected errors with no meaning change, (c) uncorrected errors with meaning change (severe/not severe), (d) corrected errors with no meaning change, (e) corrected errors with meaning change (severe/not severe), and (f) words pronounced by the examiner (pp. 280-282). Pflaum's final category—use of phonic cues (high, partial, or little)—was eliminated because of time constraints.

An initial 3-hour training session was conducted with six taped passages that had been deleted from the study because of incomplete data. Two graduate-level education students listened to the tapes, independently marked blank protocols, compared their scores to the "expert scoring" of the head trainer (the second author), and discussed scoring discrepancies. At the close of the training session, scorers were given a tape recorder and duplicate tapes of 18 additional recordings to be scored and returned within a week. The coding summary sheet is included in Appendix B.

Disappointingly low agreement between the two scorers occurred with the 18 passages. A second, 3-hour training session was scheduled to discuss the largest score discrepancies. "Expert ratings" again were presented by the trainer, scoring guidelines were clarified, and a few additional recorded readings were scored, with more satisfactory results. Finally, the two scorers were given duplicate tapes for the 231 clear recordings produced by 77 students.

Modifications of Miscue Categories

All analyses were based upon the following six miscue categories: (a) Total Number of Oral Reading Errors (Total); (b) Uncorrected Miscues (Uncorr); (c) Self-Corrected Miscues (Corr); (d) Errors Causing a Significant Sentence Meaning Change (M.Chg); (e) Errors Causing No Significant Sentence Meaning Change (No M.Chg); and (f) Meaning Change Errors that are so Severe as to Prevent Grammatical Completion of the Sentence (Severe). Two of Pflaum's main categories—"first word and heading miscues" and "examiner pronounced miscues"—were eliminated from analysis because the errors occurred in only a small percentage of recordings (5% and 8%, respectively), depressing reliability and reducing power for statistical tests. Higher error rates for these categories probably would have been found for younger students. Other miscues with insufficient data points were the "severe/not severe" sub-categories under both "uncorrected meaning-change errors" and "corrected meaning-change errors." To decrease the null responses for statistical analyses, the two "severe" sub-categories were combined.

RESULTS

Descriptive Data

Average counts for all oral reading errors (Total) and for the six miscue categories are presented in Table 1 by reading group within grade level. The overall error rate was 4% to 5% for all three grades. Differences among the three grade levels in miscue frequencies were tested through ANOVA, resulting in no significant F ratios. At all grade levels, most errors involved a change in sentence meaning and were uncorrected (Uncorr). Self-corrections (Corr) had generally low frequencies and were the least frequent miscue type in Grades 5 and 6. Most meaning-change miscues (70-80%) were perceived as "severe."

In a secondary analysis, error frequencies were totaled over three passages and plotted across the three grades as shown in Appendix C.

Table 1: Mean Counts of Total Oral Reading Errors and Five Miscue Types Produced by Grade 4, 5, and 6 Students Within 250-Word Basal Reader Selections

	Grade 4 (n=26)		Grade 5 (n=22)		Grade 6 (n=29)			
	M	SD	M	SD	M	SD	$F_{(2,74)}$	p
Total[1]	12.7	7.8	10.4	5.2	9.7	6.2	1.59	.21
Uncorr[2]	8.6	5.5	6.7	3.5	7.2	5.7	.96	.38
Corr[3]	4.4	4.4	3.7	3.0	2.5	1.5	2.46	.09
M.Chg[4]	7.2	5.4	5.5	3.1	5.0	3.9	1.90	.15
No M.Chg[5]	3.9	2.1	4.1	2.2	4.5	2.9	.35	.70
Severe[6]	6.0	5.6	4.2	3.5	3.5	3.0	2.61	.08

[1] Total number of oral reading errors.
[2] Errors not self-corrected by student.
[3] Errors self-corrected by student.
[4] Errors causing a significant change in sentence meaning.
[5] Errors causing no significant change in sentence meaning.
[6] Errors implying severe or extreme changes in meaning.

Reliability of Miscue Scoring

Following training, two of the graduate student examiners independently scored a random sample of 63 audio-taped readings. Reliabilities were calculated at the passage level from counts of each of the six error types. Four methods were used to estimate interscorer reliabilities for these raw score counts as shown in Table 2: Pearson r, Standardized Score Disagreement Index (SSDI), Cohen's Kappa, and "Hit" rates.

Pearson correlations were all significant at $p<.001$ and were all moderate to strong (range .69-.93). These reliability estimates were used to calculate the standard error of measurement (SEMeas) for each miscue type and confidence intervals for making judgments about individual scores. The ±1 SD (69%) confidence intervals for each miscue type were: Total, ±1.4; Uncorr, ±1.3; Corr, ±1.3; M.Chg, ±1.2; No M.Chg, ±1.2; and Severe, ±1.2. Translating from raw scores to percentiles, we can predict with 68% certainty (±1 SD) that an obtained 50th percentile score represents a true score within these intervals: Total, 40th-70th; Uncorr, 40th-70th; Corr, 1st-70th; M.Chg, 30th-70th; No M.Chg, 20th-80th; and Severe, 1st-70th.

The second agreement estimate, the Standard Score Disagreement Index (SSDI), reflects how discrepant two ratings are for each miscue type, averaged over all samples. First, standard (Z) scores were created from the combined raw miscue counts. Next, each pair of standard scores for a given miscue type was differenced (yielding interrater discrepancies), and the absolute differences were summed and averaged, resulting in a SSDI. The SSDI can be interpreted as the average amount of disagreement between scorers in standard deviation units. The SSDIs presented in Table 2 range from about one fourth of a standard deviation unit (.26) for Total, to over one half of a standard deviation unit (.54) for Corr. For scores near the mean, this translates to discrepancies from about 10 to 20 percentile points. The SSDI has four desirable qualities: (a) ease of computation from standard scores, (b) comparability across different miscue types and scales, (c) consideration of *degree* of discrepancy between scorers versus only hit/miss calculations, and (d) interpretability in common standard deviation units, with a functional upper limit of 3.

Because correlations only loosely reflect classification agreement (Fleiss, 1981), two classification indices also were calculated—Cohen's Kappa and Hit rates. The continuous data scales were first transformed to categorical scales, using percentile ranks, resulting in five, 20-percentile categories. Percentiles have the desirable characteristic of equally distributing the raw scores across categories.

Cohen's Kappa, a conservative index that reflects only agreement beyond chance (Fleiss, 1981), then was applied to the categorized data. Kappa coefficients ranged from .32 to .64 (all significant at $p<.001$), as shown in Table 2. The obtained Kappas of .44 and .49 for Total and Uncorr indicate "good" agreement, while the others indicate "poor" agreement beyond chance (Fleiss, 1981).

