Research in Learning Disabilities

CRITERION VALIDITY OF CURRICULUM BASED ASSESSMENT AND CORRELATION WITH TEACHER RATINGS AND ITBS SCORES

A Research Project of the
Iowa Learning Disabilities Research
Consortium

Bureau of Special Education
IOWA DEPARTMENT OF EDUCATION
January, 1989

A Research Project of the
Iowa Learning Disabilities Research
Consortium

CRITERION VALIDITY OF CURRICULUM BASED ASSESSMENT AND CORRELATION WITH TEACHER RATINGS AND ITBS SCORES

Julie Schendel
School Psychologist
Lakeland Area Education Agency 3
Cylinder, Iowa

and

Marilyn Binder-Reschly
School Psychologist
Heartland Area Education Agency 11
Johnston, Iowa

Bureau of Special Education
IOWA DEPARTMENT OF EDUCATION
January, 1989

State of Iowa
DEPARTMENT OF EDUCATION
Grimes State Office Building
Des Moines, Iowa 50319

STATE BOARD OF EDUCATION

Karen K. Goodenow, President, Spirit Lake
Dianne L. D. Paca, Vice President, Garner
Betty L. Dexter, Davenport
Thomas M. Glenn, Des Moines
Francis N. Kenkel, Defiance
Ron McGauvran, Clinton
Mary E. Robinson, Cedar Rapids
Ann W. Wickman, Atlantic
George P. Wilson III, Sioux City

ADMINISTRATION

William L. Lepley, Director and Executive Officer of the State Board of Education
David H. Bechtel, Special Assistant
Mavis E. Kelley, Special Assistant

Division of Instructional Services

Susan J. Donielson, Ed.d., Administrator
J. Frank Vance, Chief, Bureau of Special Education
James R. Reese, Consultant, Learning Disabilities

"Which is the correct criterion for identification of a mildly handicapped child: teacher perception or test data?" Given these questions and concerns, a further examination of the accuracy of, and factors related to, teacher perceptions of student performance is in order.

A second potential source of error in the current referral-identification process is norm-referenced tests. Norm-referenced tests have been criticized as being time-consuming and expensive, lacking in content validity related to material within a specific curriculum (Jenkins & Pany, 1978), and having little usefulness in instructional decision making. In addition, research suggests that current assessment practices fail to adequately discriminate between low achieving and mildly handicapped students (Ysseldyke, et al, 1982), and between the learning disabled and educable mentally retarded (Neisworth & Greer, 1975; Reynolds & Wang, 1983). In light of these concerns, researchers have suggested a need for alternatives to traditional assessment instruments.

Curriculum-based assessment (CBA) procedures have been cited as an alternative to traditional achievement tests. One form of CBA, curriculum based measurement (CBM) has been specifically constructed to offer the following advantages:

1. CBM tests are simple, efficient and inexpensive.
2. The CBM testing materials are drawn directly from a specific curriculum so that what is taught is also what is tested.
3. CBM tests allow the direct comparison of an individual's performance to that of peers through the development of local norms.
4. The CBM test results are translated into instructional goals and objectives, and facilitate the monitoring of student progress (Deno, 1985).

In addition, CBM is claimed to have many applications. Germann and Tinal (1985), for example, have described a total special education assessment which utilizes CBM in every facet of educational decision making, including screening, eligibility considerations, program planning, performance monitoring and program evaluation.

Despite its advantages and applications, CBM would not be useful for educational evaluation if it were not reliable and valid. A number of validity studies have been conducted assessing CBM procedures in reading (Deno, Mirkin & Chiang, 1982; Deno, Mirkin, Chiang, Kuehnle, Lowry, Marston & Tindal, 1980), written expression (Deno, Marston & Mirkin, 1982), and spelling (Deno, et al, 1980). In each case, researchers have used correlations between standardized achievement tests and CBM measures to demonstrate criterion-related validity. In the area of reading, for example, three studies by Deno, Mirkin & Chiang, (1982) compared student performance on CBM measures such as the oral reading of words in isolation, words in context, reading passages, cloze comprehension and word meaning with performance on selected subtests from the Stanford Diagnostic Reading Test, Woodcock Reading Mastery Test, and the Peabody Individual Achievement Test. Correlations between the CBM measures and

An examination of the recent literature in special education indicates increasing concern about the current practices used to identify and classify students as handicapped. The concern appears to be most apparent with regard to attempts to diagnose students as learning disabled on the basis of "processing deficits" or a discrepancy between ability and achievement. One question in this regard is whether the children that teachers refer for evaluation and possible placement are the same children for whom the available psychometric data would suggest a need for service. Studies by Algozzine, Christenson and Ysseldyke (1982) and Ysseldyke and Thurlow (1984) would seem to indicate that teachers are very accurate in their referral of students in need of special education, reporting that 70-90% of students referred by teachers were subsequently placed in special education.

One possible explanation for the high rate of agreement between referrals and placements would be the existence of two very accurate systems (teacher recognition and referral of student difficulties, and psychometric identification procedures) functioning in tandem. However, Ysseldyke, Algozzine, Shinn and McGue (1982) propose another explanation. Citing research that found few practical differences in test performance between learning disabled and low achieving students, the authors suggest that time-consuming, expensive, norm-referenced tests serve little more than a "confirmatory" function in validating teacher perceptions of performance difficulties even though referred students may not function that much differently from non-referred, low achieving peers. On the other hand, the authors also suggest that current identification procedures may, at times, fail to detect or validate accurately perceived deficits in classroom functioning that make an individual student significantly different from peers.

Current referral and identification procedures lend themselves to two potential sources of error. First, perhaps falsely, the procedures assume that teachers will be completely accurate, unbiased and thorough in referring students with potential learning difficulties. At least one study (Gresham, Reschly and Carey, 1987) has indicated that teachers can be very accurate predictors of both achievement, as measured by the PIAT, and classification as either learning disabled or non-handicapped. On the other hand, studies have found a high prevalence of off-task behavior and inadequate social skills to be typical of mildy handicapped children, suggesting that it is these behaviors, rather than, or along with, academic difficulties which lead to teacher referrals (Gresham & Reschly, 1986; Krupski, 1985; Morrison, Mcmillan & Kavale, 1985). Deno (1985) cites research suggesting that teachers use informal observations to make judgments regarding student performance rather than commercially available tests or teacher-made inventories. He noted that the reliability and validity of the informal observations of teachers are unknown and may be inadequate for decision making. As Gerber and Semmel (1984) have asked

standardized tests ranged from .73 to .91. CBM oral reading tests correlated most significantly with word recognition subtests such as the one from the Woodcock Reading Mastery Test, but were also moderately correlated with comprehension subtests.

