DOCUMENT RESUME

ED 228 821 EC 151 679

AUTHOR Skiba, Russell; And Others
TITLE The Effects of Training Teachers in the Use of Formative Evaluation in Reading: An Experimental-Control Comparison.
INSTITUTION Minnesota Univ., Minneapolis. Inst. for Research on Learning Disabilities.
SPONS AGENCY Office of Special Education and Rehabilitative Services (ED), Washington, DC.
REPORT NO IRLD-RR-88
PUB DATE Sep 82
CONTRACT 300-80-0622
NOTE 55p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Academic Achievement; *Disabilities; Elementary Education; *Evaluation Methods; *Formative Evaluation; Intermediate Grades; Program Implementation; *Reading Instruction; *Resource Room Programs; Teaching Methods

ABSTRACT

Twenty pairs of students from seven special education resource rooms (elementary and intermediate grades) were assessed on the structure of the instruction they received. Their reading achievement was measured three times during the year. The instructional programs for 10 experimental students were to be based on their teachers' use of continuous evaluation procedures; these students were assessed also on an accuracy of implementation scale. Results indicated that the procedures were administered and the data charted appropriately, and that teachers considered the procedures time efficient and useful in monitoring student progress. However, the procedures rarely were used to evaluate instructional programs, make instructional changes, or monitor the effectiveness of the changes. No statistically significant differences in instructional structure or achievement were found between experimental and control Ss. (Author/CL)

University of Minnesota

Research Report No. 88

THE EFFECTS OF TRAINING TEACHERS IN THE USE OF FORMATIVE EVALUATION IN READING: AN EXPERIMENTAL-CONTROL COMPARISON

Russell Skiba, Caren Wesson, and Stanley L. Deno

Institute for Research on Learning Disabilities

IRLD

Director: James E. Ysseldyke

The Institute for Research on Learning Disabilities is supported by a contract (300-80-0622) with Special Education Programs, Department of Education. Institute investigators are conducting research on the assessment/decision-making/intervention process as it relates to learning disabled students.

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

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

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

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

Research Report No. 88

THE EFFECTS OF TRAINING TEACHERS IN THE USE OF FORMATIVE EVALUATION IN READING: AN EXPERIMENTAL-CONTROL COMPARISON

Russell Skiba, Caren Wesson, and Stanley L. Deno

Institute for Research on Learning Disabilities

University of Minnesota

September, 1982

Abstract

Twenty pairs of students from seven special education resource rooms were assessed on the structure of the instruction they received and their reading achievement at three times during the year. The instructional programs for 10 experimental students were to be based on their teachers' use of continuous evaluation procedures; these students were assessed also on an accuracy of implementation scale. Results indicated that the procedures were administered and the data charted appropriately, and that teachers considered the procedures time efficient and useful in monitoring student progress. However, the procedures rarely were used to evaluate instructional programs, to make instructional changes, or to monitor the effectiveness of changes. No statistically significant differences in instructional structure or achievement were found between experimental and control group students. The implications of non-implementation of the evaluation aspects of continuous evaluation procedures are discussed.

The Effects of Training Teachers in the Use of Formative Evaluation in Reading: An Experimental-Control Comparison

In recent years, with the advent of Public Law 94-142 and increased public pressure for accountability in education, greater demands have been placed on educators, especially special educators, to be accountable for the quality of instructional decisions and the ways in which they are made. A growing body of evidence (Bohannon, 1975; Crutcher & Hofmeister, 1975; Lovitt, Schaff, & Sayre, 1970; White & Liberty, 1976) suggests that formative evaluation systems may provide viable alternatives to the traditional pre- and post-testing approach to evaluation of academic programs. Such systems provide continuous feedback to both the teacher and student, allowing educators to more closely monitor academic progress.

During the past five years, the Institute for Research on Learning Disabilities at the University of Minnesota, under federal contract, has conducted a number of studies that focus on developing and monitoring progress on IEP goals, as is intended in PL 94-142. The goal of this research has been to determine empirically the effects of using formative evaluation techniques on student achievement in reading, spelling, and written expression.

Earlier research in this area determined what measures of student performance would be ideal for use in a formative evaluation system. The search for these measures began by generating a list of desired characteristics, such as ease of administration, time efficiency, and sensitivity to growth over time (Jenkins, Deno, & Mirkin, 1979). The measures that were not reliable or valid, or those that were deemed

less suitable with respect to any of the other desired characteristics, were eliminated from consideration.

Five reading behaviors were generated from a review of the literature and placed in the original pool for consideration. A series of criterion validity studies (Deno, Mirkin, Shiang, & Lowry, 1980), showed that reading aloud from a basal reader, reading aloud from lists of isolated words, and guessing the words deleted from a reading passage (i.e., cloze comprehension) all related closely to performance on standardized tests and discriminated between program and grade placement. Such formative measures of reading also have shown high test-retest (r = .90) and alternate-forms (rs = .89 - .92) reliability (Shinn, 1981).

Related studies focused on determining the optimal duration of reading measurement and the type of data to record. Results from testing one, two, and three-minute durations indicated that reading proficiency can be indexed validly within one minute and that correct performance is a more valid measure of reading proficiency than error performance (Deno et al., 1980).

Previous studies also assessed the sensitivity of two reading measures, reading isolated word lists and reading aloud from a basal reader. Both reading measures were found to be sensitive to changes within each grade level from fall to spring and across grade levels (Marston, Lowry, Deno, & Mirkin, 1981). However, reading aloud from a basal reader was chosen as the optimal generic measure in reading because it produced a broader range of scores than isolated words, related somewhat more closely to comprehension, and required little

teacher preparation.

Rationale

Given that one-minute timed samples of reading from the curriculum have been shown to be reliable and valid measures of reading growth, there remained the need to test the practicality of such measures and any effects teacher use of such measures might have on student achievement over time. A number of questions related to these issues were posed in the current study.

First, can teachers learn to use the measurement system and will they find it practical and time-efficient? Such measures, however reliable and valid, will come into widespread use in the classroom only if teachers find them to be a savings, rather than an added burden, in time spent in evaluation.

Second, will the record of student growth presented in charts be perceived by teachers of special education as an actual representation of the student's progress in reading? Again, the extent to which teachers believe the information they obtain reflects actual growth will determine the extent to which the instrument is used in the classroom.

Third, once the measurement system is implemented, will resource room teachers use the information it provides to more closely monitor and change the educational program of the student? One of the major advantages of such a system is that it allows for continuous evaluation of the instructional program. Thus, it is critical that the information provided by the system be used.

Only if all these questions are answered affirmatively is it

possible to examine the questions concerning the effectiveness of the system. Two questions were investigated concerning the efficacy of the measures. Given that teachers can learn to use a formative evaluation system for both measurement and evaluation, will the use of such procedures have a direct effect on student achievement? One would expect that the frequent modifications in the instructional plan made possible by continuous feedback would lead to an educational program more sensitive to individual needs and thus more conducive to growth in reading. Second, will the use of such measures have an effect on the structure of the learning environment provided to the student? Because a formative evaluation system provides continuous information about the need for program changes, one might expect the use of such a system to result in a more highly structured learning environment.

Method

Subjects

The subjects were 36 elementary and 4 intermediate students in special education resource classrooms in a rural educational cooperative, representing 20 experimental-control matched pairs. The mean grade level was 3.8, with 18 students in grades 1 through 3, 18 students in grades 4 through 6, and 4 students in grades 7 and 8. All subjects were functioning dramatically below their peers in reading. Thirty of the subjects were boys, ten were girls. All subjects were studied in the resource room setting; their teachers were seven special education resource teachers whose experience ranged from two to six years, with a median of three years teaching special education.

The original pool included 46 subjects (23 matched pairs) in nine resource rooms; however, two of the teachers failed to implement the continuous evaluation procedures and the three pairs in these classrooms thus were dropped from the study.

Procedures

The resource room teachers were trained in the use of the measurement procedures during a series of three half-day workshops at the beginning of the school year. Training was based on the manual, Procedures to Develop and Monitor Progress on IEP Goals (Mirkin, Deno, Fuchs, Wesson, Tindal, Marston, & Kuehnle, 1981). The teachers continued to use the measures over the entire school year. Visits by observers in December, February, and May, and frequent phone contacts, provided feedback to the teachers on the accuracy of their implementation of the measures.