Table 2: Four Interscorer Reliability Estimates for Total Oral Reading Errors and Five Miscue Types, Based on Two Independent Scorings of 63 Audio-Tapes

	Pearson r [1]	SSDI[2]	Cohen's Kappa[3]	Hit Rates [4]
Total	.93	.26	.60	.68
Uncorr	.92	.28	.64	.71
Corr	.69	.54	.33	.50
M.Chg	.86	.40	.35	.50
No M.Chg	.82	.43	.39	.52
Severe	.75	.48	.32	.49

[1] All significant at p<.001

[2] Standard Score Disagreement Index: average disagreement or discrepancy between scorers, in standard deviation units.

[3] Cohen's Kappa: an index for agreement beyond chance. 0=Chance, <0=less than chance agreement, 1= omplete agreement. Values of .40 -.75 represent "good" agreement beyond chance (Fleiss, 1981, pp. 212-236). A synthetic 7-point classification scale was derived from the original continuous score scale. All tabled values are significant at p<.001.

[4] Hit Rates: Proportion of taped samples for which both scorers gave same score (using a 5-point, percentile category scale).

The final agreement index, included for its simplicity, was a simple proportion of Hits (a pair of scores from two raters in the same category on the 5-point percentile scale). The number of Hits was divided by the number of opportunities (63) to calculate Hit rates. Hit rates varied from 49% to 71% (see Table 2), with highest agreement for Total and Uncorr and the lowest for Severe, Corr, and M.Chg. Hit rates are not corrected for chance agreement; a 20% (1/5) Hit rate would be obtained by chance alone.

Predicting Reading Groups From Miscues

The second major question of this study, whether miscues differentiate existing reading group placements, was addressed by conducting separate univariate ANOVAs among the four reading groups for each grade. For greater stability of results, analyses were conducted on average scores, based on all three taped readings. The mean miscue scores presented in Table 3 appear to generally decrease from low to high reading groups, but only half of the F ratios for between-group differences were significant. The significant grade 4 F ratios cannot be interpreted as linear trends, as there was an unfortunate similarity between the medium-high and medium-low Grade 4 groups. For Grades 5 and 6, only M.Chg and Severe miscues consistently predicted group placement. The generalized correlation coefficient, "eta" (Hays, 1981, p. 349), for these two miscue types was of moderate size (range η=.58-.73).

To better assess the predictive accuracy of miscues in making reading group placements, the reading group variable was "effect-coded" (Pedhazur, 1982, pp. 289-296) and regressed on the strongest predictor, M.Chg miscues, at both Grade 5 and 6 levels. "Fitted" scores for each student were then examined and tallied as closer to actual group placement (a "hit") or to some other group (a "miss"). At the Grade 5 level, with η=.73, the hit rate for M.Chg miscues was 8/22, or 36%. At Grade 6, the Hit rate (with η=.66) was 12/29, or 41%. Hit rates of other miscue types were lower.

Miscue Correlations with Other Criterion Measures

The third question addressed in this study was whether miscue scores related to three other reading estimates: teacher ratings of students' reading comprehension ability, the Analytical Reading Inventory, or the Maze test. Pearson correlations between miscue raw scores and these three criteria are presented by grade level in Table 4.

Correlations between the six miscue types and four criterion measures are presented in bar graph form by grade level (Appendices D, E, and F), and across the three grades (Appendix G).

The large number of correlations (54) cautions interpretation of general patterns of significant (p<.01) correlations only. Primarily low- to high-

Table 3: Mean Miscue Counts Displayed by Four Reading Groups for Grades 4, 5, and 6, and ANOVA Results Between Groups[1]

		Miscue Types					
	N	Total	Uncorr	Corr	M.Chg	No M.Chg	Severe
Grade 4 Groups							
Low	8	61.0	37.5	25.6	35.5	12.8	34.6
Med-L	6	29.3	20.6	8.6	14.1	13.1	8.6
Med-H	6	31.5	22.6	8.8	18.6	11.0	15.0
High	6	23.5	18.5	5.0	13.5	9.6	8.8
$F_{(3,22)}$		6.0**	2.2	5.6**	3.9	0.4	6.5**
Grade 5 Groups							
Low	6	42.3	24.5	17.8	22.8	11.6	22.6
Med-L	5	35.2	25.1	10.0	22.6	11.8	12.8
Med-H	6	28.6	18.3	10.3	14.0	14.3	9.6
High	5	17.4	11.8	5.6	6.4	10.6	3.2
$F_{(3,18)}$ (Eta)	3.2	2.0	2.1	6.7(.73)**	.3	5.3(.69)*	
Grade 6 Groups							
Low	7	47.8	36.7	11.1	27.8	19.1	19.5
Med-Low	9	29.1	22.3	6.7	14.0	14.4	9.1
Med-High	7	19.5	11.4	8.1	9.5	9.2	7.5
High	6	18.1	14.1	3.8	8.0	9.8	6.1
$F_{(3,25)}$ (Eta)	5.3(.62)*	4.1(.58)*	3.8	6.4(.66)*	2.0	4.2(.58)*	

* < .01
** < .005
[1]Eta, the generalized correlation coefficient (MSbt/MStot), is presented only for those significant F ratios with mean scores properly ordered according to group level.

moderate correlations are noted, varying systematically across miscue types and across grade levels. At Grade 4, the miscue types most strongly related to the criterion measures are Corr, Total, and Severe, in that order. At Grade 5, no miscue type bore uniformly strong relationships with all three criterion measures, although Severe and M. Chg came closest. At Grade 6, Total, Uncorr, and M.Chg were most strongly related to the three criteria. Very low correlation coefficients (r=.27 and below) were produced only by No M.Chg, at the Grade 4 and 5 levels.

DISCUSSION

This study addressed three utilitarian concerns in using Oral Reading miscue analysis: interscorer reliability, reading group placements, and the relationship between miscues and other comprehension measures. From Grades 4, 5, and 6, 77 students read three, 250-word passages from the basal, completed a modified ARI (Woods & Moe, 1985), and took three Maze tests (Guthrie et al., 1974). Miscue analyses, using Pflaum's (1980a) taxonomy, were performed from audio-tapes of the Oral Reading. Scores were obtained on the *total* number of oral reading errors and five miscue categories: (a) those *corrected* by the student, (b) *uncorrected* miscues, (c) those causing significant *meaning change* in a sentence, (d) those causing *no meaning change* in sentence, and (e) those causing *severe* or extreme changes in sentence meaning.