An additional finding of the validity studies was the marked difference in performance between learning disabled and regular class students on the CBM measures. In the research on the Words in Isolation measure, regular class students scored 4 to 5 times higher than L.D. resource program students. On the Oral Reading measure, regular class students scored more than 3 times higher. This differentiation between the two groups was cited as additional evidence of the criterion-related validity of CBM. Because CBM results seemed to clearly differentiate between groups, it has been suggested that CBM oral reading measures might be used to estimate reading level and classify readers in lieu of standardized tests.

Subsequent to these initial studies suggesting that CBM performance might reliably discriminate among handicapped and non-handicapped students, additional studies have been conducted examining this question in greater detail. In 1985, Shinn and Marston evaluated 209 regular education Chapter 1 and mildly handicapped students in grades 4 through 6 on CBM materials in reading, spelling, and math and written expression. An analysis of variance for all three educational groups at each grade level yielded statistically significant differences. Regular education students were the most successful while mildly handicapped students were the least successful. Chapter 1 students performed at a level higher than their handicapped peers, but significantly lower than regular education students. Based on these findings, Shinn and Martson suggested that the difference between mildly handicapped, low achieving and regular education students might be a quantitative matter of performance rather than a qualitative difference in learning characteristics.

In 1986, Shinn, Ysseldyke, Deno and Tindal compared 71 low achieving and learning disabled students over a five-week period on CBM measures of reading, spelling and written expression. When the data was analyzed and used to predict group membership, 88.57% of the students were classified correctly on the basis of their CBM performance. Again, the researchers interpreted this data as suggesting that CBM measures might be of value in decision making.

PURPOSE

The purpose of the following research study was to examine the criterion validity of CBM reading measures in three ways. First, the relationship between CBM performance and teacher perceptions of student academic functioning was examined. Second, a comparison was made between CBM scores and reading performance on a group achievement test. Finally, the ability of CBM measures to differentiate between regular education, Chapter 1 and mildly handicapped students was examined.

METHOD

Subjects

The 1223 subjects included in this study were students attending grades one through eight in eight school districts in northwestern Iowa. Because of incomplete data sets for some of the students, sample sizes varied for the various questions addressed in the study. Table 1 shows the distribution of students by grade and the instructional program through which they received instruction. The school districts in the study had participated in a pilot project to develop Curriculum Based Measurement norms. Students were randomly selected in the larger school districts while all the students in a class in the smaller districts were included as subjects in order to achieve grade level sample sizes close to 30 for each district. The resource room sample included children identified as having mental, behavioral or learning disabilities. No attempt was made to determine the exact incidence of the handicaps. Students within self-contained or special class with integration classrooms were not included in the study.

Independent and Dependent Variables

The independent or predictor variables in this study were: (1) CBM reading probes, (2) the reading cluster of the Iowa Test of Basic Skills, and (3) a teacher rating scale. The dependent variable for the investigation was instructional group for reading; i.e., regular education, Chapter 1, or special education resource room.

The CBM probes consisted of a word list and a reading passage randomly selected from the students' reading series at their grade level. Each student was timed for one minute and both words read correctly and errors were calculated. The scores from the ITBS were vocabulary, word analysis, and total reading. The teacher rating scale was a modification of the Teacher Rating of Academic Performance (TRAP) (Gresham, Reschly & Carey, 1987; Reschly, Gresham & Graham-Clay, 1984; Reschly, Grimes and Ross-Reynolds, 1981). The TRAP-Revised (Figure 1) consisted of four questions in which the teacher was asked to rank the student on overall academic performance, study skills, reading in comparison to other class members and reading in relation to grade level expectations. The survey also included data on sex and a question regarding reading instruction group membership in regular education, Chapter 1 or resource room.

Procedure

The CBM and teacher survey data were collected during either February or May of 1988. The specific date for each district was chosen to closely follow the administration of the Iowa Tests of Basic Skills in order to provide concurrent data. While teachers were aware of students' placement as they completed the TRAP-R and were therefore possibly biased by this prior knowledge of classification, they were unaware of ITBS scores for the current year and did not participate in the CBM norming procedures.

TABLE 1

DISTRIBUTION OF SAMPLE ACCORDING TO

READING INSTRUCTIONAL GROUP AND GRADE

GRADE	REGULAR EDUCATION	CHAPTER I READING	RESOURCE ROOM READING	TOTALS BY GRADE
1	125	30	10	165
2	118	55	7	180
3	93	31	8	132
4	105	27	10	142
5	117	33	13	163
6	125	17	13	155
7	128	10	10	148
8	120	5	13	138
TOTALS BY INSTRUCTIONAL GROUP	931	208	84	1223

FIGURE 1

TEACHER RATING OF ACADEMIC PERFORMANCE (TRAP-R)

Please respond as accurately as possible to all of the items below.

1. Child's Name_____

2. Sex M F

3. Compared to other children in my classroom I would estimate the overall academic performance of this child as being in the:

Lowest 10%	Lower 30% but not lowest 10%	Middle 40%	Upper 30% but not highest 10%	Highest 10%
1	2	3	4	5

4. How does this child compare on study skills, such as staying on tasks, following directions, and assignment completion?

*Lowest 10%	Lower 30% but not lowest 10%	Middle 40%	Upper 30% but not highest 10%	Highest 10%
1	2	3	4	5

5. In the area of reading, this child is in what range in comparison to other children in your classroom?

Lowest 10%	Lower 30% but not lowest 10%	Middle 40%	Upper 30% but not highest 10%	Highest 10%
1	2	3	4	5

6. In terms of grade level expectations, this child's skills in reading are:

Well below grade level	Slightly below grade level	At grade level	Slightly above grade level	Well above grade level
1	2	3	4	5

7. This child receives reading instruction:

Only in the regular classroom	In Chapter I reading	In both the regular classroom and resource room	Reading only through the resource room
1	2	4	5

ANALYSIS

Correlational Analysis

A correlation matrix (see Table 2) was generated using all the CBM, ITBS and TRAP-R variables. The resulting range of corelations, .31 to .58, indicated a definite, but moderate relationship between the various independent variables and instructional placements. Correlations within groups were higher as would be expected (CBM = .35 to .79: ITBS = .58 to .76; TRAP-R = .73 to .92). Teacher ratings of academic achievement were very consistent per student and correlated highly with ITBS reading scores (r = .71). Correlations between CBM scores and ITBS skills were slightly higher when Iowa rather than national percentile ranks were used. All correlations were significant at the .0001 level.

