DOCUMENT RESUME

ED 256 781 TM 850 222

AUTHOR	Fuchs, Lynn S.; Fuchs, Douglas
TITLE	A Quantitative Synthesis of Effects of Formative Evaluation on Achievement.
PUB DATE	Mar 85
NOTE	29p.; Paper presented at the Annual Meeting of the American Educational Research Association (69th, Chicago, IL, March 31-April 4, 1985).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Academic Achievement; Aptitude Treatment Interaction; Behavior Modification; *Effect Size; Elementary Secondary Education; *Formative Evaluation; *Individualized Instruction; Measurement Objectives; Meta Analysis; Preschool Education; Research Methodology; Student Evaluation

ABSTRACT

While the aptitude treatment interaction (ATI) approach to educational measurement emphasizes establishing salient learner characteristics, systematic formative evaluation provides ongoing evaluation for instructional program modification. Systematic formative evaluation appears more tenable than ATI for developing individualized instructional programs. This meta-analysis investigates the effects of systematic formative evaluation of educational programs on student achievement. Twenty-one controlled studies generated 95 relevant effect sizes, with an average effect size of .72. The magnitude of effect size was associated with publication type, data evaluation methods, and use of behavior modification. Findings indicate that unlike reported ATI approaches to individualization, systematic formative evaluation procedures reliably increase academic achievement. This suggests that, given an adequate measurement methodology, practitioners can inductively formulate successful individualized educational programs. (Author/BS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED256781

A Quantitative Synthesis of Effects of Formative Evaluation on Achievement

Lynn S. Fuchs and Douglas Fuchs

Peabody College, Vanderbilt University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
L. Fuchs

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Requests for reprints should be sent to Lynn S. Fuchs, Box 328, Department of Special Education, Peabody College, Vanderbilt University, Nashville, TN 37203.

Portions of this paper were presented at the annual meeting of the American Educational Research Association, Chicago, April, 1985.

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy

Running Head: Systematic Formative Evaluation

TM 850 222

BEST COPY AVAILABLE

ERIC

Abstract

This meta-analysis investigated the effects of formative evaluation procedures on student achievement. The data source was 21 controlled studies, which generated 95 relevant effect sizes, with an average effect sizes of .72. The magnitude of the effect of formative evaluation was associated with publication type, data-evaluation method, and use of behavior modification. Implications for practice are discussed.

BEST COPY

ERIC

A Quantitative Synthesis of Effects of Formative Evaluation on Achievement

An essential purpose of educational measurement is to generate information with which to formulate instructional programs (Glaser & Nitko, 1971). Given a model of individualized instruction wherein students are taught not only at different rates but also by means of different instructional methodologies, measurement serves two major functions. First, it provides a description of the learner to guide the selection of an initial set of instructional procedures. Second, as the student engages in the instructional process, measurement generates information with which the effectiveness of the initial educational program can be evaluated and modified as required. Theoretically, these two purposes appear to complement one another. Nevertheless, in practice, they have become associated with markedly different approaches to the development of individualized instructional programs.

The first is an Aptitude-Treatment Interaction (ATI) approach. It exemplifies the first measurement function, the initial description of learners. ATI proponents presume that specific learner characteristics, or aptitudes, interact predictably with certain types of instructional programs, or treatments, to produce comparatively strong student learning. Thus, with an ATI approach, the development or selection of educational programs is derived from a prior explication of learner characteristics. It is a deductive approach to formulating educational programs.

In theory, the number of possible ATIs is limited only by our capacity to generate learner characteristics and related educational programs. Not surprisingly, this perspective has inspired much research activity into possible salient learner characteristics (see Snow & Lohman, 1984) and models of instruction (see Lloyd, 1984). Nevertheless, there are several important prob-

BEST COPY

lems in basing educational programs on initial diagnoses of learner characteristics. First, at present, there is incomplete conceptualization of students' cognitive abilities (Ysseldyke, 1979). Second, and relatedly, available tests of learner characteristics do not possess appropriate technical qualities (Salvia & Ysseldyke, 1981). Third, evidence indicates the manner in which these tests often are administered (e.g., in one sitting and by an unfamiliar examiner) may discriminate systematically against select groups of students (Fuchs & Fuchs, 1985; Fuchs, Fuchs, Power, & Dailey, in press). Fourth, knowledge concerning interactions among learner and teacher characteristics, educational treatments, and classroom environments is far from complete (Ysseldyke, 1979). These problems associated with an ATI approach appear serious. In all likelihood, they contribute to the fact that current research does not support the use of ATI approaches to improve achievement among special education students (see Lloyd, 1984).