Daily measurement consisted of one-minute timed samples of reading from the student's curriculum. Both words correct and incorrect were scored and charted on equal interval charts. Based on the results of previous research, the placement level for testing, which also became the baseline, was set at a criteria of 20-29 words per minute for grades 1 and 2, and 30-39 words per minute for grades 3 through 8.

Teachers were instructed to write IEP long-range goals using both the entry level criteria and a desired year-end mastery criteria, usually 70 words correct per minute with no more than 7 errors. The formula used in writing the long-range goal is shown in Figure 1.

Insert Figure 1 about here

Short-term objectives were based on the long-range goals (LRG). In order to compute the short-term objective, teachers first subtracted the baseline level of performance from the criterion level listed in the LRG. Dividing this difference by the number of weeks necessary until the annual review, they arrived at the number of words per week gain necessary to meet the long-range goal criteria. The format used for writing short-term objectives is given in Figure 2.

Insert Figure 2 about here

In addition, the teachers also were trained at the beginning of the year, and again at mid-year, in the use of the measurement procedures for evaluation of the instructional program. In order to monitor student growth, the baseline reading level and the long-range goal were connected by an aimline that showed the students' desired progress. Every seven data points, the teachers were to monitor student growth by means of the split-middle or quarter-intersect method (White & Haring, 1980). An example is given in Figure 3. If the student was progressing at a rate equivalent to or greater than that indicated by the aimline, the instructional program was continued; if the projected rate of growth was less than that indicated by the aimline, teachers were directed to make a substantial change in the student's program.

Insert Figure 3 about here

Measures

Four measures were used in collecting data: one each for implementation and structure, and two for achievement. The structure of the learning environment was assessed by means of the Structure of Instruction Rating Scale for both experimental and control subjects. Degree of implementation of the continuous evaluation measures--the treatment for the experimental subjects--was assessed using the Accuracy of Implementation Rating Scale. Achievement measures for both experimental and control subjects consisted of timed samples from three third grade passages, and four subtests of the Stanford Diagnostic Reading Test (SDRT). The three timed samples were collected three times during the year, in November, soon after the teachers had begun implementing the measures, in February, and in May, at the end of the school year. The Stanford Diagnostic Reading Test was administered in May. Descriptions of the measures follow.

Structure of instruction rating scale (SIRS). The Structure of Instruction Rating Scale (SIRS) was designed to measure the degree of structure of the instructional lesson that a student received, in this case in reading. The variables chosen for inclusion on the SIRS were gathered from current literature on instruction and student academic achievement (cf. Stevens & Rosenshine, 1981). A list of the variables and their operational definitions can be found in Appendix A.

The SIRS consists of 12 five-point bipolar rating scales. A

rating of 1 is low for the variable and 5 is high. Observers, trained by videotape to a criteria of .80-.90 inter-rater agreement, rate all variables on the basis of strict definitions at the end of a 20 minute observation period. The focus of each observation period for the SIRS is on the instructional environment for one student at a time. For the current study, nine research assistants were trained as observers and reached an inter-rater agreement level of .92 before actually observing in the classroom.

The reliability of the SIRS was assessed by means of Coefficient Alpha, a measure of internal consistency. For a sample of 70 students observed in November, the average inter-item correlation was .37, resulting in an alpha of .86. Thus, the SIRS seems to have a high degree of reliability as indexed by measures of homogeneity.

Factor analysis of the 12 variables on the SIRS revealed that 9 of the 12 represented one factor. Three variables--Independent Practice, Positive Consequences, and Silent Practice on Outcome Behavior--were not measuring the same factor. Thus, the nine variables were utilized in the data analyses as one factor and the other three variables were analyzed separately.

Accuracy of implementation rating scale (AIRS). The Accuracy of Implementation Rating Scale (AIRS) is an instrument that was developed in conjunction with the manual Procedures to Develop and Monitor Progress on IEP Goals (Mirkin et al., 1981). The AIRS is designed to provide a format by which to monitor the implementation of the procedures described in the manual. The AIRS consists of 12 items that are rated on a 1 to 5 scale, 1 being the lowest implementation

score and 5 being complete and accurate implementation. A copy of the scale and variable definitions is included in Appendix B.

Parts of the scale require direct observation whereas other items on the checklist can be monitored by inspection of student reading graphs and by reading IEP forms. Items 1 and 2 of the AIRS, which require direct observation, deal with the accuracy of administration of the measurement and selection of the stimulus materials. The nine observers trained in the use of the SIRS were trained concurrently on these two items, and observed and rated these variables in the classroom during the same period as they completed the SIRS. Items 3-12 of the AIRS require inspection of various written documents. Specifically, the rater examines the following documents for each student: (a) the IEP, which should specify the long-range goal and short-term objective in reading; (b) the reading graph; (c) the instructional plan for reading; and (d) the record of changes made in the instructional plan in reading. Factors included in items 3-12 pertain to accuracy of establishing: (a) the appropriate measurement level; (b) an adequate baseline; (c) an accurate long-range goal and short-term objective; (d) a detailed graph; (e) a complete instructional program; and (f) the aimline. These items also focus on the timing of instructional changes as well as the types of changes made. Items 3-12 were rated by research assistants. Frequent checks among the four research assistants completing items 3-12 assured inter-rater agreement of at least .80.

Reliability of the AIRS was assessed by means of the Cronbach's Alpha internal consistency measure. The average inter-item

correlation was .12, resulting in an alpha of .62, somewhat lower than the SIRS, but adequate for research purposes.

Results

Implementation

The mean ratings for each variable on the Accuracy of Implementation Rating Scale for all three rounds of data collection are shown in Table 1. For purposes of analyses, the scale was divided into three subscales: initial measurement (procedures necessary for setting up the measurement system and thereafter modified only if in error), ongoing measurement (procedures performed frequently that provide ongoing data), and evaluation (processes that involve evaluating instructional progress and making changes). Scores in the table represent the mean of a five-point rating scale.

Insert Table 1 about here

As can be seen in the table, relatively high scores were obtained in both initial and ongoing measurement. Initial measurement procedures such as setting up the long-range goal and short-term objective reached high levels soon after implementation and maintained an average rating of approximately 3.9 for all three rounds of data collection. Accuracy of ongoing measurement tasks initially was rated somewhat lower ($\bar{X}$ = 3.66) by observers, but increased by round two, and procedures such as administration of the measurement task were the most highly rated by the end of the year ($\bar{X}$ = 4.24).

The ratings for procedures used in evaluating the data and making

instructional changes differed from the other two sets, however. Even though a mid-year training workshop on using the measurement system to make instructional changes was conducted between data collections two and three, the mean ratings of evaluation procedures tended to remain below those of other variables ($\bar{X}$ = 3.6). Higher ratings on the variables relating to substantial and clear changes indicate that when changes were made they were in accordance with the prescriptions of the system. However, only those experimental subjects for whom changes were made could be rated on these variables, and it is clear from the low number of cases for both these variables that this represents a small subset of the subjects.

The low frequency of instructional changes implemented was striking. For the 20 students receiving the experimental treatment, only 27 changes were made in instructional programs, meaning that the average number of changes made in the instructional plan during the entire school year for any individual was 1.4. As is shown in Table 2, no changes in the instructional plan were made for seven of the experimental subjects, and for only three of the subjects were more than two instructional changes made during the course of the school year. It is clear from these data that a potentially important component of the experimental treatment, evaluation of changes in reading instruction, was implemented only partially by the resource room teachers in this study. In fact, making changes according to the data utilization rule (i.e., timing of instructional changes) received one of the lowest mean ratings (3.39). In summary, teachers made few changes in student programs and seldom when the data called for such

changes.

Insert Table 2 about here

Structure of Instruction

The mean ratings for each variable on the Structure of Instruction Rating Scale are reported in Table 3. Classroom structure, as measured by the scale, for the most part remained stable over time. Among the nine variables analyzed as one factor, the moderate to high ratings at all three points in time indicate that these aspects of classroom structure are relatively well established in most resource rooms. It is apparent, however, that the elements of structure represented by Independent Practice, Positive Consequences, and Silent Practice were much less a part of the classrooms studied. Considering the importance placed on token economies and point systems for special populations (Kazdin, 1977; O'Leary, Drabman, & Kass, 1973; Phillips, Phillips, Fixsen, & Wolf, 1971; Staats & Butterfield, 1965), the almost total absence of such systems is especially surprising. The average rating on Positive Consequences in the May observations ($\bar{X}$ = 1.13) means that, of the 38 students observed, only two were observed receiving any form of token economy in their resource room.