Discrepancy Analysis

The first discrepancy analysis involved simple discrepancy scores such as -1 standard deviation below the mean. The number and percent of students correctly classified at various degrees of discrepancy were then compared. Although relatively unsophisticated statistically, the results are more similar to how scores are used in placement decisions in the real world.

In this study, discrepancy analyses required some further decisions. The first was the choice of a mean from which to determine discrepancies. Due to the variations in the CBM measures by grade level and by school district, use of means computed by grade/by district seemed most appropriate. Thus, all CBM raw scores were linearly transformed to Z scores. The transformations were carried out by grade/by district; that is, the CBM raw scores for District A third graders were transformed to Z scores according to the standard formula. Similarly, the raw scores were transformed separately for all other combinations of grades and districts. The transformations did not change the character of the distributions.

The discrepancy analyses using the CBM Z-Scores (based on transformation of the raw scores for each combination of grade and district), the ITBS standard scores (transformed to standard scores from Iowa percentile ranks), and the TRAP indices are presented in Table 3. A student was regarded as correctly classified if his/her score was in an appropriate range, consistent with their placement. Determination of what constituted an appropriate range is somewhat arbitrary, but the same criteria were used for all measures. Thus the comparisons are valid, even if there are disagreements with the precise criteria.

Students were regarded as correctly classified using the following criteria. Regular education students with scores above the mean or within one standard deviation below the mean were regarded as correctly classified. Chapter 1 students were regarded correctly classified if their scores were in the range of the mean to one and one-half standard

TABLE 2

PEARSON CORRELATION COEFFICIENTS
ACROSS ALL DEPENDENT VARIABLES

	1	2	3	4	5	6	7	8	9	10	11	12
1. Words Correct Per Minute	1.00	-0.53	0.79	-0.41	0.45	0.53	0.57	0.55	0.48	0.58	0.63	-0.44
2. Errors Per Minute	-	1.00	-0.35	0.57	-0.42	-0.45	-0.42	-0.43	-0.35	-0.43	-0.46	0.31
3. Passage Correct Per Minute	-	-	1.00	-0.46	0.52	0.39	0.50	0.46	0.43	0.50	0.54	-0.43
4. Passage Errors Per Minute	-	-	-	1.00	-0.41	-0.36	-0.40	-0.39	-0.33	-0.39	-0.41	0.36
5. ITBS Word Analysis IPR	-	-	-	-	1.00	0.59	0.62	0.62	0.59	0.61	0.59	-0.52
6. ITBS Vocabulary IPR	-	-	-	-	-	1.00	0.76	0.64	0.53	0.64	0.66	-0.43
7. ITBS Reading IPR	-	-	-	-	-	-	1.00	0.71	0.63	0.72	0.75	-0.50
8. General Academic Achievement	-	-	-	-	-	-	-	1.00	0.84	0.93	0.86	-0.58
9. Study Skills	-	-	-	-	-	-	-	-	1.00	0.81	0.73	-0.48
10. Reading Comparison	-	-	-	-	-	-	-	-	-	1.00	0.88	-0.57
11. Grade Level Expectations	-	-	-	-	-	-	-	-	-	-	1.00	-0.58
12. Instructional Group	-	-	-	-	-	-	-	-	-	-	-	1.00

TABLE 3

PEARSON PRODUCT CORRELATIONS WITH

INSTRUCTIONAL PLACEMENT[a]

CBM	TOTAL SAMPLES	MALE	FEMALE
Word List Words Correct	-.44	-.49	-.36
Word List Errors	.31	.35	.27
Passage Correct	-.43	-.47	-.38
Passage Errors	.36	.41	.29
ITBS			
Word Analysis[b]	-.52	-.57	-.43
Vocabulary	-.44	-.46	-.40
Reading	-.50	-.51	0.48
TRAP			
Academic Performance	-.58	-.60	-.55
Study Skills	-.48	-.51	-.43
Reading Skills	-.57	-.58	-.55
Grade Level Work	-.58	-.59	-.59

[a]All correlations were significant at .0001.

[b]Only administered to first and second grades.

deviations below the mean. Finally, resource students with scores in the range of one standard deviation or more below the mean were regarded as correctly classified.

Results of the discrepancy analysis are presented in Table 4. Several trends apparent in Table 4 deserve further comment. First, the CBM indices vary markedly in accuracy of classification depending on instructional group and measure. All of the measures were relatively successful with regular education students where the variation was from a respectable 73% for Word List Errors to a very impressive 92% for Passage Words Correct per minute. However, CBM words correct per minute measures were far superior to the errors measures with the Chapter 1 and resource students. The errors measures were quite inaccurate, particularly errors per minute on passages which accurately classified only 4% of resource students.

The results indicate substantial variations in the CBM indices regarding classification accuracy. The best measure was Passage Words correct per minute; the weakest was passage errors per minute.

The ITBS classification accuracy of ITBS scores were similar to CBM words correct measures. The ITBS Reading Subtest correctly classified 89%, 74%, and 74% of the regular education, Chapter 1, and resource students respectively.

The range of scores on the TRAP raw scores for students in the different instructional groups were then analyzed. Scores were in the appropriate discrepancy ranges for very high percentages of students, particularly on the two TRAP ratings involving reading skills (denoted by COMP and EXPT in the table).

Discrimination Function Analyses

A second analysis of correct classification used highly sophisticated statistical procedures, principally discriminant function analyses. Although this approach is impractical in everyday decision making, it does provide an estimate of the upper limit of the possible degree of accurate classification using these measures, singly or in combination.

In this study, the procedure determined the rate of correct classification using weightings of each variable. The use of ideal weights improves the accuracy of classification through construction of complex combinations of the variables to maximize accuracy. In contrast to the prior analysis, the exact instructional grouping had to be predicted rather than an appropriate score range. Thus, the overall rates of correct classification were all lower than in the previous analysis.