The second, contrasting approach to developing individualized educational programs is systematic formative evaluation. Whereas an ATI approach emphasizes the importance of the first purpose of educational measurement, establishing salient learner characteristics, systematic formative evaluation embodies educational measurement's second major function: ongoing evaluation and modification of proposed programs. Specifically, this approach employs regular monitoring of student performance under different instructional procedures. The purpose of this monitoring is to provide a data base with which individualized programs may be developed empirically. Thus, systematic formative evaluation is an inductive, rather than deductive, approach to developing instructional programs.

There are at least three reasons why, at present, systematic formative

BEST COPY

evaluation appears more tenable than ATI as a general strategy to develop individualized instructional programs. First, its inductive nature avoids reliance on initial diagnoses of learner characteristics when there are incomplete conceptualizations of the relation between students' abilities and educational treatments. Second, its measurement procedures have been shown to be psychometrically acceptable, whereas many ATI-related measures are seemingly inadequate. Third, and relatedly, it requires repeated measurement by classroom teachers in familiar classroom settings, which appears more ecologically valid and less reactive than the use of traditional assessment procedures associated with typical ATI approaches.

Moreover, systematic formative evaluation's repeated use of technically adequate measurement procedures appears consonant with public demand for accountability in the schools, as reflected in legislative action such as PL 94-142 (Deno & Mirkin, 1977). Nevertheless, there has been no attempt to integrate available research on systematic formative evaluation or to quantify the magnitude of effect associated with such an approach to formulating individualized programs. This lack of research contrasts sharply with the numerous integrations of research on ATI-related strategies in special education (e.g., Arter & Jenkins, 1977, 1979; Hammill & Larsen, 1974; Hammill & Wiederholt, 1973; Kavale, 1981; Tarver & Dawson, 1978). Consequently, the purpose of the current investigation was to conduct a meta-analysis of studies exploring the effects of systematic formative evaluation of educational programs on academic achievement.

A previous meta-analysis investigated a related aspect of formative evaluation, corrective feedback (Lysakowski & Walberg, 1982). However, the studies constituting that meta-analysis addressed only the effects of student feedback.

BEST COPY

As Linn (1983) has noted, teacher feedback also is critical to the use of testing for formative purposes. Therefore, the present study contributes to the previous data base by quantifying the effects of formative feedback to teachers for the purpose of empirically developing individualized instructional programs.

Method

Search Procedure

The search for pertinent studies comprised four steps. First, employing the Thesaurus of Psychological Index Terms (APA, 1982), multiple descriptors were generated from key topic-related terms. For example, student achievement alternately was identified by "student progress," "goal attainment," and "educational effects." Second, these terms facilitated a computer search of three on-line data bases: (a) ERIC, a data base of educational materials from the Educational Resources Information Center consisting of abstracts from Research in Education and Current Index to Journals in Education; (b) Comprehensive Dissertation Abstracts; and (c) Psychological Abstracts. Third, employing similar key descriptors, a manual search was conducted of five educational journals for the years 1973 through 1983. These journals were: American Educational Research Journal, Journal of Learning Disabilities, Journal of Precision Teaching, Journal of Special Education, and Learning Disability Quarterly. Fourth, titles in the reference sections of investigations discovered by these efforts were explored for additional studies.

BEST COPY

Criteria for Relevant Studies

A study was considered for inclusion if it employed a control group to evaluate the effects of providing systematic formative evaluation to teachers concerning the academic performance of preschool, elementary, and/or secondary students. Studies were excluded that (a) monitored nonacademic behaviors, (b) primarily focused on the use of behavior modification, while employing time series to test experimental effects, (c) provided test feedback only to students, and/or (d) employed college-age subjects.