Insert Table 3 about here

Although one would not expect any significant differences to emerge between experimental and control subjects on the SIRS since the

treatment was not fully implemented for experimental subjects, mean differences for experimental and control subjects are reported in Table 4. None of the differences were statistically significant. Still, it is worthwhile to note that differences in sample means that do appear between experimental and control subjects are in favor of the experimental subjects on 9 of 12 rated variables.

Insert Table 4 about here

Achievement

The achievement measures showed high degrees of correlation among themselves. The high correlations among the third grade passages (r = .61 to .93, with the majority of coefficients in the high .70s and .80s) suggest that the passages possess a high degree of alternate-forms reliability. The correlations ranging from .47 to .68 between the passages and the Stanford subtests reaffirm the validity of the passage data for purposes of measuring reading growth.

Overall descriptive data for the achievement measures are presented in Tables 5 and 6. As can be seen in Table 5, each of the reading passages showed performance gains over time, although passage 3 proved less sensitive to growth than passage 1 or 2. The average percentile rank for the Structural Analysis subtest of the SDRT was 40.5 (see Table 6); the average percentile rank for the Comprehension subtest of the SDRT was 32.1. To control for grade differences, Z-transformations were performed on all achievement data prior to analysis.

Insert Tables 5 and 6 about here

Again, since the evaluation procedures were not fully implemented for the experimental subjects, one must be cautious in interpreting the results of experimental-control comparisons. Nevertheless the pattern of experimental-control achievement differences (see Tables 7 and 8) is consistent with that obtained for structure. Although none of the mean differences reach the level required for statistical significance, the mean differences for passage data showed gains in favor of the experimental subjects, and three of the four SDRT subtests that exhibited mean differences at the end of the year were in favor of the experimental subjects.

Insert Tables 7 and 8 about here

Useability

At the end of the year, the resource room teachers who participated in the current study were interviewed regarding their reactions to the continuous evaluation procedures they implemented. Preliminary analysis of the interview data indicated that, in general, the teachers reacted positively to the use of formative evaluation. All but one of the teachers interviewed felt that the timed reading sample represented a good measure of reading proficiency. In general, they believed that the slope of the graph reflected the actual rate of reading growth. After a year of measuring, none expressed feelings

that the procedures were too time-consuming.

Other benefits were noted by the teachers as well. One stated that, "the data was more useful to me and the classroom teacher." Another stressed the motivational aspects of the system, stating "It greatly improved the students' confidence which spurred him/her on to improve." Others noted improved reading performance in even those students who failed to reach their reading goal. The only reservation about the continuous evaluation system concerned its relationship to comprehension; one teacher felt that at least one of her students "improved in reading words, but not in comprehension."

Discussion

The current investigation focused on a number of research questions relating to the practicality and efficacy of formative evaluation systems in reading in the resource room. Can teachers learn to use such a measurement system, and will they find it practical, time efficient, and a believable record of student growth? Will they use the continuous information provided by such a system in order to make frequent changes in the educational plan, and monitor the effectiveness of such changes? Does such a system of measurement and evaluation, once implemented, have an effect on the structure of instruction or on reading achievement?

Ratings of the accuracy of implementation indicated that teachers can learn to administer timed reading samples and chart the results in order to provide a continuous record of student growth in reading. A majority of the teachers involved in the project felt that such measures were time efficient and helpful in monitoring student

progress. In addition, all of the teachers had confidence by the end of the year that the student's growth (or lack thereof) in reading was accurately reflected in his or her graph. Thus the teachers can learn to use the measurement system and find it both practical and time efficient in the classroom.

Although able to accurately measure student reading progress, the resource room teachers rarely utilized the procedures to evaluate the instructional program, make instructional changes, and monitor the effectiveness of these changes. Ratings of evaluation procedures tended to be lower than either initial or ongoing measurement variables. More importantly, very few changes were made in the instructional programs of the experimental subjects. Thus, for all intents and purposes, the goal that teachers use the measurement to direct and test changes in student programs was only partially realized.

Why teachers can master a measurement system but not utilize the data for educational planning and change is a matter for some speculation. One explanation is that the technical requirements of the evaluation procedures are more demanding than the initial or ongoing measurement procedures. Teachers must learn to compute slopes in time series data and make judgments regarding program effectiveness. This requires not only new technical skills, but is also time consuming.

A second explanation may lie in the nature of the demands of the measurement and evaluation components. While changes in the procedures used for assessment may be relatively easy for most

teachers to make, the evaluation procedures prescribed by continuous evaluation may require changes in teaching procedures that teachers are unable or unwilling to make in the space of a year. It may well be that special education teachers, accustomed to making few changes in educational plans, find making frequent changes in the instructional plan based on graphed data to be in conflict with their current mode of teaching. Further, the teachers may well be doing all that they know how to do to teach reading and may not have a repertoire of alternative tactics from which to draw. Finally, few prompts or reinforcers may be available for making changes. In fact, making changes may increase work--a punishment contingency.

While it would be unrealistic to expect improved achievement with the experimental treatment only partially in effect, the data provide some basis for the belief that an appropriately implemented continuous evaluation system could have some effect on both achievement and program structure. Although none of the experimental-control differences were reliable, most of the mean differences favored the experimental condition. Such results provide a hopeful sign that continued implementation of such a system, with special attention paid to its evaluative components, might indeed accelerate the rate of learning to read. Also, the reports of the resource teachers support the technical adequacy and the useability of the measures in the classroom.

The results of this study corroborate earlier findings indicating that measuring and graphing a student's performance do not ensure that those data will be used to make instructional decisions. They suggest

that further study is necessary to determine how best to ensure that teachers will use the data in order to adjust instructional programs and maximize growth toward goals in reading.

References

Bohannon, R. Direct and daily measurement procedures in the identification and treatment of reading behaviors of children in special education. Doctoral dissertation, University of Washington, 1975.

Crutcher, C. E., & Hofmeister, A. M. Effective use of objectives and monitoring. Teaching Exceptional Children, 1975, 7(2), 78-80.

Deno, S. L., Mirkin, P. K., Chiang, B., & Lowry, L. Relationships among simple measures of reading and performance on standardized achievement tests (Research Report No. 20). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1980.

Jenkins, J. R., Deno, S. L., & Mirkin, P. K. Measuring pupil progress toward the least restrictive environment (Monograph No. 10). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1979.

Kazdin, A. E. The token economy: A review and evaluation. New York: Plenum, 1977.

Lovitt, T., Schaff, M., & Sayre, E. The use of direct and continuous measurement to evaluate reading materials and procedures. Focus on Exceptional Children, 1970, 2, 1-11.

Marston, D., Lowry, L., Deno, S. L., & Mirkin, P. K. An analysis of learning trends in simple measures of reading, spelling, and written expression: A longitudinal study (Research Report No. 49). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1981.

Mirkin, P., Deno, S., Fuchs, L., Wesson, C., Tindal, G., Marston, D., & Kuehnle, K. Procedures to develop and monitor progress on IEP goals. Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1981.

O'Leary, K. D., Drabman, R., & Kass, R. E. Maintenance of appropriate behavior in a token program. Journal of Abnormal Child Psychology, 1973, 1, 127-138.

Phillips, E. L., Phillips, E. A., Fixsen, D. L., & Wolf, M. M. Achievement place: Modification of the behaviors of pre-delinquent boys within a token economy. Journal of Applied Behavior Analysis, 1971, 4, 45-59.

Shinn, M. A comparison of psychometric and functional differences between students labeled learning disabled and low achieving. Doctoral dissertation, University of Minnesota, 1981.

Staats, A. W., & Butterfield, W. H. Treatment of nonreading in a culturally deprived juvenile delinquent: An application of reinforcement principles. Child Development, 1965, 36, 925-942.

Stevens, R., & Rosenshine, B. Advances in research on teaching. Exceptional Education Quarterly, 1981, 2(1), 1-9.

White, O. R., & Haring, N. G. Exceptional teaching (2nd ed.). Columbus, OH: Charles E. Merrill, 1980.