TABLE 4

NUMBER AND PERCENT OF CORRECT CLASSIFICATIONS WITH CBM AND ITBS INDICES OF READING

CBM		REGULAR EDUCATION	CHAPTER I	RTP
Word List Correct	N	731	139	66
	%	77%	66%	78%
Word List Errors	N	694	57	24
	%	73%	27%	29%
Passage Words Correct	N	868	159	67
	%	92%	75%	80%
Passage Errors	N	820	78	3
	%	87%	37%	4%

ITBS		REGULAR EDUCATION	CHAPTER I	RTP
Word Analysis	N	.....Not		
	%		Available.....	
Vocabulary	N	825	137	57
	%	87%	65%	68%
Reading	N	842	157	62
	%	89%	74%	74%

TRAP		REGULAR EDUCATION	CHAPTER I	RTP
Acad	N	916	155	58
	%	97%	73%	69%
Study	N	912	145	46
	%	96%	68%	55%
Comp	N	824	158	72
	%	87%	75%	86%
Expt	N	811	165	74
	%	86%	78%	88%

NOTE: A student was regarded as correctly classified by the measure according to the following criteria: 1) Regular Education students with scores above -1 SD below the mean; 2) Chapter I students with scores in the range of the mean to 1.5 SD below the mean; 3) Resource Teaching Program students were regarded as classified accurately if their scores were greater than one standard deviation below the mean.

The discriminant function results for the four CBM measures, using the Grade by District Z scores are provided in Table 5. Overall correct classification was 68.25%. Correct classification for the regular education, Chapter 1, and resource instructional groups was 73.0%, 49.1%, and 60.0% respectively. The correct classification in a similar analysis using the CBM raw scores (results not shown) was 68.45%, a nearly identical result.

The ITBS Group by Grade Z scores for Vocabulary and Reading were then used in a discriminant function analysis (see Table 6). Once again, the CBM and ITBS results were quite similar. The overall correct classification was about 65% (vs 68% for CBM). Correct classification for the instructional grouping varied as follows: Regular education = 68.7%; Chapter 1 = 44.5%; and resource = 71.3% Virtually identical results were obtained for the ITS standard scores (data not shown).

Discriminant function analysis on the four teacher ratings raw scores are presented in Table 7. In this analysis, the correct classification over all groups was 77.7%, a level somewhat above the CBM and ITBS measures. The same pattern of correct classification was apparent with more accurate grouping of regular education and resource students (82.4% and 74.7% respectively) versus Chapter 1 students (57.7%).

A final discriminant function analysis was conducted combining all variables. The results in Table 8 show correct classification improved over ITBS and CBM, but slightly below the level reported in Table 7 for TRAP variables alone. It should be noted that the teacher ratings of the four characteristics (e.g., study skills, reading level in comparison to the class, etc.) may have been influenced by the teachers' prior knowledge of the students' instructional grouping. Even with this caveat, the TRAP relationships to instructional grouping have to be regarded as impressive.

Analysis of Instructional Group Means

Analysis of variance (ANOVA) procedures were used to determine whether differences existed between the three instructional groups on the various indices of reading. F-tests with associated probability values are presented for each variable. Differences at the .05 level or lower were followed by Neuman - Keuls (N-K) post hoc tests to determine which pairs of the three instructional group means were different.

The first four rows of Table 9 provide information on the Z-Scores for the four CBM measures: Words Correct Per Minute, Errors Per Minute, Passage Words Correct Per Minute, and Passage Errors Per Minute. All overall comparisons of instructional group means on all CBM variables were statistically significant. All pairs of instructional group means were also different statistically as indicated in the last column of Table 7.

TABLE 5

DISCRIMINANT FUNCTION CLASSIFICATION BY INSTRUCTIONAL GROUP

USING FOUR CBM GRADE BY DISTRICT Z SCORES

ACTUAL GROUP	N		PREDICTED GROUP R	C	RTP
Regular Education (R)	794	N	580	162	52
		%	73.0%	20.4%	6.5%
Chapter I (C)	171	N	34	84	53
		%	19.9%	49.1%	31.0%
Resource Program (RTP)	65	N	7	19	39
		%	10.8%	29.2%	60.0%

Overall correct classification = 68.25%.

NOTE: Overall correct classification with the CBM raw scores was 68.45%.

TABLE 6

DISCRIMINANT FUNCTION CLASSIFICATION BY INSTRUCTIONAL GROUP

USING TWO ITBS GROUP BY GRADE Z SCORES

ACTUAL GROUP	N		PREDICTED GROUP R	C	RPT
Regular Education (R)	927	N	637	213	77
		%	68.7%	23%	8.3%
Chapter I (C)	209	N	40	93	76
		%	19.1%	44.5%	36.4%
Resource Program (RTP)	80	N	3	20	57
		%	3.8%	25%	71.3%

Overall correct classification = 64.7%.

NOTE: Overall correct classification using the ITBS standard scores was 64.1%.

TABLE 7

DISCRIMINANT FUNCTION CLASSIFICATION BY INSTRUCTIONAL GROUP

USING FOUR TRAP RATINGS RAW SCORES

ACTUAL GROUP	N		PREDICTED GROUP R	C	RPT
Regular Education (R)	932	N	768	120	44
		%	82.4%	12.9%	4.7%
Chapter I (C)	208	N	44	120	47
		%	19.7%	57.7%	22.6%
Resource Program (RTP)	83	N	5	16	62
		%	6.0%	19.3%	74.7%

Overall correct classification = 77.7%.

TABLE 8

DISCRIMINANT FUNCTION CLASSIFICATION BY INSTRUCTIONAL GROUP

USING ITBS Z SCORES, CBM Z SCORES, AND TRAP RAW SCORES

ACTUAL GROUP	N		PREDICTED GROUP R	C	RPT
Regular Education (R)	765	N	612	117	36
		%	80.0%	15.3%	4.7%
Chapter I (C)	164	N	21	100	43
		%	12.8%	61.0%	26.2%
Resource Program (RTP)	60	N	6	14	40
		%	10.0%	23.3%	66.7%

Overall correct classification = 76.04%.