The search yielded 29 studies that met the inclusion criteria. From these studies, 8 were eliminated because of insufficient data for calculating meta-analytic statistics.

Data Extracted from Each Study

Guidelines were established to ensure that each relevant effect was counted only once in analyses and that papers reporting results of the same study were grouped within analyses as one investigation.[1]

Effect size. Results of the studies were transformed to a common metric, effect size, defined here as the difference between the treatment means, divided by the control group standard deviation. For purpose of analysis, an effect was given a positive sign if subjects achieved greater scores in the systematic formative evaluation treatment. For studies reporting relevant means and standard deviations for the systematic formative evaluation and control groups, effect sizes were calculated from these statistics. For studies not reporting means and standard deviations, effect sizes were calculated from other statistics, such as F or p-values (see Glass, McGaw, & Smith, 1981). Be-

BEST COPY

fore averaging effect sizes, each one was converted to an unbiased effect size (UES) to correct for the inconsistency in estimating true from observed effect sizes (Hedges, 1981). The difference between the observed and unbiased effect sizes was neglible ($X = .019$, $SD = .025$) as has been demonstrated elsewhere (Bangert-Drowns, Kulik, & Kulik, 1983). Nevertheless UESs were employed to insure the mathematical tractability of the data.

Meta-analytic Z. Results from the 21 studies were combined to determine the unweighted Stouffer meta-analytic Z (Rosenthal, 1978). This statistic permits computation of the probability that the combined effect of children's greater achievement scores in the systematic formative evaluation treatment would occur by chance. It was derived by changing the p-values of all effects to z scores, summing them, and dividing this sum by the square root of the number of studies included. When calculating a z score for studies in which multiple dependent variables were analyzed, a median p-value was calculated for each study and its associated z score was used in the meta-analysis (see Rosenthal & Rubin, 1978).

Methodological and Substantive Study Features

Methodological study features. The effects of systematic formative evaluation of pupils' academic progress were related to three methodological variables that were coded for each study.

1. Publication type. This refers to a description of the kind of literature in which the studies were found. Coded values included "journals," "dissertations," and "nonpublished" studies such as ERIC reports, conference presentations, and solicited manuscripts.

2. Publication year. This variable was coded "before 1975," "between

BEST COPY

1975 and 1979," and "between 1980 and 1984."

3. Quality of Study. Each study was coded as "poor," "fair," or "good." To accomplish this, raters analyzed studies to identify "serious" and "less serious" threats to internal validity. "Serious" threats included (a) unequivalent subject groups, (b) confounded experimental treatments, and (c) nonrandom assignment of subjects to treatments. Examples of "less serious" threats were (a) the use of technically inadequate dependent measures, (b) uncontrolled examiner expectancy, (c) unchecked fidelity of treatment, (d) the employment of inappropriate statistical unit of analysis, and (e) inadequate teacher training. "Poor" quality studies were identified on the basis of at least one serious threat and/or because of a minimum of at least three less serious design flaws. Investigations were considered "fair" in quality if they were free of serious threats and evidenced no more than two less serious methodological problems. "Good" quality denoted studies displaying no more than one less serious methodologial problem.

Interrater agreement on each of the three methodological variables, based on two raters' evaluations of eight randomly selected studies (38% of the sample), ranged from 75% to 100%. Average agreement across the three methodological features was 92%.

Substantive study features. There were six substantive variables.

1. Behavior modification. Studies incorporating behavior modification as part of a formative evaluation treatment were distinguished from those investigations that did not use this adjunct treatment.

2. Data display. Investigations in which teachers were required to graph student performance data were differentiated from those in which teachers were asked simply to record data.

BEST COPY

3. Data evaluation. Studies were identified on the basis of whether participants (a) were required to employ explicit, systematic data-evaluation rules that indicted when and/or how they were to introduce programmatic changes, or (b) were permitted to judge for themselves when and how to make changes in students' programs.