Interscorer reliability of the miscues was quite low, in contrast to Pflaum's (1980a) report that "average levels of agreement ranged from 83 to 100%" (p. 280). Using Cohen's Kappa and SEMeas only, *total* errors and *uncorrected* miscues demon-

Table 4: Correlations Among Six Miscue Types and Four Criterion Measures, Presented for Grades 4, 5, and 6

		Criterion Measures		
Grade	Miscue Types	T. Rate	ARI-c	Maze
4 (N=26)	Total	.61*	.61*	.67*
	Uncorr	.37	.49*	.41
	Corr	.80*	.58*	.78*
	M. Chg	.26	.47	.47
	No M. Chg	.27	.20	.23
	Severe	.62*	.61*	.59*
5 (N=22)	Total	.54*	.67*	.46
	Uncorr	.42	.45	.20
	Corr	.44	.63*	.56*
	M. Chg	.62*	.65*	.49
	No M. Chg	.08	.13	.18
	Severe	.50	.74*	.66*
6 (N=29)	Total	.61*	.64*	.74*
	Uncorr	.57*	.60*	.72*
	Corr	.38	.41	.38
	M. Chg	.55*	.63*	.73*
	No M. Chg	.55*	.48	.60*
	Severe	.50*	.50*	.59*

* $p<.01$

Note: All correlations are actually negative.

strated reasonable reliability. If individual student-level decisions were based on miscue scores, a large error band of more than 30 to 60 percentile points would be necessary—too broad for most practical uses. Of the six miscue types, the most reliable were *total* number of errors and *uncorrected* miscues, while the least reliable were *corrected* and *severe meaning change* miscues, regardless of which reliability index was computed.

The problem of unreliability is likely because of the small number of miscue counts per typical taped reading relative to the large variance among students' scores. The students sampled for this study spanned the full range of achievement, but, as reported earlier, the average district achievement level was high. Improved reliability would likely be noted among more deficient readers with higher miscue rates. These low reliability levels resulted despite 6 hours of training by the second author, an experienced teacher-trainer and reading teacher of 14 years. Although scorers had the advantage of using only audio-tapes, which had been first screened for clarity, they reported that most tapes needed to be stopped periodically during scoring and many replayed to ensure accuracy; Pflaum (1980a) concurs that audio-taping for later analysis is essential. From a utilitarian view, then, miscue analysis may be too cumbersome to implement in the classroom. Given the low reliability under optimal training and scoring conditions, we could expect it to be even lower in the classroom.

This study also addressed the utility of miscues for making reading group placements. No significant between-grade differences were obtained for any miscue type. Although certain miscue types reached significance in predicting within-grade reading group placements, no single miscue type was effective at all three grades. A grouping anomaly (the Medium-Low and Medium-High groups were reversed in achievement level) was noted in the Grade 4 data, which made those data difficult to interpretate. The strongest reading group predictors were *meaning change* miscues at Grades 5 and 6, and both *total* errors and *corrected* miscues at the Grade 4 level. The corresponding generalized

correlation coefficients (η) were of moderate size (range η=.58-.73), but when translated to accuracy in group placement of individual students, the best Hit rates were 36% and 41% (for *meaning change* miscues). Therefore, only certain miscue types appeared effective for each grade level, and considerable error would be involved in their use for placing individual students in reading groups.

The third major question posed in this study was how miscues related to other measures of reading comprehension, namely teacher ratings, a modified ARI, and the Maze. There was considerable consistency among the three criterion measures in miscue correlations. As in predicting reading group placements, however, the miscue types with most predictive efficiency (r=.60-.80 range) varied from one grade to another. The only miscue type that produced at least moderate-size correlations (r=.54-.74 for 8 of 9 coefficients) across the three grade levels was *total* number of errors. The best grade-specific matches with these three comprehension measures were *corrected* miscues at Grade 4, *severe* miscues at Grade 5, and *total* number of errors at Grades 4 and 6. The fact that there were no correlations above .80 argues against the use of any miscue type as a replacement of any of the three existing criterion measures.

This study had several limitations. First, some of Pflaum's (1980a) original categories were deleted because of low incidence rates and training time constraints, thus limiting the conclusions. Second, this study dealt only with Grades 4 to 6; such large variation in results were obtained from one grade to the next that conclusions should not be overgeneralized even to all elementary grades. Third, the anomalous grouping situation noted in Grade 4 reduces the usefulness of some of the group-prediction data. Fourth, miscues were not clinically interpreted, contrary to Goodman's original focus and intent; this study is mute on the subject of miscues' usefulness in investigating the *process* of gaining meaning from text. Finally, Oral Reading fluency was excluded from the study to maintain the focus on miscues. Fluency scores, however, were obtained from these data and showed stronger reliability and uniformly stronger prediction of reading groups and of the criterion measures than did any of the miscue types. Results of these supplementary analyses are presented in bar graph form in Appendix H.

In summary, no miscue type matched the reliability or the power in predicting reading group placements, teacher ratings, ARI scores, or Maze tests as did the simple measure, *total* number of oral reading errors. Lack of interscorer reliability, the relatively small number of Oral Reading miscues demonstrated by most students, and the large variability of miscue counts among students all reduced the utility of the theoretically well-grounded miscue categories. In addition, the modest practical utility of any miscue type appears to vary greatly from one grade level to the next. These writers hypothesize that certain miscue categories, particularly severe meaning change miscues, might be more profitably studied in Grades 1 to 4.

REFERENCES

Allen, P. D. (1976). The miscue research studies. In P. D. Allen (Ed.), *Findings in research in miscue analysis: Classroom implications* (pp. 3-10). Urbana, IL: ERIC Clearinghouse on Reading and Communication Skills.

Ekwall, E. E. (1976). *Diagnosis and remediation of the disabled reader*. Boston, MA: Allyn & Bacon.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York: John Wiley & Sons.

Goodman, K. S. (1969). Analysis of oral reading miscues: Applied psycholinguistics. *Reading Research Quarterly, 5*, 9-30.

Goodman, Y. L., & Burke, C. L. (1972). *Reading Miscue Inventory*. New York: Macmillan.

Goodman, Y. L., & Burke, C. L. (1973). *Theoretically based studies of pattrens of miscues in oral reading performance*. U.S. Department of Health, Education and Welfare, Office of Education.

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.

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.

Hays, W. L. (1981). *Statistics*. New York: CBS College Publishing.

Hood, J. (1978). Is miscue analysis practical for teachers? *Reading Teacher, 32*(3), 260-266.

Marr, M. B., & Lyon, K. R. (1980). Passage independency: 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.