TABLE 9

EXAMINATION OF INSTRUCTIONAL GROUP MEANS ON VARIOUS INDICES OF READING

READING MEASURES	N	INSTRUCTIONAL MEANS GROUP 1 Reg	2 Chapt	3 RTP	F	P	N-K
CBM (Z-Scores)							
Word Lists Words Correct/Minute	1032	0.25	-0.71	-1.20	158.22	.0001	1 2 3
Word Lists Errors Per Minute	1032	-0.178	0.53	0.88	71.48	.0001	1 2 3
Passage Correct Per Minute	1230	0.267	-0.69	-1.31	229.62	.0001	1 2 3
Passage Errors Per Minute	1230	-0.19	0.39	1.22	113.67	.0001	1 2 3
ITBS (Z-Scores)							
ITBS Word Analysis IPR	340	0.3004	-0.69	-1.06	54.25	.0001	1 2 3
ITBS Word Analysis (NPR)	340	0.30	-0.68	-1.22	58.24	.0001	1 2 3
ITBS Vocabulary IPR	1216	0.24	-0.64	-1.11	154.26	.0001	1 2 3
ITBS Vocabulary NPR	1216	0.25	-0.62	-1.21	168.21	.0001	1 2 3
ITBS Reading IPR	1219	0.28	-0.73	-1.32	230.76	.0001	1 2 3
ITBS Reading NPR	1219	0.36	-.73	-1.38	230.51	.0001	1 2 3
TRAP (Raw Scores)							
Study	1237	3.56	2.31	1.74	197.64	.0001	1 2 3
Acad	1234	3.61	2.15	1.46	332.13	.0001	1 2 3
Comp	1235	3.67	2.12	1.59	327.03	.0001	1 2 3
EXPT	1232	3.62	2.08	1.56	339.81	.0001	1 2 3

The differences were consistently in the pattern of regular education better than Chapter 1 and both better than resource room. The sizes of the differences were substantial. The differences between regular education and Chapter 1 were about one standard deviation on both words correct per minute indices. The differences between regular education and resource on the words correct measures were even larger, about 1.4 to 1.6 standard deviations. The differences on the errors measures were in the same direction, but the magnitude was not as large. On errors measures the three groups differed by about .6 SD to 1.0 SD. All of the differences are substantial and practically significant.

Similar results were obtained with the Z-Scores computed from the Iowa and National percentile rank scores for the ITBS reading measures. All overall and pair-by-pair comparisons of instructional group means were statistically significant. Again, the differences were substantial and, in our judgment, practically significant. The size of the regular education and Chapter 1 differences for the ITBS measures were in the range of 0.9 SD to 1.1 SD and 1.3 SD to 1.7 SD for regular education and resource. It should be noted that the magnitude of the differences between Regular Education, Chapter 1, and resource were very similar for the CBM words correct measures and the various ITBS scores. The CBM errors measures yielded smaller differences.

Examination of TRAP variables using raw scores found each of the teacher rating items to differentiate the instructional groups. The differences were statistically significant and in the same direction as previous results: i.e., regular education greater than Chapter 1, and Chapter 1 greater than resource.

Analysis of Means by Sex

Comparisons of means by sex yielded somewhat inconsistent differences. Results in Table 10 indicate that females obtained significantly higher scores than males on the words correct CBM indices, but not on the CBM errors measure. The magnitude of the differences on the words correct measures were .25 and .55 SD, a level that approaches the casual rule of thumb for practical significance. In contrast, the errors means for males and females were virtually identical.

Comparable results are reported in Table 8 for the ITBS measures. Four of the six scores (reflecting three measures across two standardization groups) were statistically significant, the Word Analysis measure used with younger students and the Reading measure used with older students. The Vocabulary subtest means were not different for males and females. The male-female differences on the ITBS measures were about .20 to .25 SD, slightly smaller than the male-female differences on the CBM variables. The reader should be cautioned that the male-female differences on both sets of measures were modest in size.

TABLE 10

EXAMINATION OF SEX MEANS ON VARIOUS INDICES OF READING

READING MEASURES	N	MEANS OF SEX		F	P	N-K
		Males	Females			
CBM (Z-Scores)						
Word Lists Words Correct	1032	-0.127	0.144	25.86	.0001	1 2
Word Lists Errors	1032	0.042	-0.037	1.86	.1726	N S
Passage Words Correct	1230	-0.24	0.26	115.66	.0001	1 2
Passage Errors	1230	0.05	-0.04	3.44	.064	N S
ITBS (Z-Scores)						
ITBS Word Analysis IPR	340	-0.12	0.129	6.69	.0101	2 1
ITBS Word Analysis NPR	360	-0.12	0.11	6.10	.014	2 1
ITBS Vocabulary IPR	1216	-0.043	0.046	3.10	.079	2 1
ITBS Vocabulary NPR	1216	-0.038	0.048	2.94	.087	N S
ITBS Reading IPR	1219	-0.10	0.11	19.50	.0001	2 1
ITBS Reading NPR	1219	-0.047	0.178	18.75	.0001	2 1

Prediction of Instructional Group

Extensive analyses were conducted to determine the degree to which the three sets of variables could predict instructional group placement. Two procedures were used: multiple regression and stepwise regression. The former addresses the question regarding the maximum relationship of the measures to the dependent variable of instructional placement. All variables are entered into the prediction regardless of relative contribution or whether they are needed to improve the accuracy of the prediction. Thus, multiple regression provides a maximum estimate of the prediction and does not provide an indication of which variables in the analysis contribute more than others. In contrast, stepwise regression enters variables into the equation in order of their relative contribution to prediction of the dependent variable, i.e., instructional placement. Stepwise regression provides a basis for determining the best predictors among an array of variables while multiple regression indicates the maximum prediction using all possible information.

Multiple Regression Analysis

The multiple regression analyses using four sets of variables are presented in Table 11. The four sets were: (1) The ITBS vocabulary and reading scores; (2) the CBM scores; (3) the TRAP ratings; and, (4) all scores combined. All sets of measures had substantial, statistically significant relationships to instructional grouping. All F-tests were significant at the .0001 level indicating that the overall relationships were not attributable to chance factors. The R^2 index varied from about .25 for the CBM raw scores to about .40 for the combination of all variables. The multiple R^2, interpreted much like simple Pearson correlations, varied from about .5 to .6.