4. Grade level. Subjects' average grade levels were aggregated into "preschool[3] through primary," "intermediate," or "junior and senior high" groups.

5. Measurement frequency. Studies were noted for the frequency with which student performance was measured: 2, 3, or 5 times per week.

6. Treatment duration. Study length was coded in terms of "less than 3 weeks," "3 to 10 weeks," or "greater than 10 weeks."

Two raters independently coded the six substantive features in eight randomly selected studies (38% of the sample). Interrater agreement[2] for the substantive features ranged from 75% to 100%. Average agreement across all six substantive variables was 86%.

Results

Overall Effects

Results of the 21 studies were combined to provide three interrelated aggregate descriptions of the effects of systematic formative evaluation: unbiased effect size (UES), percentage of distribution nonoverlap, and meta-analytic Z.

The overall mean UES was .72 (SD = .88; SE = .09, t (94) = 7.97, $p < .001$. In terms of the percentage of nonoverlap between experimental and control group distributions, U_3 (Cohen, 1977), a UES of .72 indicates that the upper 50% of the experimental group distribution exceeds approximately 76% of the control group distribution. In terms of the standard normal curve and an achievement test scale with a population mean of 100 and a standard deviation of 15, the integration of formative evaluation with instruction would raise the typical achievement outcome score from 100 to 110.80, or from the 50th to 76th percentile.

The meta-analytic Z was 4.43, $p < .001$, indicating that it is highly unlikely that the combined effect of students' greater achievement scores in the systematic formative evaluation treatment occurred by chance. Credence in a statistically reliable meta-analytic Z may be compromised by the suspicion that researchers do not report nonsignificant results (Greenwald, 1975). Rosenthal (1979) described a method for determining the number of unreported null effects that would be needed to reduce a meta-analytic Z to nonsignificance. The larger this "fail-safe N," the more confidence one can have in the reliability of a meta-analytic result. This investigation's fail-safe N was 131, indicating that it would take 131 studies summing to a null result to raise the probability of the meta-analytic Z beyond .05.

Relation Between UESs and Study Features

Methodological features. Table 1 displays data for the UESs by methodological features of the effect sizes. There was one significant effect for type of publication. As indicated in Table 1, a follow-up Scheffe analysis indicated that the mean UES associated with reports published in journals was statistically significantly greater than the average UES associated with unpublished studies.

BEST COPY

Insert Table 1 about here

Substantive features. Table 1 also shows UESs by substantive features. There were two statistically significant effects. First, the data-evaluation variable yielded a significant F value, with the mean UES greater for the use of data-evaluation rules than for teacher judgment. Second, the factor behavior modification resulted in a significant difference; the average UES was greater when behavior modification procedures were incorporated as part of the experimental treatment.

Discussion

The purpose of this meta-analysis was to determine the effects of systematic formative evaluation of educational programs on academic achievement. Results indicated the use of systematic formative evaluation procedures significantly increased students' school achievement, both statistically and pratically. The mean effect size of .72 was reliably different from zero. It suggests that one can expect students whose programs are monitored systematically and developed formatively over time to achieve, on average, almost three-quarters of a standard deviation higher than students whose programs are not systematically monitored and developed formatively.

Moreover, this finding generally was robust over several methodological features associated with the effect sizes: Neither quality of study nor publication year appeared to mediate or moderate formative evaluation effects. Only publication type yielded a statistically significant difference, wherein effect sizes associated with studies published in journals were higher than those de-

BEST COPY

rived from unpublished manuscripts. Such a finding might be anticipated given the tendency of journals to reject studies that fail to yield reliable results (Rosenthal, 1978), and given the related suspicion that researchers do not report nonsignificant results (Greenwald, 1975). Nevertheless, the meta-analytic Z analysis indicated that it would require the addition to this meta-analysis of as many as 131 studies summing to a null result to reduce findings to nonsignificance.