McGinnis, D. J., & Smith, D. E. (1982). *Analyzing and treating reading problems.* New York: Macmillan.

Page, W. D. (1976). The three cue systems: #3—semantic. In P. D. Allen (Ed.), *Findings in research in miscue analysis: Classroom implications* (pp. 88-93). Urbana, IL: ERIC Clearinghouse on Reading and Communicaton Skills.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research* (2nd ed.). New York: Holt, Rinehart & Winston.

Pflaum, S. W. (1980a). Diagnosis of oral reading. *The Reading Teacher, 33*(3), 278-284.

Pflaum, S. W. (1980b). The predictability of oral reading behaviors on comprehension in learning disabled and normal readers. *Journal of Reading Behavior, 12*(3), 231-236.

Pikulski, J. (1974). A critical review: Informal reading inventories. *Reading Teacher, 28*(2), 141-152.

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.

Powell, W.R. (1970). 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.

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.

Schwartz, S. (1984). *Measuring reading competence: A theoretical-perspective approach*. New York: Plenum Press.

Scott, Foresman Reading Series (1987). *An American tradition*. Chicago, IL: Scott, Foresman.

Smith, L., & Weaver, C. (1978). A psycholinguistic look at the informal reading inventory, Part 1: Looking at the quality of readers' miscues: A rationale and an easy method. *Reading Horizons, 19*, 12-22.

Weaver, C. (1980). *Psycholinguistics and reading: From process to practice* (Chapter 7). Cambridge, MA: Winthrop.

Woods, M. L., & Moe, A. J. (1985). *Analytical Reading Inventory* (3rd ed.). Columbus, OH: Charles E. Merrill.

APPENDIX A
Miscue Bibliography

Allen, P. D. (1976). The miscue research studies. In P. D. Allen (Ed.), *Findings in research in miscue analysis: Classroom implications* (pp. 3-10). Urbana, IL: ERIC Clearinghouse on Reading and Communication Skills.

Allen, P. D. (1976). Implications for reading instruction. In P. D. Allen (Ed.), *Findings in research in miscue analysis: Classroom implications* (pp. 107-112). Urbana, IL: ERIC Clearinghouse on Reading and Communication Skills.

Biemiller, A. (1970). The development of the use of graphic and contextual information as children learn to read. *Reading Research Quarterly, 6*, 75-96.

Burke, C. L. (1976). Reading miscue research: A theoretical position. In P. D. Allen (Ed.), *Findings in research in miscue analysis: Classroom implications* (pp. 10-23). Urbana, IL: ERIC Clearinghouse on Reading and Communication Skills.

Burke, C. L. (1976). The three cue systems: #2— syntactic. In P. D. Allen (Ed.), *Findings in research in miscue analysis: Classroom implications* (pp. 79-88). Urbana, IL: ERIC Clearinghouse on Reading and Communication Skills.

Goodman, K. S. (1965). A linguisitic study of cues and miscues in reading. *Elementary English, 42*(6), 639-643.

Goodman, K. S. (1976). What we know about reading. In P. D. Allen (Ed.), *Findings in research in miscue analysis: Classroom implications* (pp. 57-70). Urbana, IL: ERIC Clearinghouse on Reading and Communication Skills.

Goodman, K. S. (1980). Revisiting research revisited. *Reading Psychology, 1*(3), 192-194.

Goodman, Y. L., & Burke, C. L. (1972). *Reading Miscue Inventory*. New York: Macmillan.

Goodman, Y. L., & Burke, C. L. (1973). *Theoretically based studies of pattrens of miscues in oral reading performance.* U.S. Department of Health, Education & Welfare, Office of Education.

Guthrie, J. T. (1973). Models of reading and reading disability. *Journal of Educational Psychology, 65*, 9-18.

Guthrie, J. T. & Tyler, S. J., (1968). Cognition and instruction of poor readers. *Journal of Reading Behavior, 10*(1), 57-79.

Hood, J. (1978). Is miscue analysis practical for teachers? *Reading Teacher, 32*(3), 260-266.

MacLean, M. (1980). Qasor: A framework for qualitatively analyzing silent and oral reading. In M. L. Kamil (Ed.), *Perspectives on reading research and instruction: 29th yearbook of the National Reading Conference* (pp. 276-286). Washington, DC: National Reading Conference.

Marzano, R.J., Hagerty, P.J., Valencia, S. W., & SiStefano, P. P. (1987). *Reading diagnosis and instruction: Theory into practice.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

McGinnis, D. J., & Smith, D. E. (1982). *Analyzing and treating reading problems.* New York: Macmillan.

Menosky, D. M. (1976). The three cue systems: #1— graphophonic. In P. D. Allen (Ed.), *Findings in research in miscue analysis: Classroom implications* (pp. 75-79). Urbana, IL: ERIC Clearinghouse on Reading and Communication Skills.

Negin, G. A. (1981). Students' abilities to recongise words in isolation and in context. *Reading Improvement, 18*, 73-80.

Nicholson, T.,& Hill, D. (1985). Research revisited: Good readers don't guess—taking another look at the issue of whether children read words better in context or in isolation. *Reading Psychology, 4*(3-4), 181-198.

Nicholson, T., Lillas, C., & Rzoska, M. A. (1988). Have we been misled by miscues? *The Reading Teacher, 42*(1), 6-10.

Page, W. D. (1976). Miscue research and diagnosis. In P. D. Allen(Ed.), *Findings in research in miscue analysis: Classroom implications* (pp. 139-146). Urbana, IL: ERIC Clearinghouse on Reading and Communication Skills.

Page, W. D. (1976). The three cue systems: #3— semantic. In P. D. Allen (Ed.), *Findings in research in miscue analysis: Classroom implications* (pp. 88-93). Urbana, IL: ERIC Clearinghouse on Reading and Communication Skills.

Pearson, D. P. (1978). On bridging gaps and spanning chasms. *Curriculum Inquiry, 8*, 353-362.

Pflaum, S. W. (1980). Diagnosis of oral reading. *The Reading Teacher, 33*(3), 278-284.

Pflaum, S. W. (1980). The predictability of oral reading behaviors on comprehension in learning disabled and normal readers. *Journal of Reading Behavior, 12*(3), 231-236.

Pikulski, J. (1974). A critical review: Informal reading inventories. *Reading Teacher, 28*(2), 141-152.

Salvia, J., & Ysseldyke, J. (1985). *Assessment in special and remedial education:* 3rd. ed. Boston: Houghton Mifflin.

Schwartz, S. (1984). *Measuring reading competence: A theoretical-perspective approach.* New York: Plenum Press.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*(4), 360-407.