In what is now a familiar pattern, the CBM and ITBS relationships were very similar. The relationships were especially apparent for the Grade by District Z-scores from CBM and ITBS where the R^2 of .3004 and .3046 are virtually identical. Clearly, both ITBS and CBM sets of measures provide strong and comparable predictors of instructional grouping.

The strength of the TRAP ratings as predictors of instructional grouping were, perhaps, surprising. The TRAP ratings alone were better than either ITBS or CBM, and only slightly less efficient than the overall combination of TRAP, ITBS, and CBM.

Stepwise Regression Analysis

The stepwise regression analyses are reported in Tables 12 and 13. The analyses are intended to show which among an array of variables are the best predictors of some criterion. However, the stepwise regression analysis provided relatively little additional information. The essential results from the prior multiple regression analyses, that the three sets of variables are about equally good predictors, are not modified by the stepwise regression findings.

TABLE 11

MULTIPLE REGRESSION ANALYSIS PREDICTING INSTRUCTIONAL GROUPING

Predictor Variables	F	P	R^2	Multiple R
A. ITBS Measures				
1. Standard Scores on Reading & Vocab.	118.5	.0001	.2814	.5305
2. Grade by District Z-Scores on Reading & Vocab.	132.6	.0001	.3046	.5519
B. CBM Measures				
1. Raw Scores, Words Correct & Errors on Passage & List	78.58	.0001	.2347	.4844
2. Grade by District Z-Scores on Words Correct & Errors on Passage & List	110.05	.0001	.3004	.5481
C. TRAP - Teaching Ratings				
1. Raw Scores on Overall Academics, Study Skills & Reading Level	175.12	.0001	.3649	.6041
D. All Variables				
1. ITBS Standard Scores, CBM Raw Scores, TRAP Raw Scores	51.67	.0001	.3883	.6231

The stepwise regressions carried out separately for the three sets of variables are reported in Table 12. The results show that the best single variables were: (1) for ITBS, the Reading standard score; (2) for CBM, the Passage Words Correct Z-Score; and (3) for TRAP, the Overall Academic Rating raw score. The three variables would appear to be the best predictors within each of the sets of variables. Moreover, the three variables are about as good as all of the variables in the respective sets of measures. The difference in the R^2 for each variable alone with all of the variable in the set was only .03 to .04 (see Table 10). Adding more information from the same set apparently does little good. For example, adding the rest of the ITBS scores to the Reading standard score increases the multiple R by only .03, a very small, probably trivial increment. In like fashion, additions of other CBM variables to the Passage Words Correct Z-Score or more TRAP variables to the Overall Academic Rating have similar inconsequential effects.

The stepwise regression with all variables from the three sets is summarized in Table 13. Six variables met the criterion of a statistically significant addition to the R^2, but the addition of each variable to the others had a quite small effect. The three best variables were TRAP Overall Academic Rating, CBM Passage Words Correct, and ITBS Reading.

Although the TRAP Overall Academic Rating was the first variable entered by the multiple regression analysis, the difference between this and the next two variables, ITBS Reading and CBM Passage Word Correct, was rather small. The magnitude of the differences is shown in Table 13. Again, clearest interpretation of the results is that the best measures from the three sets of variables are about equally good as predictors of instructional group, all showing strong, consistent relationships to instructional group.

Analysis of CBM Incidence of Below Average Performance

Table 14 shows the distribution of the Passage Words Correct Z-Scores (the single best Curriculum Based Measurement predictor of instructional group membership) in regards to the percentage of scores which fall below specific standard deviation cut offs. By using only Passage Words Correct, 33.01% of the scores fell between 0 and -1 standard deviation below the mean. Likewise, 11.51% of the obtained scores were found between -1 and -1.5 standard deviations below the mean. Between -1.5 and -2.0 standard deviations, 3.79% of the sample scores were found. The remaining 2.9% of the sample obtained Passage Words Correct scores below -2.0 standard deviations.

DISCUSSION

This study found three sets of measures which predicted reading instructional groups equally well. While gathered or provided by separate sources, the data are highly interrelated as shown by the fact that all correlations (225 in the entire matrix) were significant at the point .0001 level. However, the independent variables examined in this study

TABLE 12

STEPWISE REGRESSION BY THE THREE SET OF VARIABLES

	Variable	R^2	(R)
Best Single ITBS	Reading Standard Score	.2726	.5221
All ITBS	Grade by District Z-Scores	.3046	.5519
Best Single CBM	Passage Words Correct	.2620	.5119
All CBM	Grade by Distrct Z-Scores	.3004	.5481
Best Single TRAP	Overall Academic Raw Scores	.3332	.5772
All TRAP	Raw Scores	.3649	.6041

TABLE 13

STEPWISE REGRESSION FOR ALL VARIABLES

STEP	VARIABLE	NUMBER IN	R^2	F	P
1.	TRAP, Overall Academic Raw Score	1	.3332	493.61	.0001
2.	CBM Passage Words Correct Raw Score	2	.3601	41.47	.0001
3.	ITBS Reading Standard Score - Nat'l.	3	.3693	14.48	.0002
4.	CBM Passage Errors Raw Score	4	.3749	8.79	.0031
5.	ITBS Vocabulary Standard Score	5	.3800	8.09	.0045
6.	TRAP Reading Grade Level Raw Score	6	.3855	8.81	.0031

TABLE 14

PERCENTAGE OF PASSAGE WORDS CORRECT (Z-SCORES) WITHIN CERTAIN STANDARD DEVIATION CUT OFFS

LEVELS	PERCENT
Above the Mean	48.79
Between the Mean and -1 Standard Deviation	33.01
Between -1 and -1.5 Standard Deviations	11.51
Between -1.5 and -2 Standard Deviations	3.78
Below -2 Standard Deviations	2.90
TOTAL	100.00

are not the usual criteria used for special education instructional classifications.

TRAP-P

Teacher judgment plays a part in the initial referral of students and in forming within-class reading groups, but a the present time, the final decision to place a student in special education is largely based on normed referenced, standardized test data in most cases. Teachers from whom the TRAP ratings were obtained were aware of the student's instructional grouping. Teacher awareness of student instructional grouping may have contaminated the ratings of the student behaviors or status on the four TRAP items; study skills, overall academics, reading in relation to others in the classroom, and reading in relation to grade expectations. Knowledge that the student was receiving instruction in Chapter 1 or resource program might have influenced the teacher to rate the student lower than otherwise might have been the case. The point is speculation.