Findings were robust across not only methodological features, but also substantive variables. Specifically, systematic formative evaluation was similarly effective regardless of students' age, treatment duration, the frequency with which measurements were taken, or whether student data were graphed. Nonsignificant findings associated with some of these variables may be explained at least partially by the low number of effect sizes for certain coded variables. For example, the mean effect size associated with the practice of graphing data was 3.5 times greater than the average effect size associated with simply recording such information; however, there were only seven effect sizes for recording.

Despite the general robustness of findings for substantive variables, two substantive study features produced reliably different effect sizes, and appear to reflect critical dimensions of effective formative evaluation systems. Specifically, effect sizes connected with the use of behavior modification in addition to systematic formative evaluation were reliably higher than those representative of systematic formative evaluation only. This finding is consonant with previous research. For example, Bloom (1984) reported an effect size of 1.20 on student achievement for reinforcing students' academic behavior. Thus, it is not surprising that incorporating reinforcement as part of system-

BEST COPY

atic monitoring procedures would produce differentially greater student achievement.

A less predictable finding of the current study was the significant difference associated with data-evaluation methods. When teachers were required to employ data-utilization rules, effect sizes were higher than when data were evaluated by teacher judgment. Data-evaluation rules required practitioners to analyze student performance at regular intervals and, if the data suggested certain patterns, to introduce instructional changes into a student's educational program. For example, Fuchs, Deno, and Mirkin (1984) required teachers to calculate a line of best fit through every 7 to 10 data points. If a line of best fit was less steep than the goal line, running from baseline to the intersection of the criterion performance and the goal date, teachers were required to institute a programmatic change. Results suggest that, in order to effect greater learning for pupils, teachers might employ explicit, systematic rules to evaluate the data they collect. This finding is in concert with previous work (Baldwin, 1976; Tindal, Fuchs, Christenson, Mirkin, & Deno, 1981; White, 1974), demonstrating that although teachers may collect student performance data according to designated time schedules, they frequently do not employ those data meaningfully to modify students' educational programs.

Therefore, findings of the current study indicate that the use of systematic formative evaluation procedures reliably increases academic achievement, and that effects may be enhanced when teachers also employ behavior modification and data-evaluation rules. The apparent effectiveness of systematic formative evaluation suggests that, given an adequate measurement methodology, practitioners can inductively formulate successful individualized educational programs. This conclusion contrasts with a body of literature indicating that

ATI approaches to individualization, wherein different instructional programs are <u>deductively</u> formulated from explications of learner characteristics, fail to enhance achievement. The use of systematic formative evaluation and resulting development of effective individualized programs might be considered by those who, in their astute criticisms of ATI approaches, also have questioned the validity of individualized instruction (see, for example, Lloyd [1984]). Given results of this meta-analysis, we believe such questioning of the legitimacy of individualized instruction may represent a case of "throwing the baby out with the bath."

Current findings must be considered in light of at least two possible methodological problems. First, a limited number of researchers dominate the group experimental literature in systematic formative evaluation. Specifically, one team of investigators (Deno et al.) accounted for eight reports employed in the meta-analysis and one researcher (Beck) accounted for four studies. While such a pattern may be problematic because it may inflate meta-analytic findings (see Slavin, 1984), post-hoc analysis indicates that the effect sizes associated with these sets of researchers did not inflate results, but rather tended to underestimate the average effect size. A second concern in this quantitative synthesis is that a relatively small number of investigations produced a large number of effect sizes, a situation that results in dependency among effect sizes. This methodological problem commonly is associated with meta-analytic research. It warrants additional attention by those concerned with the development of meta-analytic methodology, and current findings must be understood within the confines of this limitation.

Footnotes

[1] One paper authored by Haring (1971) and two additional reports by Haring and Krug (1975a, 1975b) described aspects of the same investigation. Only non-redundant effect sizes were extracted from these reports and, when analyses required that effect sizes be grouped by investigation, such as the meta-analytic Z, these effect sizes were grouped as one investigation. Therefore, although it is reported that 21 studies were employed in the meta-analysis, 23 appear in the appendix due to the separate listings of the Haring and Haring and Krug papers.

[2] Interrater agreement was calculated using the following formula (Coulter in Thompson, White, & Morgan, 1982): Percentage agreement = agreements between rater A & rater B / (agreements between A & B + disagreements between A & B + omissions by A + omissions by B).