Stanovich, K. E. (1980). Towards an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly, 16*(1), 32-71.

Stanovich, K. E., Cunningham, A. E., & Feeman, D. J. (1984). Intelligence, cognitive skills, and early reading progress. *Reading Research Quarterly, 19*(3), 278-303.

Weber, R. M. (1970). First-graders' use of grammatical context in reading. In H. Levin & J. P. Williams (Eds.), *Basis studies in reading.* New York: Basic Books, Inc.

APPENDIX B

MISCUE ANALYSIS CODING SUMMARY SHEET

Student____________________ Grade__________________ Date__________

Passage	1	2	3		4	5		6
	First Word Heading Miscue	Uncorrected No Meaning Change	Uncorrected Meaning Change		Corrected No Meaning Change	Corrected Meaning Change		Examiner Pronoun.
			Severe	Not Severe		Severe	Not Severe	

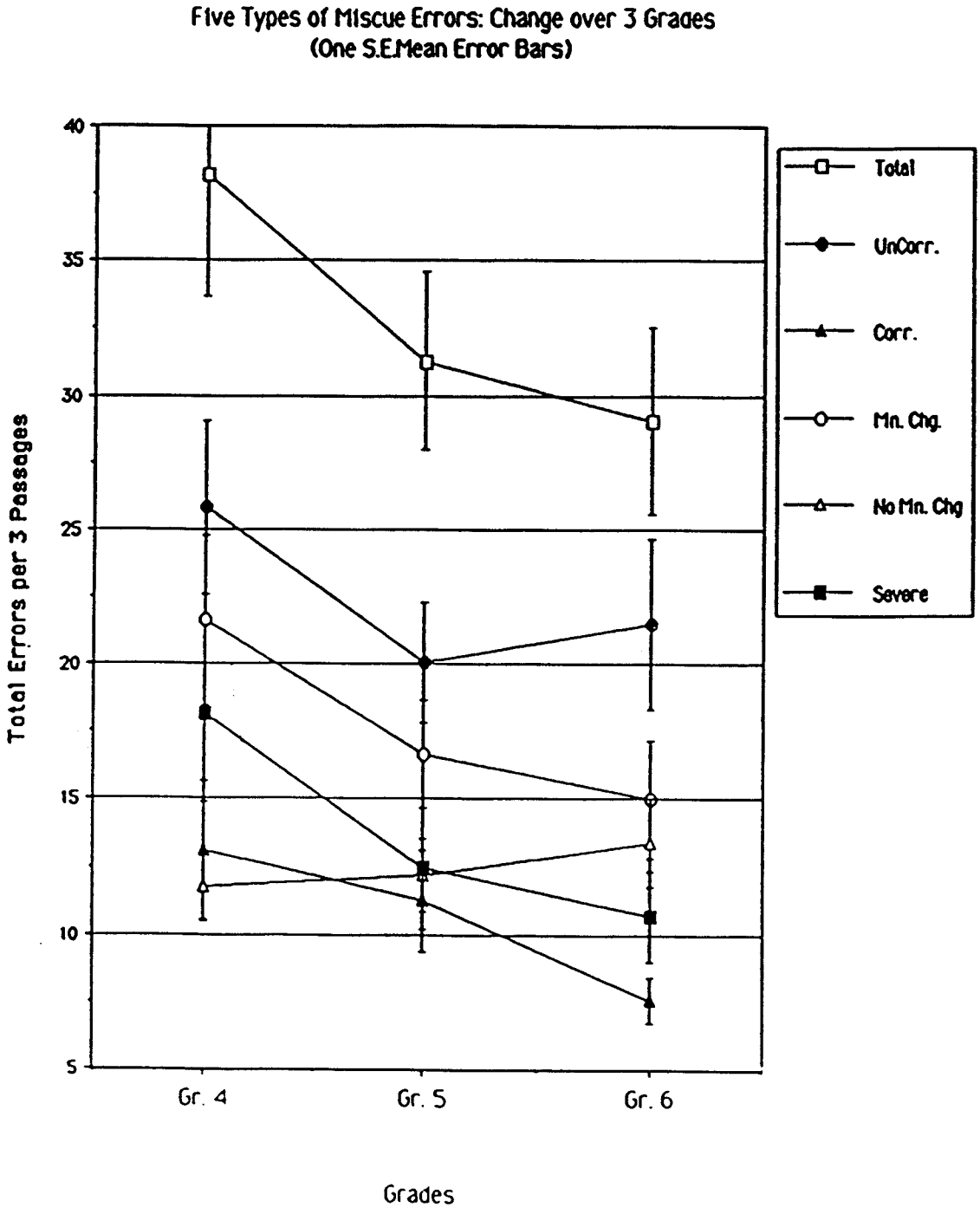
Student____________________ Grade__________________ Date__________

Passage	1	2	3		4	5		6
	First Word Heading Miscue	Uncorrected No Meaning Change	Uncorrected Meaning Change		Corrected No Meaning Change	Corrected Meaning Change		Examiner Pronoun.
			Severe	Not Severe		Severe	Not Severe	