White, O. R., & Liberty, K. A. Behavioral assessment and precise educational measurement. In N. G. Haring & R. L. Schiefelbusch (Eds.), Teaching special children. New York: McGraw-Hill, 1976.

Table 1

Accuracy of Implementation Rating Scale: Mean Ratings for Experimental Subjects[a]

	November	N of cases	February	N of cases	May	N of cases	X
Initial Measurement							
Sampling for Instructional Level	3.55	(20)	3.75	(20)	3.40	(20)	3.57
Baseline	3.80	(20)	3.50	(20)	3.60	(20)	3.63
Aimline	4.80	(20)	4.55	(20)	4.50	(20)	4.62
Long-Range Goal	4.10	(20)	4.05	(20)	4.25	(20)	4.13
Short-Term Objective	4.00	(20)	4.20	(20)	4.40	(20)	4.20
Instructional Plan	3.30	(20)	3.35	(20)	3.45	(20)	3.33
Mean	3.93		3.90		3.93		3.92
Ongoing Measurement							
Administering the Measurement Task	3.73	(19)	4.50	(20)	4.71	(17)	4.31
Selecting the Stimulus Material	3.95	(19)	4.50	(20)	4.94	(17)	4.46
Graph Set-up	3.30	(20)	3.30	(20)	3.30	(17)	3.30
Mean	3.66		4.10		4.24		4.02
Evaluation							
Timing of Instructional Changes	3.33	(15)	3.25	(20)	3.60	(20)	3.39
Substantial Changes	4.20	(2)	3.33	(9)	3.88	(8)	3.90
Clear Change	3.50	(2)	3.77	(9)	4.00	(8)	3.57
Mean	3.78		3.45		3.83		3.62
AIRS Scale Mean							
All Variables	3.79		3.82		4.00		3.85

[a]Ratings are on a scale from 1 (low) to 5 (high).

Table 2

Number of Changes Made in the Instructional Program During the School Year

	No Changes	One Change	Two Changes	Three Changes	Four Changes
No. of Experimental Subjects	7	6	5	1	2

Table 3

Structure of Instruction Rating Scale: Mean Ratings for Experimental Subjects[a]

	November	(N)	February	(N)	May	(N)	X
Instructional Grouping	3.43	(37)	2.95	(40)	3.29	(38)	3.22
Teacher-Directed Learning	4.08	(37)	4.24	(38)	4.08	(38)	4.13
Active Academic Responding	4.16	(37)	4.15	(40)	3.92	(38)	4.08
Demonstration and Prompting	3.24	(37)	3.55	(40)	3.50	(38)	3.43
Controlled Practice	3.57	(37)	3.26	(39)	3.43	(37)	3.42
Frequency of Correct Answers	3.78	(36)	4.44	(36)	4.23	(35)	4.15
Corrections	4.03	(36)	4.51	(37)	4.31	(36)	4.28
Pacing	3.57	(37)	3.59	(39)	3.72	(36)	3.63
Oral Practice	--	--	2.23	(40)	2.90	(38)	2.57
Mean	3.73		3.66		3.71		3.66
Independent Practice	2.17	(18)	1.41	(22)	1.56	(18)	1.71
Positive Consequences	1.62	(37)	1.51	(40)	1.13	(38)	1.42
Silent Practice	--	--	1.98	(40)	1.92	(38)	1.95
Mean	1.90		1.63		1.54		1.69

[a]Ratings are on a scale from 1 (low) to 5 (high).

Table 4

Structure of Instruction Rating Scale: Experimental - Control Means

	November		February		May	
	Experimental	Control	Experimental	Control	Experimental	Control
Instructional Grouping	3.76	3.06	2.95	2.95	3.39	3.11
Teacher-Directed Learning	4.29	3.82	4.50	4.06	4.28	3.89
Active Academic Responding	4.18	4.18	4.15	4.15	4.22	3.61
Demonstration and Prompting	3.18	3.24	3.70	3.40	3.56	3.33
Controlled Practice	3.47	3.47	3.05	3.47	3.47	3.29
Frequency of Correct Answers	3.69	3.81	4.53	4.35	4.38	4.13
Corrections	3.94	3.94	4.71	4.59	4.35	4.17
Pacing	3.64	3.41	3.74	3.58	3.93	3.44
Oral Practice	--	--	2.15	2.30	3.00	2.89
Mean SIRS Rating	3.77	3.61	3.64	3.54	3.77	3.46
Independent Practice	2.00	2.00	1.17	1.67	1.60	1.60
Positive Consequences	1.76	1.53	1.20	1.10	1.11	1.17
Silent Practice	--	--	1.95	2.00	1.89	2.00

Table 5

Third Grade Passage Data: Means and Standard Deviations for All Subjects

	November		February		May	
	Mean	(SD)	Mean	(SD)	Mean	(SD)
Words Read Correctly in One Minute						
Passage 1	32.74	(24.40)	44.10	(25.49)	62.65	(34.37)
Passage 2	32.28	(22.67)	45.92	(26.98)	65.83	(34.27)
Passage 3	23.51	(18.26)	32.38	(19.61)	47.98	(28.00)
Mean	29.51	(21.26)	40.80	(23.23)	58.82	(31.60)
Words Read Incorrectly in One Minute						
Passage 1	8.86	(4.07)	8.34	(4.92)	6.30	(4.95)
Passage 2	8.45	(3.69)	8.21	(5.31)	6.23	(5.45)
Passage 3	9.93	(5.09)	8.90	(4.34)	7.80	(5.50)
Mean	9.08	(3.94)	8.48	(4.45)	6.78	(4.96)

Table 6

Stanford Diagnostic Reading Test: Means and Standard Deviations for All Subjects by Grade

Grade	N*	Mean	(Standard Deviation)	Mean % Rank
Structural Analysis Raw Score (Word Division & Word Blending)				
1		19.50	(2.12)	*
2		28.75	(10.00)	*
3		40.13	(10.26)	48.0
4		39.60	(10.61)	31.8
5		52.00	(4.97)	47.3
6		44.00	(18.22)	*
7		57.67	(1.15)	*
8		56.00	[one case]	*
Total Comprehension Raw Score (Literal & Inferential Comprehension				
1		18.00	(5.66)	*
2		27.38	(8.96)	*
3		36.50	(12.20)	39.5
4		38.70	(13.23)	29.3
5		46.75	(3.30)	24.5
6		47.00	(20.70)	*
7		52.67	(5.03)	*
8		57.00	[one case]	*

*Norms not available.

Table 7

Third Grade Passage Data: Experimental - Control Mean Comparisons

	November		February		May	
	Experimental	Control	Experimental	Control	Experimental	Control
Words Correct Per Minute						
Passage 1	32.84	30.89	45.42	41.37	64.90	60.40
Passage 2	32.05	31.05	46.37	44.10	67.50	64.15
Passage 3	23.11	22.94	32.95	30.37	48.70	47.25
Mean	29.33	28.29	41.58	38.61	60.37	57.27
Errors in One Minute						
Passage 1	8.14	9.57	7.21	10.00	5.95	6.65
Passage 2	8.14	8.78	7.50	9.43	6.25	6.20
Passage 3	10.29	9.79	8.50	9.71	7.10	8.50
Mean	8.86	9.38	7.74	9.71	6.43	7.11

Table 8

Stanford Diagnostic Reading Test:

Experimental-Control Mean Comparisons[a]

Subtest	May	
	Experimental	Control
Word Division	20.75	19.70
Word Blending	20.55	18.95
Structural Analysis	41.30	38.65
Literal Comprehension	20.85	20.65
Inferential Comprehension	17.60	17.80
Total Comprehension	37.75	38.45

[a]Entries in tables are mean raw scores.

	Condition	Behavior	Criteria
LRG:	In _____ (total # weeks) weeks, when presented with stories from Level _____ (#), _____ (reading series),	student will read aloud	at the rate of 50 wpm or better 5 or fewer errors.

Figure 1. Format for Long-Range Goal: Reading

	Condition	Behavior	Criteria
STO	Each successive week, when presented with a random selection from Level _____ (#) of _____ (reading series),	student will read aloud	at an average increase of _____ (repeated-actual performance/total # weeks) words correct/ minute and no increase in errors.