[3] Only one study included subjects whose average age was at the preschool level; thus, effect sizes for preschool children were grouped with those associated with primary grade students.

BEST COPY

References

American Psychological Association. (1982). Thesaurus of psychological index terms (3rd ed.). Washington, DC: Author.

Arter, J.A., & Jenkins, J.R. (1977). Examining the benefits and prevalence of modality considerations in special education. Journal of Special Education, 11, 281-298.

Arter, J.A., & Jenkins, J.R. (1979). Differential diagnosis-prescriptive teaching: A critical appraisal. Review of Educational Research, 49, 517-555.

Baldwin, V. (1976). Curriculum concerns. In M.A. Thomas (Ed.), Hey, don't forget about me. Reston, VA: Council for Exceptional Children.

Bangert-Drowns, R.L., Kulik, J.A., & Kulik, C.C. (1983). Effects of coaching programs on achievement test performance. Review of Educational Research, 53, 571-585.

Bloom, B.S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher, 13, (6), 4-16.

Cohen, J. (1977). Statistical power analysis for the behavioral sciences. New York: Academic Press.

Deno, S.L., & Mirkin, P.K. (1977). Data-based program modification: A manual. Reston, VA: Council for Exceptional Children.

Fuchs, D., & Fuchs, L.S. (1985). Test procedure bias: A meta-analysis. Paper to be presented at the annual meeting of the American Educational Research Association, Chicago.

Fuchs, D., Fuchs, L.S., Power, M.H., & Dailey, A.M. (in press). Bias in the assessment of handicapped children. American Educational Research Journal.

BEST COPY

Fuchs, L.S., Deno, S.L., & Mirkin, P.K. (1984). The effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning. American Educational Research Journal, 21, 449-460.

Glaser, R., & Nitko, J. (1971). Measurement in learning and instruction. In R. Thorndike (Ed.), Educational measurement (2nd ed.). Washington, D.C.: American Council on Education.

Glass, G., McGaw, B., & Smith, M.L. (1981). Meta-analysis in social research. Beverly Hills: Sage.

Greenwald, A.G. (1975). Consequences of prejudice against the null hypothesis. Psychological Bulletin, 82, 1-20.

Hammill, D.D., & Larsen, S. (1974). The effectiveness of psycholinguistic training. Exceptional Children, 41, 5-14.

Hammill, D.D., & Wiederholt, J.L. (1973). Review of the Frostig visual perception test and the related training program. In L. Mann & D.A. Sabatino (Eds.), First review of speical education (Vol. 1, pp. 33-48). Philadelphia: JSE Press.

Hedges, L. (1981). Distribution theory for Glass's estimator of effect size and related estimators. Journal of Educational Statistics, 6, 359-361.

Kavale, K. (1981). Functions in the Illinois Test of Psycholinguistic Abilities (ITPA): Are they trainable? Exceptional Children, 47, 496-510.

Linn, R.L. (1983). Testing and instruction: Links and distinctions. Journal of Educational Measurement, 20, 179-191.

Lloyd, J.W. (1984). How shall we individualize instruction - Or should we? Remedial and Special Education, 5, 7-15.

Lysakcwski, R.S., & Walberg, H.J. (1981). Classroom reinforcement: A quantitative synthesis. Journal of Educational Research, 75, 69-77.

BEST COPY

Rosenthal, R. (1978). Combining results of independent studies. Psychological Bulletin, 85, 185-193.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null effects. Psychological Bulletin, 86, 638-641.

Rosenthal, R., & Rubin D.B. (1978). Interpersonal expectancy effects: The first 345 studies. The Behavioral and Brain Sciences, 3, 377-415.

Salvia, J., & Ysseldyke, J. (1981). Assessment in special and remedial education (2nd ed.). Boston: Houghton-Mifflin.

Slavin, R.E. (1984). Meta-analysis in education: How has it been used? Educational Researcher, 13 (8), 6-15.