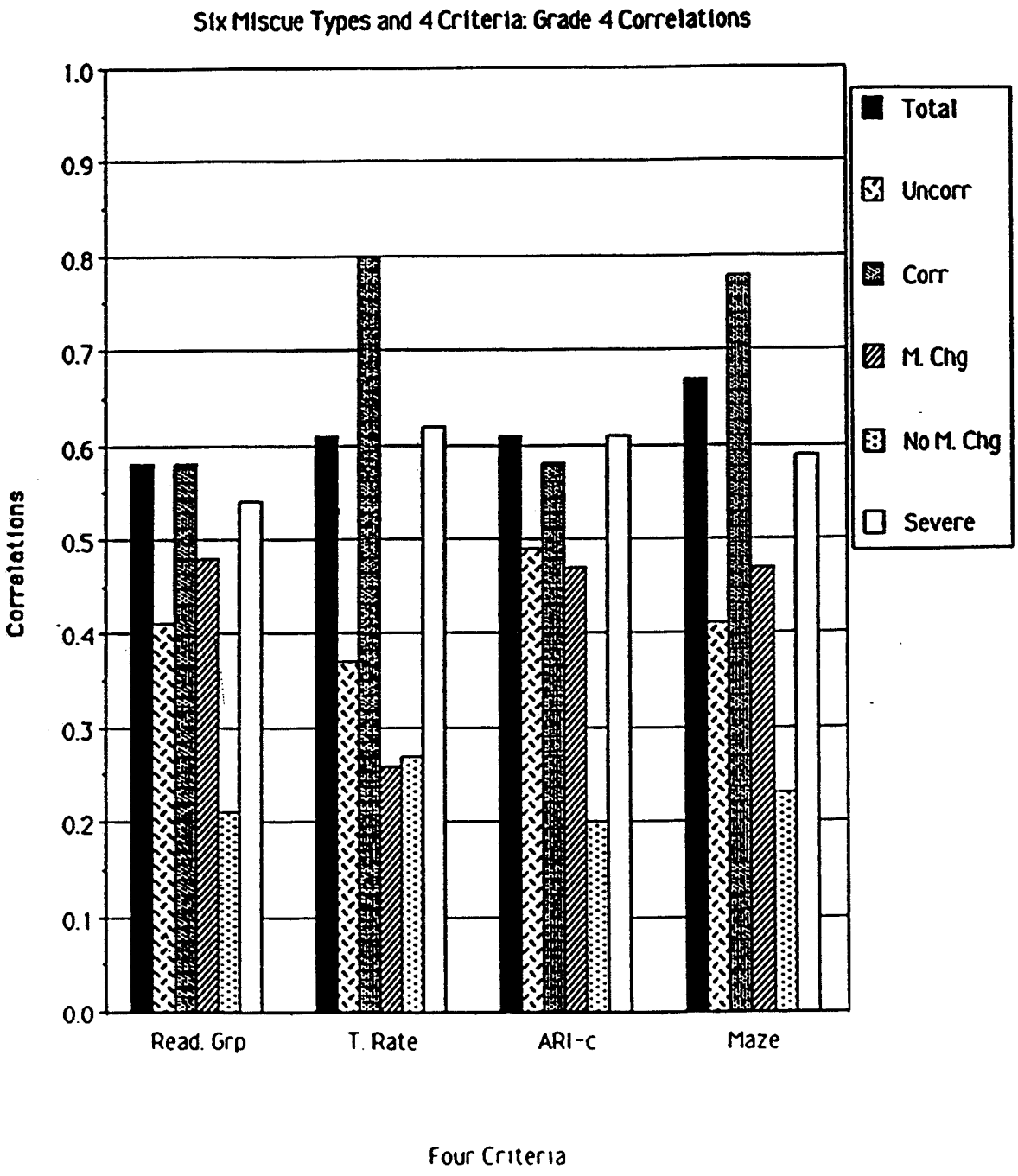
Student____________________ Grade__________________ Date__________

Passage	1	2	3		4	5		6
	First Word Heading Miscue	Uncorrected No Meaning Change	Uncorrected Meaning Change		Corrected No Meaning Change	Corrected Meaning Change		Examiner Pronoun.
			Severe	Not Severe		Severe	Not Severe	

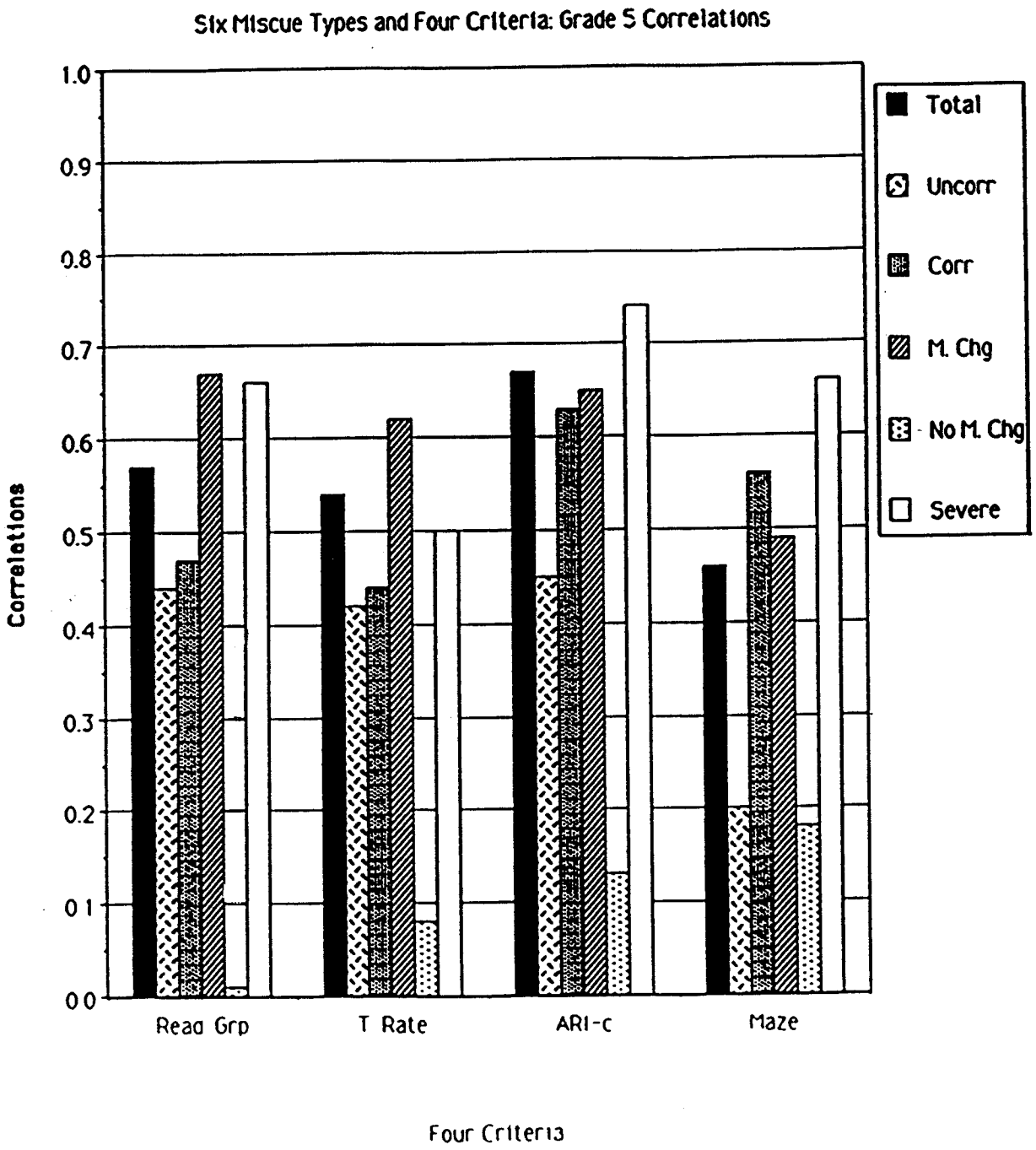
APPENDIX C


Five Types of Miscue Errors: Change over 3 Grades
(One S.E.Mean Error Bars)
Total Errors per 3 Passages
40
35
30
25
20
15
10
5
Gr. 4
Gr. 5
Gr. 6
Grades
Total
UnCorr.
Corr.
Mn. Chg.
No Mn. Chg
Severe