Figure 2. Format for Short-Term Objective: Reading

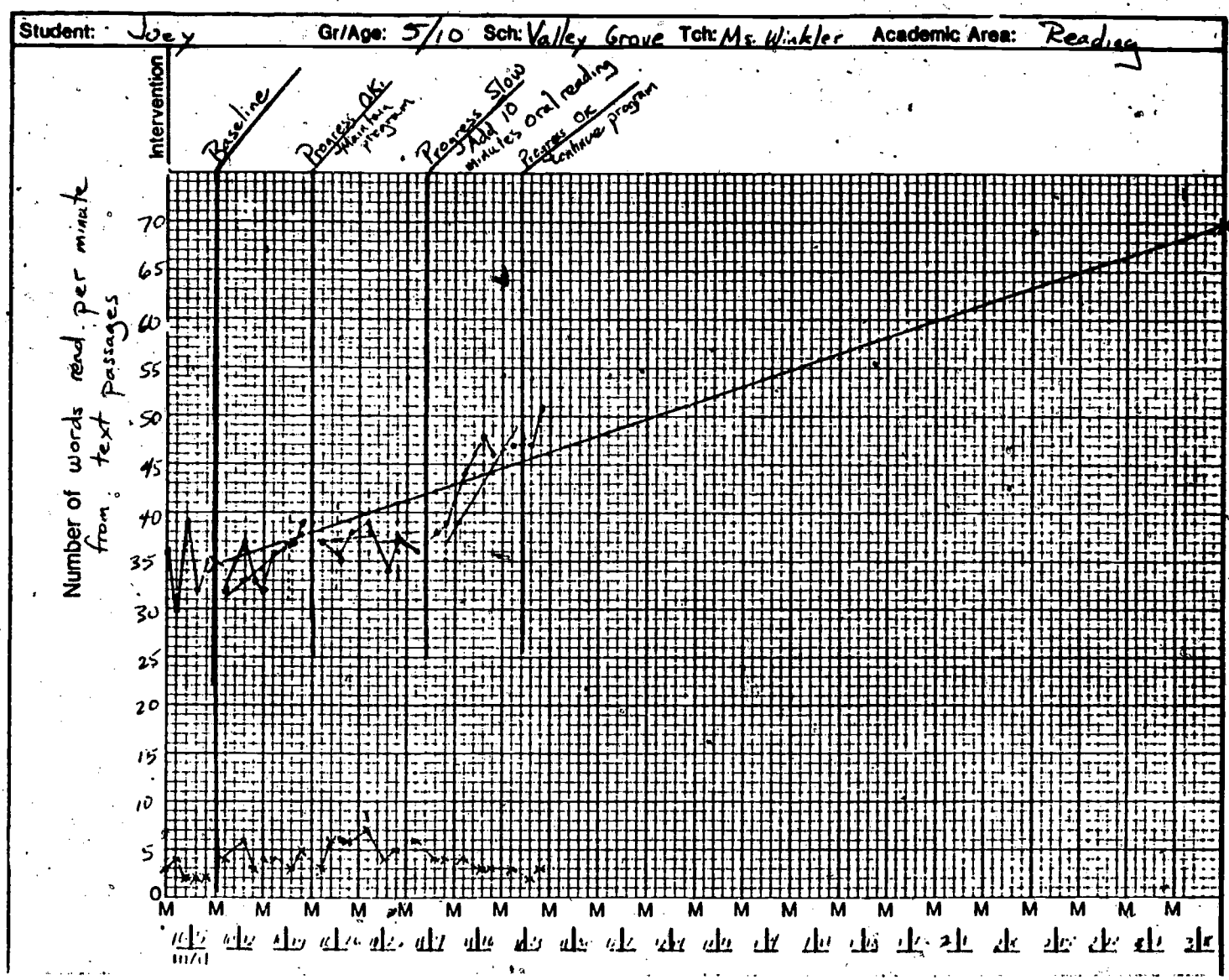

Figure 3. Using the split-middle technique to monitor student progress.

Appendix A

Structure of Instruction Rating Scale (SIRS)

School:_____ Student:_____

Date:_____ Teacher:_____

Observer:_____ Number of Students in Group:_____

Number of observations prior to rating:_____

Time observation begins:_____ Time observation ends:_____

Time allocated to reading instruction per day:_____

Curriculum used for instruction: Publisher_____

Series_____ Level_____

Instructions

Circle the number that accurately reflects your rating for each variable. Only one number may be circled per variable. If you are unable to evaluate a certain variable, mark N/A (not applicable) next to the left-hand column.

1.	Instructional Grouping	1	2	3	4	5
2.	Teacher-directed Learning	1	2	3	4	5
3.	Active Academic Responding	1	2	3	4	5
4.	Demonstration/Prompting	1	2	3	4	5
5.	Controlled Practice	1	2	3	4	5
6.	Frequency of Correct Answers	1	2	3	4	5
7.	Independent Practice	1	2	3	4	5
8.	Corrections	1	2	3	4	5
9.	Positive Consequences	1	2	3	4	5
10.	Pacing	1	2	3	4	5
11.	Oral Practice on Outcome Behavior	1	2	3	4	5
12.	Silent Practice on Outcome Behavior	1	2	3	4	5

SIRS

Operational Definitions Codebook

1. Instructional Grouping

   5 - 90% or more of the instruction this student receives from the teacher is on an individual basis.

   1 - 10% or less of the instruction this student receives from the teacher is on an individual basis.

2. Teacher-Directed Learning

   5 - Student's instruction is extremely organized, businesslike, and teacher is firm in direction and control of activities. For example, student is presented with questions, student has material to cover, etc.

   1 - Student's instruction is casually organized and very spontaneous. Teacher is not committed to having the student work on a particular set of material. Instructional materials do not determine what activities student engages in and the lessons change according to problems or mood of this student.

3. Active Academic Responding

   5 - The student is actively practicing the academic skills to be learned more than 75% of the time observed. Specifically, the student is engaged in oral or written responding to teacher questions or written material, e.g., reading aloud, answering questions, writing, or computing. Student rarely is involved in non-academic conversations with teacher or other students. Attending to the lesson without responding, such as sitting, looking, listening, and/or following along in a book does not apply. The student must make an active, written or oral response.

   1 - The student is actively practicing the skills to be learned less than 10% of the time observed. Instructional lessons may be interrupted or shortened to include "process" and other non-academic activities, e.g., clarifying feelings, opinions, and working on arts and crafts.

4. Demonstration and Prompting

   5 - Appropriate steps of the desired behavior to be performed are demonstrated for the student. Student is given an opportunity to practice the step(s) as teacher provides prompts for correct behavior that approximates or achives desired response.

   1 - Teacher attempts to teach the student a behavior without using demonstration and prompting techniques.

SIRS

5. Controlled Practice

5 - Student's practice of material is actively controlled by teacher who frequently asks questions to clarify that the student understands what has just been demonstrated. Questions are convergent (single factual answer) and the student's answers consistently follow the questions and are given teacher feedback.

1 - Student is rarely questioned by teacher following demonstration of new materials. Questions are more divergent (open-ended, several interpretations) than convergent (single factual answer). Student's response is not consistently followed by teacher feedback. The type of questions are such that several answers are acceptable, i.e., questions are abstract or ambiguous.

Examples:

If during an oral reading session:

a) the teacher frequently attempts to clarify the material with convergent questions ("what color hat was John wearing?"), a 5 would be recorded.

b) the teacher asks few questions, most of which are divergent ("What do you think this means?"), a 1 would be recorded.

c) the teacher asks few convergent questions or many divergent questions, the appropriate rating would be a 3.

6. Frequency of Correct Answers

5 - Academic lessons are conducted in such a way that the difficulty of the material allows the student to achieve mean accuracy of 80% or higher.

1 - Academic material is difficult for student, component steps are large or unsequenced, and mean accuracy for student is less than 55%.

(Note: If the student has no opportunity for oral or written response during the observational period, item 6 would be rated N/A - not applicable, while items 3 and 5 would most likely be rated 1).

7. Independent Practice

5 - When engaged in independent seatwork, the student frequently is monitored by the teacher who assists, clarifies, and praises the student for academic engaged tasks.

(Note: Independent seatwork is defined here as a student working on an assigned task for at least 5 minutes. [If no such 5-minute block of time is observed, Item 7 is rated N/A].)