Even allowing for possible contamination, by teacher awareness of instructional group, the TRAP instructional grouping relationship was impressive. Specifically, the TRAP study found teacher judgment, especially of reading level and general grade level expectations, to predict instructional group placement and to correlate well with other independently administered measures, especially the reading score on the ITBS ($r = .71$ and $.75$ respectively). While half the students placed in regular education were rated by their teachers as doing average or above average in all academic areas, only 3% of the Chapter 1 or resource students received that rating. In fact, 70% of the resource students were said to be working at least 1.5 standard deviation below average.

The fact that the resource students are doing poorly in all subject areas is noteworthy. Students who receive their reading instruction through special education resource programs are still spending most of the school day in the regular classroom and may need modifications in assignments and some extra help in subjects other than reading. Current practice relies on the regular education teacher to do this in many cases. Regardless of whether are new proposals are embraced that promote more services for low achieving students in the regular classroom, it must be recognized that low achieving students are already there. Research and pilot programs should focus on ways to provide more support for the regular education teacher, promoting additional teaching techniques, innovative materials, and the use of additional volunteer and professional personnel in the classroom to help all students experiencing academic difficulties.

Study skills, as rated by the classroom teacher, were moderately correlated to reading achievement as measured by CBM words correct measures and all ITBS reading scores. Are students poor readers because they have inadequate study and attending skills? Or, are the students frequently off task because of frustration and boredom in classes in which they cannot keep up? Prior research has indicated students are often referred because of poor behavior. Based on this study, such referrals would frequently identify students in need of assistance in reading instruction.

ITBS

The study also investigated the relationship of readily available ITBS scores to CBM reading scores. Many schools are spending time and money every year to administer the ITBS to students in all grade levels, yet the schools frequently seem troubled with the multitude of data that is analyzed, tabled, and returned and unsure of how to pass information on to parents or how to use it to improve educational programs. The fact that ITBS reading scores were as effective in predicting instructional group placement as either CBM words correct or teacher ratings was, quite frankly, a surprising side effect of the study. Reading scores are used as one criterion for Chapter 1 screening, and this fact no doubt inflated the results. However, the percentile cutoff used for Chapter 1 programming should merely serve to separate regular education students from those receiving some kind of assistance; the ITBS data was instead able to discriminate between Chapter 1 students and special education resource program students. Perhaps some lower percentile cutoff score could be used as screening for special education referrals.

Only ITBS subtest scores were examined in this study. ITBS results are now broken down into specific criteria or as goals and objectives in planning reading programs. The present results would support more reliance on ITBS scores as valid indications of current student achievement. We found a respectable correlation between ITBS reading scores and reading achievement as judged by teachers, measured by CBM, and the resulting instructional placement for reading.

CBM

The initial focus of this study was to compare CBM data, which were being gathered as part of a CBM norming project, with student instructional placement. The present study differs from two earlier examinations of the classification issue (Shinn & Marston, 1985; Shinn, et al., 1986) in several important ways. First, the sample size represented in the current study was significantly larger than those reported in the earlier research (209 students in the Shinn & Marston study; 70 in the Shinn, et al. study; 1288 students in the current study). A second important difference was the number of CBM probes administered. In the Shinn & Marston study, students read from three randomly selected reading passages. In the Shinn, et al. study, students read four different word lists over the course of five weeks. In the current study, students read from a single reading passage and a single word list administered on the same day. Despite the more limited sampling of performance in the present study, the data shows the same pattern of results reported in previous research: the ability to stratify three groups (regular education, Chapter 1, and resource program) at a statistically significant level. The present study also differs from other studies in that subjects were from eight different school districts using five different textbook series, Further analysis is needed to know the consistency across grades or schools or texts and correlations with schools' ITBS scores in order to judge the feasibility of establishing "standard" CBM norms for specific grades, reading series, or ITBS score ranges.

A significant finding was that CBM raw data was just as robust as data which was transformed to standard Z scores, implying that CBM data can be useful without complicated statistical transformations. However, for research purposes, discriminant function analyses were run on the current data and the current classification of students was 68.25% compared to 88.56% in the Shinn and Marston study (1986).

The agreement of CBM data with current instructional placement was computed using discrepancy analyses. Passage Words Correct Per Minute was found through the stepwise regression to be the most effective CBM measure with 6.68% of the students in this study scoring at least 1.5 standard deviation below the mean of their class. In the present samples, 6.76% were receiving instruction in the resource program. If the results are typical of the student population, such a cutoff point would identify approximately the same number of students as are now being served.

In view of the discriminative validity, ease of administration, and time and cost efficiency, CBM measures certainly appear to be an attractive alternative to the current assessment/identification model. However, some inherent dangers lie within the very factors that make CMB such an attractive alternative. One is the possibility of over generalizing about individual student performance on the basis of too few samples of behavior. The fact that a single one-minute reading test was sufficient to accurately classify a high percentage of a large group sample does not in any way imply that such an approach would be the best practice when making a decision about an individual student. In sites where CBM data is used to make screening and eligibility decisions (such as in Pine County, Minnesota, or in the Minneapolis Public Schools), multiple samples of student performance (e.g., nine probes) are collected over the course of several days (Shinn 1988). The multiple sampling of performance has the potential effect of increasing the reliability of data and decreasing the standard error of measurement (Marston & Magnusson, 1987). Since CBM measures are so brief, multiple samplings of performance are still time and cost efficient.

The data presented in Table 9 incorporate a broader range of scores for each instructional group. CBM could thus be used as one criterion and further discrimination could be garnered from behavioral observations, teacher ratings and ITBS-Scores. While a stepwise regression did not find additional measures to improve predictions significantly, from a practical point of view, a multiple-factored assessment of a student before making any eligibility decision is necessary. The research found that the use of only one CBM measure, Passage Words Correct, could be sufficient when incorporated with other readily available data, such as ITBS results and teacher judgment, in making placement decisions and resulting in similar numbers of students receiving services. The use of such measures could restructure professional roles and should provide more useful information for goal and objective development, and progress monitoring than current assessment instruments and procedures.