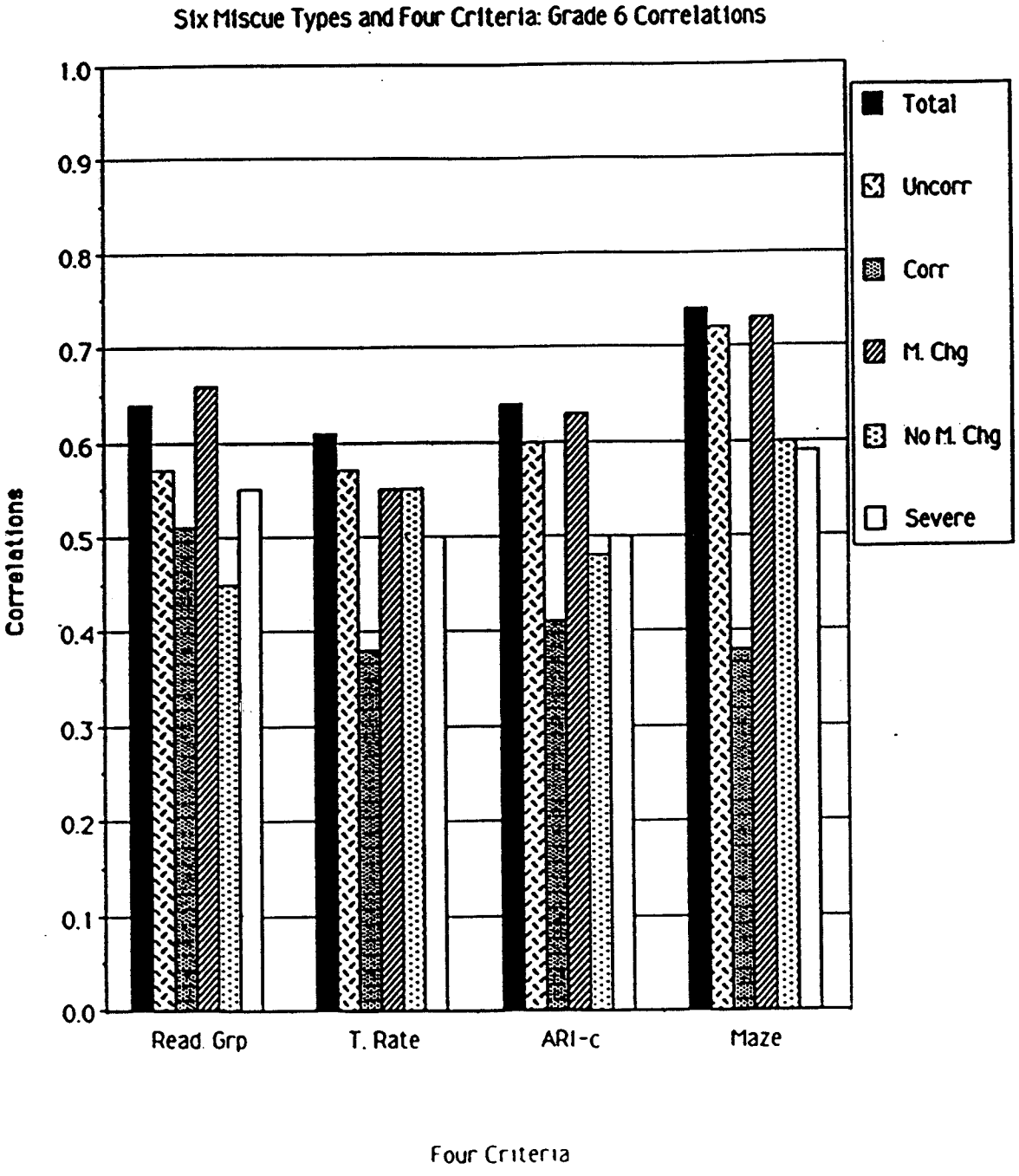
APPENDIX D


Six Miscue Types and 4 Criteria: Grade 4 Correlations
Correlations
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Read. Grp
T. Rate
ARI-c
Maze
Four Criteria
Total
Uncorr
Corr
M. Chg
No M. Chg
Severe