SIRS

1 - When student is engaged in academic seat-work activities, little attention is given by teacher who directs seat-work activities from a distance or engages in work separate from the assigned seat work. Teacher is generally not helpful or supportive to student during independent practice time.

8. Corrections

5 - The student's errors are consistently corrected by the teacher. When the student either does not respond, responds incorrectly, or does not respond in unison if the activity is group directed and requires such responding, the teacher will systematically attempt to correct the student by asking a simpler question, refocusing student's attention to elicit correct response from the student or provide general rules by which to determine the correct answer 90% or more of the time.

1 - Student's errors are rarely and inconsistently corrected by the teacher. The student responses are not systematically corrected. Student's errors are corrected 50% or less of the time.

For example: In oral reading this includes teacher correction of skips and mispronunciations, or help in sounding out hesitations.

9. Positive Consequences

5 - Positive events (tokens, points, activities, etc.) are given to the student when performing the desired behavior. When learning a new skill the student receives positive consequence for approximations of the desired behavior. Consequences are consistently received during academic training time. Praise and compliments, e.g., "good working, nice job," are not included in this definition.

1 - Student rarely receives positive consequences for academic work. When student receives consequences they usually are for social behavior, rather than for behaviors occurring under systematic academic training.

10. Pacing

5 - The pace of the lesson is rapid, providing many opportunities for response by the student. As a result, attention is high and off-task behavior is low.

1 - The pace of the lesson is slow and the student's rate of responding is low. Lesson format frequently varies, is not highly structured, and student attention may be low.

SIRS

11. Oral Practice on Outcome Behavior

5 - Student reads aloud from context nearly all the time (85-100% or 12-15 min. of a 15 min. observation).

1 - Student does not read aloud during the observation (0% of the time).

(Note: Reading aloud for measurement purposes should not be considered when rating this variable. Reading in context is defined as reading phrases, sentences, paragraphs, or story selections.)

Examples:

If the student is reading isolated words nearly the entire time, the appropriate rating is a 3.

If the student is reading aloud from a text about half the time, a 3 would be recorded.

12. Silent Practice on Outcome Behavior

5 - Student reads silently from context nearly all the time (85-100% or 12-15 min. of a 15 min. observation).

1 - Student does not read silently during the observation (0% of the time).

(Note: Reading in context is defined as the same as #11. The examples of #11 are the same for #12, with silent reading.)

Appendix B

Accuracy of Implementation Rating Scale

School:_____ Student:_____

Date:_____ Teacher:_____

Observer (Items 1 and 2):_____

Rater (Items 3-13):_____

Number of observations prior to rating:_____

Time observation begins:_____ Time observation ends:_____

Time allocated to reading instruction per day:_____

Curriculum used for measurement: Publisher_____

Series_____ Level_____

Instructions

Circle the number that accurately reflects your rating for each variable. Only one number may be circled per variable. 1 reflects a low level of implementation and 5 means total implementation of the Procedures to Develop and Monitor Progress on IEP Goals. See Operational Definitions. Items 1 and 2 require direct observation of the measurement administration. Items 3, 4, 5, 6, and 7 require inspection of the student graph. Items 8, 9, and 10 require inspection of the student's IEP form. The Instructional Plan must be inspected to rate item 11. The Change Record must be inspected to rate items 12 and 13.

Dir. Obs.	1. Administering the Measurement Task	1	2	3	4	5
	2. Selecting the Stimulus Material	1	2	3	4	5
Inspect Graph	3. Sampling for Instructional Level	1	2	3	4	5
	4. Baseline	1	2	3	4	5
	5. Graph Set-up	1	2	3	4	5
	6. Aimline	1	2	3	4	5
	7. Timing of Instructional Changes	1	2	3	4	5
Inspect IEP	8. Long-Range Goal	1	2	3	4	5
	9. Short-Term Objective	1	2	3	4	5
	10. Measurement System	1	2	3	4	5
Insp. Inst. Plan	11. Instructional Plan	1	2	3	4	5
Insp. Change Rec.	12. Substantial Changes	1	2	3	4	5
	13. One, Clear Change	1	2	3	4	5

AIRS

Operational Definitions

Accuracy of Implementation Rating Scale

1. Administering the Measurement Task

   5 - The measurement task is administered correctly: teacher brings stopwatch and pencil to measurement area; gives correct directions for the task; administers the measurement procedure for one minute; correctly marks the teacher copy; correctly counts words correct and incorrect; correctly counts words correct and incorrect; correctly plots the data point.

   1 - The teacher: forgets necessary materials; does not give directions; does not time the task accurately; fails to mark the teacher copy or incorrectly marks errors; miscounts correct and incorrect words; and inaccurately plots the data point.

2. Selecting the Stimulus Material

   5 - The teacher has followed these procedures: Uses passages selected from the level that represents the annual goal. Observers should record the book from which the passage was selected and later check this with the long-range goal level. At this level find the pages in these stories that do not have excessive dialogue, indentations, and/or unusual pronouns. Write these page numbers on equal size slips of paper.

   - Put the slips of paper into a drawbag and shake it.

   - Randomly pick a slip of paper.

   - The page number chosen is the page where the student begins reading. If the page chosen is a passage that was read earlier during the week, draw another page number.

   Other completely random procedures are also rated a 5. If, however, not all passages have an equal chance of being selected, a 4 rating would be indicated.

   1 - The teacher fails to randomly pick the passage or the sample is taken from a domain which is greater or smaller than the one indicated in the goal.

3. Sampling for Instructional Level

   5 - The teacher has sampled from higher or lower reading levels to find the level in which the student reads 20-29 wpm (grades 1 & 2) or 30-39 wpm (grades 3 and up).

1 - The teacher is measuring at a level which is too high or too low.

4. Baseline

5 - The student's performance has been measured at least 3 times to establish a stable baseline. A stable baseline means that all data points fall within a range of 10.

1 - The teacher has not found a level for which a stable baseline has been established or has failed to collect 3 data points during the baseline phase.

5. Graph Set-Up

5 - The graph is accurately set up: The dates filled in on the horizontal axis; the vertical axis is correctly labeled words read per minute from _____ material; the units of measurement are specified; the student's name and subject area are certified; a key identifies the symbols for correct (.) and incorrect (x); symbols are placed at the intersection of date and score; the data points are connected with straight lines; and absences are recorded on the graph as (abs.).

1 - The graph does not include many of the items mentioned above.

6. Aimline

5 - The long-range goal is marked on the graph with an X at the intersection of the desired performance level and date of attainment and a line of desired progress connects the point representing the student's median score of the last 3 data points from baseline and the LRG.

1 - The long-range goal is not marked on the graph and/or the median and LRG are not connected.

7. Timing of Instructional Changes

5 - All the adjustments in the student's program are made at the appropriate time given the rules for data utilization:

(1) Compare the actual slope based on 7 to 10 data points to the slope required to attain the Annual Goal.

(2) If the actual slope is equal to, or steeper than, the Annual Goal slope, continue the program.

(3) If the actual slope is flatter than the Annual Goal slope, change the program.

1 - None of the adjustments in the student's program are made at the appropriate time.

8. Long-Range Goal

5 - The long-range goal is accurately written; goal specifies the number of weeks until next review; stimulus materials for the goal represents the level in which the student is performing at entry level criterion; goal specifies student behavior; goal specifies mastery criterion of 50 wpm with fewer than 5 errors (grades 1 & 2) or 70 wpm with fewer than 7 errors (grades 3-5) when there are 36 weeks until the annual review. If there are fewer than 36 weeks, the criteria can be lowered proportionately.

1 - The long-range goal contains none of the above criteria.

9. Short-Term Objective

5 - The short-term objective is accurately written; stimulus material and behavior is specified; and the average increase in performance is the desired performance minus the actual performance divided by the number of weeks until the annual review.

1 - The short-term objective contains none of the above criteria.

10. Measurement System

5 - The teacher has indicated how the material is organized, the frequency of measurement, and what is to be recorded on the graph.

1 - The measurement system is not specified.

11. Instructional Plan

5 - The instructional plan includes clear and specific descriptions of the instructional procedures, the time spent in each activity, the pertinent materials, the arrangements, and the motivational strategies.

1 - The instructional plan is unclear and lacks specific descriptions of the instructional procedures, the time spent in each activity, the pertinent materials, the arrangements, and the motivational strategies.