Snow, R.E. & Lohman, D.F. (1984). Toward a theory of cognitive aptitude for learning from instruction. Journal of Educational Psychology, 76, 347-376.

Tarver, S.G., & Dawson, M.M. (1978). Modality preference and the teaching of reading: A review. Journal of Learning Disabilities, 11, 5-17.

Thompson, R.H., White, K.R., & Morgan, D.P. (1982). Teacher-student interaction patterns in classrooms with mainstreamed mildly handicapped students. American Educational Research Journal, 19, 220-236.

Tindal, G., Fuchs, L.S., Christenson, S., Mirkin, P.K., & Deno, S.L. (1981). The relationship between student achievement and teacher assessment of short- or long-term goals (Research Report No. 61). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities. (ERIC Document Reproduction Service No. ED 218 846)

White, O.R. (1974). Evaluating educational process (Working paper). Seattle: University of Washington, Child Development and Mental Retardation Center, Experimental Education Unit.

BEST COPY

Ysseldyke, J.E. (1979). Psychoeducational assessment and decision making. In J.E. Ysseldyke & P.K. Mirkin (Eds.), Proceedings of the Minnesota Round-table Conference on Assessment of Learning Disabled Children (Monograph No. 8). Minneapolis: Univeristy of Minnesota, Institute for Research on Learning Disabilities. (ERIC Document Reproduction Service No. ED 185 765)

BEST COPY

Table 1

Means, Standard Deviations, F values, and Significant Scheffe Contrasts on UESs by Methodological and Substantive Study Features

Feature	Mean	SD	N[a]	F	df	Scheffe Contrast
<u>Methodological</u>						
Publication Type				7.94[b]	2,93	A > C
A. Journal	1.27	1.09	20			
B. Dissertation	.80	.99	32			
C. Unpublished	.41	.45	44			
Publication Year				.67	2,93	
A. Before 1975	.33	.52	17			
B. Between 1975 and 1979	.56	.83	32			
C. Between 1980 and 1984	.58	.80	47			
Quality of Study				.43	2,93	
A. Good	.85	.45	19			
B. Fair	.71	.99	66			
C. Poor	.54	.63	11			
<u>Substantive</u>						
Behavior Modification				11.96[b]	1,94	A > B
A. With behavior modification	1.15	1.31	30			
B. Without behavior modification	.52	.48	66			

(Table 1 continued on p. 20.)

BEST COPY

Table 1 (Continued)

Feature	Mean	SD	N[a]	F	df	Scheffe Contrast
Data Display				2.54	1,94	
A. Graphed	.75	.89	89			
B. Recorded	.21	.35	7[d]			
Data Evaluation				9.11[c]	1,94	A > B
A. By rule	.95	1.08	50			
B. By judgment	.44	.43	46			
Grade Level				1.65	2,93	
A. N-3	.51	.39	31			
B. 4-6	.71	.63	30			
C. 7-12	.90	1.26	35			
Measurement Frequency				2.51	2,93	
A. Twice per week	1.00	.49	11			
B. Three times per week	.31	.46	16			
C. Daily	.77	.96	69			
Treatment Duration				.88	2,93	
A. Fewer than 3 weeks	.60	.66	2[d]			
B. 3-10 weeks	.46	.52	16			
C. More than 10 weeks	.78	.93	78			

[a] N represents number of UESs not number of studies.
[b] $p < .001$.
[c] $p < .01$.
[d] The small N in these categories may result in unstable estimates of UES.

BEST COPY

Appendix

Reports Included in the Meta-Analysis

Beck, R. (1976). Report for the office of education dissemination review panel. (Unpublished manuscript available at Precision Teaching Project, 3300 Third St. N.E., Great Falls, MT 59404.)

Beck, R. (1979). Report for the office of education dissemination review panel. (Unpublished manuscript available at Precision Teaching Project, 3300 Third St. N.E., Great Falls, MT 59404.)

Beck, R. (1981). Curriculum management through a data base. (Unpublished manuscript available at Precision Teaching Project, 3300 Third St. N.E., Great Falls, MT 59404.)