Finally, an area of concern with regard to the interpretation of this study lies in the potential danger of divorcing CBM screening and eligibility functions from other applications in the areas of goal setting and performance monitoring. Limiting the application of CBM solely to

screening and eligibility decision making would fail to take advantage of its usefulness in the development of more effective educational programs for children. While there might be some time and cost benefits to substituting CBM procedures for current identification practices, there is no evidence to suggest that there would be any concomitant improvement in the outcomes of special programming.

The true value of CBM lies in its sensitivity to small changes in instruction over short periods of time. This factor, along with CBM's other advantages, allows teachers to design an effective, child-specific instructional program which has the greatest likelihood of returning the below-average child to a more normal level of academic functioning. It seems clear that CBM has tremendous potential to improve educational programming for children, but only when it is adopted as a total system and applied at all educational decision points.

Conclusion

In conclusion, this study found several non-traditional assessment tools to be good predictors of current reading group instruction for regular education and mildly handicapped children. ITBS reading scores, CBM reading probes, and a teacher rating scale (TRAP) were able to stratify a large sample of elementary school-age into Chapter 1, resource room, or regular education reading groups with 76% accuracy.

The possible utility of such measures lies in their efficiency and availability. Rather than employing special education support service personnel to conduct lengthy individual assessments using traditional psychometric measures, this study would indicate that the same students could be selected to receive additional instruction in reading by less invasive and time-consuming measures. ITBS scores are updated annually and readily available in most Iowa schools. CBM probes can be administered by regular education teachers or paraprofessionals.

The more extensive use of these selection criteria, along with more reliance on teacher referrals and judgment, could free special education personnel to spend more time on consultation with teachers and parents with the ultimate goal of solving many more behavior and academic problems, and guiding and documenting student progress through management plans, rather than testing to maintain the status quo of the current system.

Screening for eligibility by any measures needs to be closely tied to program planning and implementation. CBM measures provide a quick way for teachers to measure and document progress and make appropriate program changes to sustain learning momentum. Future studies might focus on other areas of CBM and ITBS; e.g. written language and math, and how well such measures in these areas can identify students in need of assistance, and provide concurrent error analysis to begin prompt remediation of student deficits without the current test-retest-qualify-IEP time lag and personnel drain.

REFERENCES

Algozzine, B., Christenson, S., & Ysseldyke, J. (1982). Probabilities associated with the referral to placement process. Teacher Education and Special Education, 5, 19-23.

Deno, S. (1985). Curriculum-based measurement: The emerging alternative. Exceptional Children, 52, 219-232.

Deno, S., Marston, D., & Mirkin, P. (1982). Valid measurement procedures for continuous evaluation of written expression. Exceptional Children, 48, 368-371.

Deno, S., Mirkin, P., and Chiang, B. (1982). Identifying valid measures of reading. Exceptional Children, 49, 36-45.

Deno, S., Mirkin, P., Chiang, B., Kuehnle, K., Lowry, L., Marston, D., & Tindal, G. (1980). Current status of research on the development of a formative evaluation system for learning disabilities programs. Bridges to Tomorrow: The Best of ACLD, 125-147.

Gerber, M. M. & Semmel, M. I. (1984). Teacher as imperfect test: Reconceptualizing the referral process. Educational Psychologist, 19, 137-148.

Germann, G., & Tindal, G. (1985). An application of curriculum-based assessment: The use of direct and repeated measurement. Exceptional Children, 52, 244-265.

Gresham, F. M. (1983). Multitrait-multimethod approach to multifactored assessment: Theoretical rationale and practical application. School Psychological Review, 12(1), 26-34

Gresham, F. M., & Reschly, D. J. (1986). Social skills deficits and low peer acceptance on mainstreamed learning disabled children. Learning Disabilities Quarterly, 9, 23-32.

Gresham, F. M., Reschly, D. J. & Carey, M. P. (1987). Teachers as "test": Classification accuracy and concurrent validation in the identification of learning disabled children. School Psychology Review, 16(4), 543-553.

Jenkins, J., & Pany, D. (1978). Standardized achievement tests: How useful for special education? Exceptional Children, 44, 448-453.

Krupski, A. (1985). Variations in attention as function of classroom task demands in learning handicapped and CA-matched non-handicapped children. Exceptional Children, 52(1), 52-56.

Marston, D., & Magnusson, D. (1987). Curriculum-based measurement: An introduction. Minneapolis: Minneapolis Public Schools, Special Education Department.

Morrison, G. M., Macmillan, L., & Kavale, (1985). System identification of learning disabled children: Implications for research sampling. Learning Disability Quarterly, 8(1), 2-10.

Neisworth, J., & Greer, J. (1975). Functional similarities of learning disabilities and mild retardation. Exceptional Children, 42, 17-21.

Reschly, D. J. Gresham, F. M. and Graham-Clay, S. (1984). Multifactored nonbiased assessment: Convergent and discriminant validity of social education students. Washington, DC United States Department of Education, Grant No. G00 811 01156, Assistance Catalog No. CFDA: 84-023E.

Reschly, D. J., Grimes, J., and Ross-Reynolds, J. (1981). State norms for IQ, adaptive behavior, and sociocultural background: Implications for nonbiased assessment. Des Moines, IA: Iowa Department of Public Instruction.

Reynolds, M., & Wang, M. (1983). Restructuring "special" school programs. Policy Studies Review, 2, 189-212.

Shinn, M. R. (1988). Development of curriculum-based local norms for use in special education decision making. School Psychology Review, 17, 61-80.

Shinn, M. R., & Marston, D. (1985). Differentiating mildly handicapped low achieving, and regular education students: A curriculum-based approach. Remedial and Special Education, 6, 31-38.

Shinn, M. R. Ysseldyke, J. E., Deno, S. L., & Tindal, G. A. (1986). A comparison of differences between students labeled learning disabled and low achieving on measures of classroom performance. Journal of Learning Disabilities, 19(9), 545-552.

Wilson, B. J. (1986). On a hot summer Iowa day. In Timothy Hollinger, Teresa Sagen, & James Reese (Eds.), Perspectives on research: Recent findings and future directions. Des Moines, IA: Iowa Department of Education.

Ysseldyke, J., Algozzine, B., Shinn, M., & McGue, M. (1982). Similarities and difference between underachievers and students classified learning disabled. Journal of Special Education, 16, 73-85.

Ysseldyke, J., & Thurlow, Martha. (1984). Assessment practices in special education: Adequacy and appropriateness. Educational Psychologists, 9(3), 123-136.