12. Substantial Changes

5 - The adjustments in the student's program are always substantial (have a good chance of being effective; see Unit XIV).

1 - The adjustments are never substantial.

13. Clear Change

5 - All the adjustments made introduce only one, clear program change.

1 - All the adjustments made introduce more than one change and/or the change is unclear.

PUBLICATIONS

Institute for Research on Learning Disabilities
University of Minnesota

The Institute is not funded for the distribution of its publications. Publications may be obtained for $3.00 per document, a fee designed to cover printing and postage costs. Only checks and money orders payable to the University of Minnesota can be accepted. All orders must be prepaid.

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

Ysseldyke, J. E. Assessing the learning disabled youngster: The state of the art (Research Report No. 1). November, 1977.

Ysseldyke, J. E., & Regan, R. R. Nondiscriminatory assessment and decision making (Monograph No. 7). February, 1979.

Foster, G., Algozzine, B., & Ysseldyke, J. Susceptibility to stereotypic bias (Research Report No. 3). March, 1979.

Algozzine, B. An analysis of the disturbingness and acceptability of behaviors as a function of diagnostic label (Research Report No. 4). March, 1979.

Algozzine, B., & McGraw, K. Diagnostic testing in mathematics: An extension of the PIAT? (Research Report No. 5). March, 1979.

Deno, S. L. A direct observation approach to measuring classroom behavior: Procedures and application (Research Report No. 6). April, 1979.

Ysseldyke, J. E., & Mirkin, P. K. Proceedings of the Minnesota round-table conference on assessment of learning disabled children (Monograph No. 8). April, 1979.

Somwaru, J. P. A new approach to the assessment of learning disabilities (Monograph No. 9). April, 1979.

Algozzine, B., Forgnone, C., Mercer, C. D., & Trifiletti, J. J. Toward defining discrepancies for specific learning disabilities: An analysis and alternatives (Research Report No. 7). June, 1979.

Algozzine, B. The disturbing child: A validation report (Research Report No. 8). June, 1979.

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

Ysseldyke, J. E., Algozzine, B., Regan, R., & Potter, M. Technical adequacy of tests used by professionals in simulated decision making (Research Report No. 9). July, 1979.

Jenkins, J. R., Deno, S. L., & Mirkin, P. K. Measuring pupil progress toward the least restrictive environment (Monograph No. 10). August, 1979.

Mirkin, P. K., & Deno, S. L. Formative evaluation in the classroom: An approach to improving instruction (Research Report No. 10). August, 1979.

Thurlow, M. L., & Ysseldyke, J. E. Current assessment and decision-making practices in model programs for the learning disabled (Research Report No. 11). August, 1979.

Deno, S. L., Chiang, B., Tindal, G., & Blackburn, M. Experimental analysis of program components: An approach to research in CSDC's (Research Report No. 12). August, 1979.

Ysseldyke, J. E., Algozzine, B., Shinn, M., & McGue, M. Similarities and differences between underachievers and students labeled learning disabled: Identical twins with different mothers (Research Report No. 13). September, 1979.

Ysseldyke, J., & Algozzine, R. Perspectives on assessment of learning disabled students (Monograph No. 11). October, 1979.

Poland, S. F., Ysseldyke, J. E., Thurlow, M. L., & Mirkin, P. K. Current assessment and decision-making practices in school settings as reported by directors of special education (Research Report No. 14). November, 1979.

McGue, M., Shinn, M., & Ysseldyke, J. Validity of the Woodcock-Johnson psycho-educational battery with learning disabled students (Research Report No. 15). November, 1979.

Deno, S., Mirkin, P., & Shinn, M. Behavioral perspectives on the assessment of learning disabled children (Monograph No. 12). November, 1979.

Sutherland, J. H., Algozzine, B., Ysseldyke, J. E., & Young, S. What can I say after I say LD? (Research Report No. 16). December, 1979.

Deno, S. L., & Mirkin, P. K. Data-based IEP development: An approach to substantive compliance (Monograph No. 13). December, 1979.

Ysseldyke, J., Algozzine, B., Regan, R., & McGue, M. The influence of test scores and naturally-occurring pupil characteristics on psycho-educational decision making with children (Research Report No. 17). December, 1979.

Algozzine, B., & Ysseldyke, J. E. Decision makers' prediction of students' academic difficulties as a function of referral information (Research Report No. 18). December, 1979.

Ysseldyke, J. E., & Algozzine, B. Diagnostic classification decisions as a function of referral information (Research Report No. 19). January, 1980.

Deno, S. L., Mirkin, P. K., Chiang, B., & Lowry, L. Relationships among simple measures of reading and performance on standardized achievement tests (Research Report No. 20). January, 1980.

Deno, S. L., Mirkin, P. K., Lowry, L., & Kuehnle, K. Relationships among simple measures of spelling and performance on standardized achievement tests (Research Report No. 21). January, 1980.

Deno, S. L., Mirkin, P. K., & Marston, D. Relationships among simple measures of written expression and performance on standardized achievement tests (Research Report No. 22). January, 1980.

Mirkin, P. K., Deno, S. L., Tindal, G., & Kuehnle, K. Formative evaluation: Continued development of data utilization systems (Research Report No. 23). January, 1980.

Deno, S. L., Mirkin, P. K., Robinson, S., & Evans, P. Relationships among classroom observations of social adjustment and sociometric rating scales (Research Report No. 24). January, 1980.

Thurlow, M. L., & Ysseldyke, J. E. Factors influential on the psychoeducational decisions reached by teams of educators (Research Report No. 25). February, 1980.

Ysseldyke, J. E., & Algozzine, B. Diagnostic decision making in individuals susceptible to biasing information presented in the referral case folder (Research Report No. 26). March, 1980.

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

Ysseldyke, J. E., Regan, R. R., & Schwartz, S. Z. The use of technically adequate tests in psychoeducational decision making (Research Report No. 28). April, 1980.

Richey, L., Potter, M., & Ysseldyke, J. Teachers' expectations for the siblings of learning disabled and non-learning disabled students: A pilot study (Research Report No. 29). May, 1980.

Thurlow, M. L., & Ysseldyke, J. E. Instructional planning: Information collected by school psychologists vs. information considered useful by teachers (Research Report No. 30). June, 1980.

Algozzine, B., Webber, J., Campbell, M., Moore, S., & Gilliam, J. Classroom decision making as a function of diagnostic labels and perceived competence (Research Report No. 31). June, 1980.

Ysseldyke, J. E., Algozzine, B., Regan, R. R., Potter, M., Richey, L., & Thurlow, M. L. Psychoeducational assessment and decision making: A computer-simulated investigation (Research Report No. 32). July, 1980.

Ysseldyke, J. E., Algozzine, B., Regan, R. R., Potter, M., & Richey, L. Psychoeducational assessment and decision making: Individual case studies (Research Report No. 33). July, 1980.

Ysseldyke, J. E., Algozzine, B., Regan, R., Potter, M., & Richey, L. Technical supplement for computer-simulated investigations of the psychoeducational assessment and decision-making process (Research Report No. 34). July, 1980.

Algozzine, B., Stevens, L., Costello, C., Beattie, J., & Schmid, R. Classroom perspectives of LD and other special education teachers (Research Report No. 35). July, 1980.

Algozzine, B., Siders, J., Siders, J., & Beattie, J. Using assessment information to plan reading instructional programs: Error analysis and word attack skills (Monograph No. 14). July, 1980.

Ysseldyke, J., Shinn, M., & Epps, S. A comparison of the WISC-R and the Woodcock-Johnson Tests of Cognitive Ability (Research Report No. 36). July, 1980.

Algozzine, B., & Ysseldyke, J. E. An analysis of difference score reliabilities on three measures with a sample of low achieving youngsters (Research Report No. 37). August, 1980.

Shinn, M., Algozzine, B., Marston, D., & Ysseldyke, J. A theoretical analysis of the performance of learning disabled students on the Woodcock-Johnson Psycho-Educational Battery (Research Report No. 38). August, 1980.

Richey, L. S., Ysseldyke, J., Potter, M., Regan, R. R., & Greener, J. Teachers' attitudes and expectations for siblings of learning disabled children (Research Report No. 39). August, 1980.