Beck, R. (1981). High school basic skills improvement project. (Unpublished manuscript available at Precision Teaching Project, 3300 Third St. N.E., Great Falls, MT 59404.)

Bohannon, R.M. (1975). Direct and daily measurement procedures in the identification and treatment of reading behaviors in children in special education. Unpublished doctoral dissertation, University of Washington.

Bradfield, R.H., Brown, J., Kaplan, P., Rickert, E., & Stannard, R. (1973). The special children in the regular classroom. Exceptional Children, 39, 384-391.

Brandstetter, G., & Merz, C. (1978). Charting scores in precision teaching for skill acquisition. Exceptional Children, 45, 42-48.

Bruening, S.E. (1978). Precision teaching in the high school classroom: A necessary step towards maximizing teacher effectiveness and student performance. American Educational Research Journal, 15, 125-140.

Crutcher, C.E., & Hofmeister, A.M. (1975). Effective use of objectives and

BEST COPY

monitoring. Teaching Exceptional Children, 7 (3), 78-79.

Dubrule, M.N. (1984). The study of precision teaching as a remedial method. Unpublished doctoral dissertation, Clark University.

Fuchs, L.S., Deno, S.L., & Mirkin, P.K. (1984). The effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning. American Educational Research Journal, 21, 449-460.

Fuchs, L.S., Wesson, C., Tindal, G., Mirkin, P.K., & Deno, S.L. (1982). Instructional changes, student performance, and teacher preferences: The effects of specific measurement and evaluation procedures (Research Report No. 64). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities. (ERIC Document Reproduction Service No. ED 218 849)

Frumess, S.C. (1973). A comparison of management groups involving the use of the standard behavior chart and setting performance aims. Unpublished doctoral dissertation, University of Houston.

Haring, N.G. (1971). Investigation of systematic instructional procedures to facilitate academic achievement in mentally retarded disadvantaged children. Final Report. (ERIC Document Reproduction No. ED 071 248)

Haring, N.G., & Krug, D.A. (1975a). Evaluation of a program of systematic instructional procedures for extremely poor retarded children. American Journal of Mental Deficiency, 79, 627-631.

Haring, N.G., & Krug, D.A. (1975b). Placement in regular programs: Procedures and results. Exceptional Children, 41, 413-417.

King, R., Deno, S.L., Mirkin, P.K., & Wesson, C. (1983). The effects of training teachers in the use of formative evaluation in reading: An

BEST COPY

experimental-control comparison (Research Report No. 111). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities.

Lovitt, T.C., & Fantasia, K. (1983). A precision teaching project with learning disabled chidlren. Journal of Precision Teaching, 3, 85-91.

Mirkin, P.K., & Deno, S.L. (1979). Formative evaluation in the classroom: An approach to improving instruction (Research Report No. 10). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities. (ERIC Document Reproduction Service No. ED 185 754)

Mirkin, P.K., Deno, S.L., Tindal, G., & Kuehnle, K. (1980). Formative evaluation: Continued development of data utilization systems (Research Report No. 23). University of Minnesota, Institute for Research on Learning Disabilities. (ERIC Document Reproduction Service No. ED 197 510)

Peniston, E.G. (1975). An evaluation of the Portage Project: A Comparison of a home-visit program for multiply handicapped preschoolers and Head-start program. (ERIC Document Reproduction Service No. ED 112 570)

Sevcik, B., Skiba, R., Tindal, G., King, R., Wesson, C., Mirkin, P.k., & Deno, S.L. (1983). Curriculum-based measurement: Effects on instruction, teaching estimates of student progress, and student knowledge of performance. (Research Report No. 124). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities.

Skiba, R., Wesson, C., & Deno, S.L. (1982). The effects of training teachers in the use of formative evaluation in reading: An experimental-control comparison (Research Report No. 88). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities.

Tindal, G., Fuchs, L.S., Christenson, S., Mirkin, P.K., & Deno, S.L. (1981). The relationship between student achievement and teacher assessment of short- or long-term goals (Research Report No. 61). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities. (ERIC Document Reproduction Service No. ED 218 846)

BEST COPY