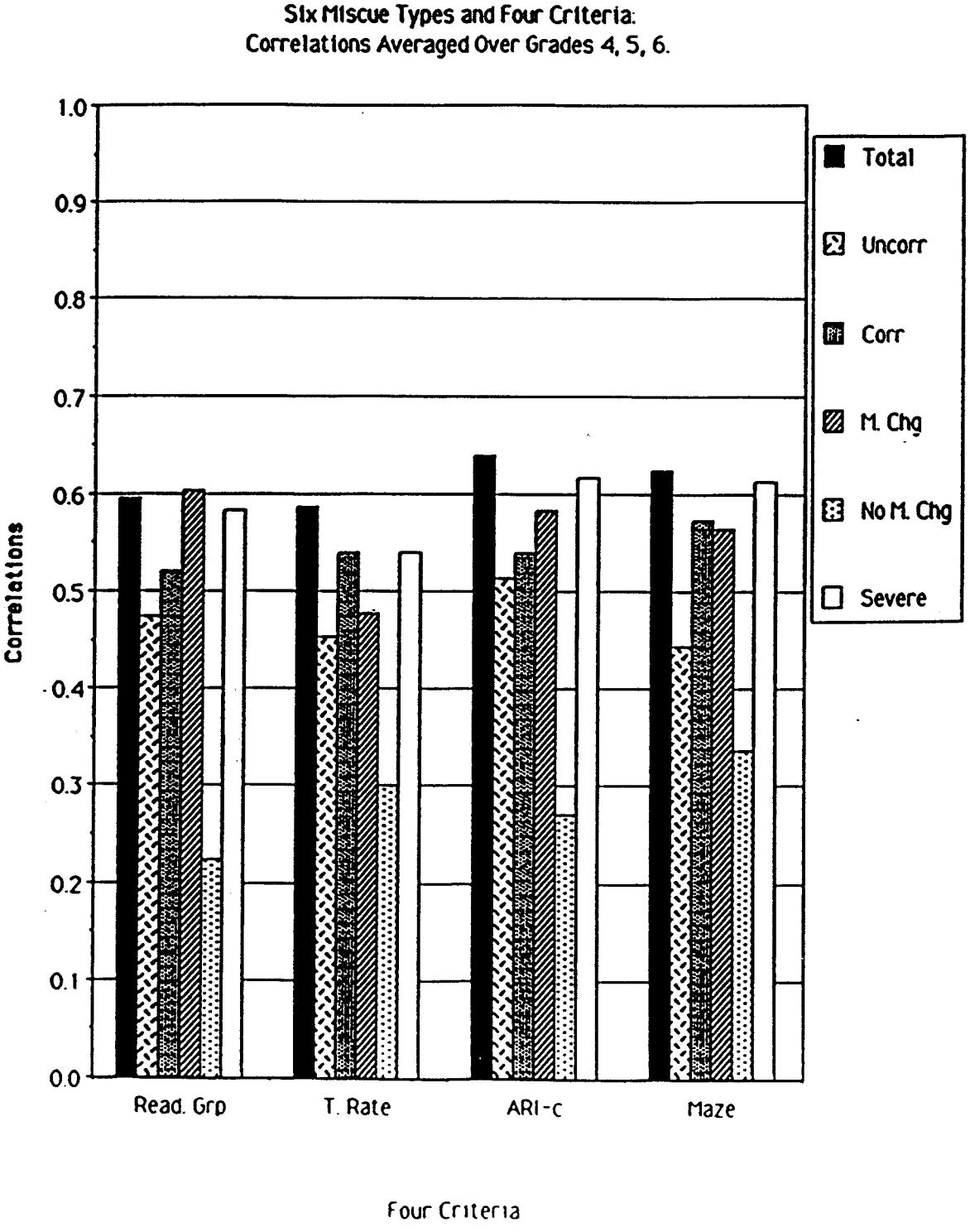
APPENDIX E


Six Miscue Types and Four Criteria: Grade 5 Correlations
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0 1
0 0
Correlations
Total
Uncorr
Corr
M. Chg
No M. Chg
Severe
Read Grp
T Rate
ARI-c
Maze
Four Criteria

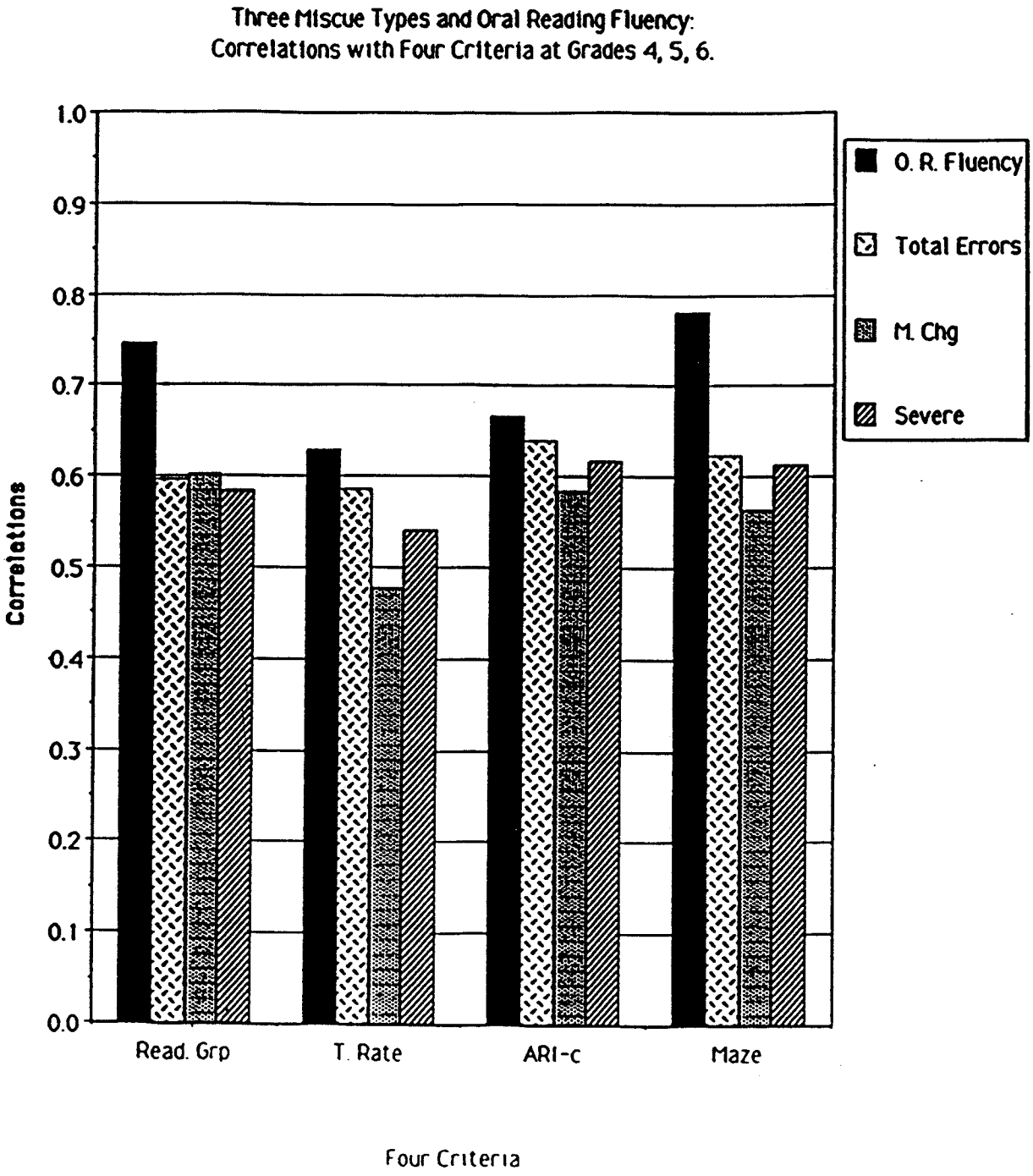
APPENDIX F


Six Miscue Types and Four Criteria: Grade 6 Correlations
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Correlations
Total
Uncorr
Corr
M. Chg
No M. Chg
Severe
Read. Grp
T. Rate
ARI-c
Maze
Four Criteria

APPENDIX G


Six Miscue Types and Four Criteria:
Correlations Averaged Over Grades 4, 5, 6.
Correlations
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Total
Uncorr
Corr
M. Chg
No M. Chg
Severe
Read. Grp
T. Rate
ARI-C
Maze
Four Criteria

APPENDIX H


Three Miscue Types and Oral Reading Fluency:
Correlations with Four Criteria at Grades 4, 5, 6.
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Correlations
O. R. Fluency
Total Errors
M. Chg
Severe
Read. Grp
T. Rate
ARI-c
Maze
Four Criteria