Ysseldyke, J. E., Algozzine, B., & Thurlow, M. L. (Eds.). A naturalistic investigation of special education team meetings (Research Report No. 40). August, 1980.

Meyers, B., Meyers, J., & Deno, S. Formative evaluation and teacher decision making: A follow-up investigation (Research Report No. 41). September, 1980.

Fuchs, D., Garwick, D. R., Featherstone, N., & Fuchs, L. S. On the determinants and prediction of handicapped children's differential test performance with familiar and unfamiliar examiners (Research Report No. 42). September, 1980.

Algozzine, B., & Stoller, L. Effects of labels and competence on teachers' attributions for a student (Research Report No. 43). September, 1980.

Ysseldyke, J. E., & Thurlow, M. L. (Eds.). The special education assessment and decision-making process: Seven case studies (Research Report No. 44). September, 1980.

Ysseldyke, J. E., Algozzine, B., Potter, M., & Regan, R. A descriptive study of students enrolled in a program for the severely learning disabled (Research Report No. 45). September, 1980.

Marston, D. Analysis of subtest scatter on the tests of cognitive ability from the Woodcock-Johnson Psycho-Educational Battery (Research Report No. 46). October, 1980.

Algozzine, B., Ysseldyke, J. E., & Shinn, M. Identifying children with learning disabilities: When is a discrepancy severe? (Research Report No. 47). November, 1980.

Fuchs, L., Tindal, J., & Deno, S. Effects of varying item domain and sample duration on technical characteristics of daily measures in reading (Research Report No. 48). January, 1981.

Marston, D., Lowry, L., Deno, S., & Mirkin, P. An analysis of learning trends in simple measures of reading, spelling, and written expression: A longitudinal study (Research Report No. 49). January, 1981.

Marston, D., & Deno, S. The reliability of simple, direct measures of written expression (Research Report No. 50). January, 1981.

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

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

Evans, P. R., & Peham, M. A. S. Testing and measurement in occupational therapy. A review of current practice with special emphasis on the Southern California Sensory Integration Tests (Monograph No. 15). April, 1981.

Fuchs, L., Wesson, C., Tindal, G., & Mirkin, P. Teacher efficiency in continuous evaluation of IEP goals (Research Report No. 53). June, 1981.

Fuchs, D., Featherstone, N., Garwick, D. R., & Fuchs, L. S. The importance of situational factors and task demands to handicapped children's test performance (Research Report No. 54). June, 1981.

Tindal, G., & Deno, S. L. Daily measurement of reading: Effects of varying the size of the item pool (Research Report No. 55). July, 1981.

Fuchs, L. S., & Deno, S. L. A comparison of teacher judgment, standardized tests, and curriculum-based approaches to reading placement (Research Report No. 56). August, 1981.

Fuchs, L., & Deno, S. The relationship between curriculum-based mastery measures and standardized achievement tests in reading (Research Report No. 57). August, 1981.

Christenson, S., Graden, J., Potter, M., & Ysseldyke, J. Current research on psychoeducational assessment and decision making: Implications for training and practice (Monograph No. 16). September, 1981.

Christenson, S., Ysseldyke, J., & Algozzine, B. Institutional constraints and external pressures influencing referral decisions (Research Report No. 58). October, 1981.

Fuchs, L., Fuchs, D., & Deno, S. Reliability and validity of curriculum-based informal reading inventories (Research Report No. 59). October, 1981.

Algozzine, B., Christenson, S., & Ysseldyke, J. Probabilities associated with the referral-to-placement process (Research Report No. 60). November, 1981.

Tindal, G., Fuchs, L. Christenson, S., Mirkin, P., & Deno, S. The relationship between student achievement and teacher assessment of short- or long-term goals (Research Report No. 61). November, 1981.

Mirkin, P., Fuchs, L., Tindal, G., Christenson, S., & Deno, S. The effect of IEP monitoring strategies on teacher behavior (Research Report No. 62). December, 1981.

Wesson, C., Mirkin, P., & Deno, S. Teachers' use of self instructional materials for learning procedures for developing and monitoring progress on IEP goals (Research Report No. 63). January, 1982.

Fuchs, L., Wesson, C., Tindal, G., Mirkin, P., & Deno, S. Instructional changes, student performance, and teacher preferences: The effects of specific measurement and evaluation procedures (Research Report No. 64). January, 1982.

Potter, M., & Mirkin, P. Instructional planning and implementation practices of elementary and secondary resource room teachers: Is there a difference? (Research Report No. 65). January, 1982.

Thurlow, M. L., & Ysseldyke, J. E. Teachers' beliefs about LD students (Research Report No. 66). January, 1982.

Graden, J., Thurlow, M. L., & Ysseldyke, J. E. Academic engaged time and its relationship to learning: A review of the literature (Monograph No. 17). January, 1982.

King, R., Wesson, C., & Deno, S. Direct and frequent measurement of student performance: Does it take too much time? (Research Report No. 67). February, 1982.

Greener, J. W., & Thurlow, M. L. Teacher opinions about professional education training programs (Research Report No. 68). March, 1982.

Algozzine, B., & Ysseldyke, J. Learning disabilities as a subset of school failure: The oversophistication of a concept (Research Report No. 69). March, 1982.

Fuchs, D., Zern, D. S., & Fuchs, L. S. A microanalysis of participant behavior in familiar and unfamiliar test conditions (Research Report No. 70). March, 1982.

Shinn, M. R., Ysseldyke, J., Deno, S., & Tindal, G. A comparison of psychometric and functional differences between students labeled learning disabled and low achieving (Research Report No. 71). March, 1982.

Thurlow, M. L. Graden, J., Greener, J. W., & Ysseldyke, J. E. Academic responding time for LD and non-LD students (Research Report No. 72). April, 1982.

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

Algozzine, B., Ysseldyke, J., & Christenson, S. The influence of teachers' tolerances for specific kinds of behaviors on their ratings of a third grade student (Research Report No. 74). April, 1982.

Wesson, C., Deno, S., & Mirkin, P. Research on developing and monitoring progress on IEP goals: Current findings and implications for practice (Monograph No. 18). April, 1982.

Mirkin, P., Marston, D., & Deno, S. L. Direct and repeated measurement of academic skills: An alternative to traditional screening, referral, and identification of learning disabled students (Research Report No. 75). May, 1982.

Algozzine, B., Ysseldyke, J., Christenson, S., & Thurlow, M. Teachers' intervention choices for children exhibiting different behaviors in school (Research Report No. 76). June, 1982.

Tucker, J., Stevens, L. J., & Ysseldyke, J. E. Learning disabilities: The experts speak out (Research Report No. 77). June, 1982.

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

Graden, J. L., Thurlow, M. L., Ysseldyke, J. E., & Algozzine, B. Instructional ecology and academic responding time for students in different reading groups (Research Report No. 79). July, 1982.

Mirkin, P. K., & Potter, M. L. A survey of program planning and implementation practices of LD teachers (Research Report No. 80). July, 1982.

Fuchs, L. S., Fuchs, D., & Warren, L. M. Special education practice in evaluating student progress toward goals (Research Report No. 81). July, 1982.

Kuehnle, K., Deno, S. L., & Mirkin, P. K. Behavioral measurement of social adjustment: What behaviors? What setting? (Research Report No. 82). July, 1982.

Fuchs, D., Dailey, Ann Madsen, & Fuchs, L. S. Examiner familiarity and the relation between qualitative and quantitative indices of expressive language (Research Report No. 83). July, 1982.

Videen, J., Deno, S., & Marston, D. Correct word sequences: A valid indicator of proficiency in written expression (Research Report No. 84). July, 1982.

Potter, M. L. Application of a decision theory model to eligibility and classification decisions in special education (Research Report No. 85). July, 1982.

Greener, J. E., Thurlow, M. L., Graden, J. L., & Ysseldyke, J. E. The educational environment and students' responding times as a function of students' teacher-perceived academic competence (Research Report No. 86). August, 1982.

Deno, S., Marston, D., Mirkin, P., Lowry, L., Sindelar, P., & Jenkins, J. The use of standard tasks to measure achievement in reading, spelling, and written expression: A normative and developmental study (Research Report No. 87). August, 1982.

Skiba, R., Wesson, C., & Deno, S. L. The effects of training teachers in the use of formative evaluation in reading: An experimental-control comparison (Research Report No. 88). September, 1982.