Exceptional Children, Vol. 53, No. 3, pp. 199–208.

Effects of Systematic Formative Evaluation: A Meta-Analysis

LYNN S. FUCHS
DOUGLAS FUCHS

ABSTRACT: This meta-analysis investigated the effects of formative evaluation procedures on student achievement. The data source was 21 controlled studies, which generated 96 relevant effect sizes, with an average weighted effect size of .70. The magnitude of the effect of formative evaluation was associated with publication type, data-evaluation method, data display, and use of behavior modification. Implications for special education practice are discussed.

■ A basic assumption in special education is that individualized instruction results in improved student achievement. This belief is reflected in special education literature (e.g., Schenck, 1980; Wang & Reynolds, 1985); teacher preparation standards (e.g., Corrigan, 1978; Council for Exceptional Children Delegate Assembly, 1983); teacher training materials (e.g., Mann, Suiter, & McClung, 1979; Turnbull, Strickland, & Brantley, 1978; White & Haring, 1980); educators' attitudes about teaching (e.g., Spears, 1973; West & Bates, 1977); and federal law (PL 94-142), which mandates that an individualized educational program (IEP) be devised for each handicapped pupil.

Despite the pervasiveness of this assumption, recently some have questioned the effectiveness of individualized instruction (see, for example, Lloyd, 1984). The cause for such concern is the apparent inadequacy of the predominant methodology for formulating individualized educational programs: an *aptitude treatment interaction* (ATI) approach.

LYNN S. FUCHS *and* DOUGLAS FUCHS *are Assistant Professors, Department of Special Education, Peabody College, Vanderbilt University, Nashville, Tennessee.*

THE ATI APPROACH

An ATI approach to developing instructional programs focuses on initial descriptions 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 ATI, 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 the 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, several important problems in basing educational programs on initial

diagnoses of learner characteristics have been identified. First, at present, there is incomplete conceptualization of students' cognitive abilities (Ysseldyke, 1979). Second, available tests of learner characteristics do not possess appropriate technical qualities (Salvia & Ysseldyke, 1981). Third, evidence indicates that the manner in which these tests are often administered (e.g., in one sitting and by an unfamiliar examiner) may discriminate systematically against select groups of students (Fuchs & Fuchs, 1986; Fuchs, Fuchs, Power, & Dailey, 1985). 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 specify individualized instructional programs and, concurrently, to improve achievement among special education students (see Lloyd, 1984).

SYSTEMATIC FORMATIVE EVALUATION

Although the lack of success associated with the ATI approach has led some to reject individualized instruction, such rejections largely ignore a second, contrasting approach to developing individualized educational programs, *systematic formative evaluation*. Whereas an ATI approach emphasizes the importance of describing salient learner characteristics, systematic formative evaluation focuses on ongoing evaluation and modification of proposed programs. Specifically, this approach employs regular monitoring of handicapped students' 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 four reasons why, at present, systematic formative evaluation may be more tenable than ATI as a general strategy to develop individualized instructional programs for handicapped pupils. 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 seem to be inadequate. Third, 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. Finally, systematic formative evaluation's repeated use of technically adequate measurement procedures appears consonant with public demand for accountability in special education, as reflected in legislative action such as PL 94-142 (Deno & Mirkin, 1977).

PURPOSE

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.

Findings may be useful in addressing at least two practical questions: first, whether individualized instruction is effective, a question prompted largely by the apparent failure of ATI approaches to develop useful individualized instructional plans; second, whether the quantified effect, or "benefit," of systematic formative evaluation is sufficiently large to offset the additional time, or "cost," required to implement this evaluation procedure (see Wesson, King, & Deno, 1984).

A previous meta-analysis investigated a related aspect of formative evaluation, corrective feedback (Lysakowski & Walberg, 1982). However, the studies constituting the meta-analysis addressed only the effects of *student* feedback. As Linn (1983) has noted, *teacher* feedback is also 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* was alternately 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.

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. Systematic formative evaluation was operationalized as curriculum-based data collection that occurred at least twice each week, with decisions concerning the adequacy of programs formulated on an individual, not a group, basis. Studies were excluded that (a) monitored nonacademic behaviors, (b) focused primarily 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. (See *Note* in Appendix.)

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).

Each effect size 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 negligible ($X = .019$, $SD = .025$) as has been demonstrated elsewhere (Bangert-Drowns, Kulik, & Kulik, 1983). Nevertheless, UESs were employed to ensure the mathematical tractability of the data. There were 96 effect sizes, with between 1 and 12 comparisons per study. Analysis indicated no statistical dependency between effect size magnitude and number of comparisons per study ($r = .12$). Therefore, UESs were aggregated at the individual effect size level. In combining these UESs, weighted averages were calculated to account for the variances of the UESs (see Hedges, 1984).

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 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 or 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 methodological 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%. (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].)

Substantive Study Features. There were seven 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 where they were simply to record data.
3. *Data evaluation*. Studies were identified on the basis of whether participants (a) were required to employ explicit, systematic data-evaluation rules that indicated 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 through primary," "intermediate," or "junior and senior high" groups. (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.)
5. *Handicapped status*. Handicapped and nonhandicapped subjects were differentiated.
6. *Measurement frequency*. Studies were noted for the frequency with which student performance was measured: 2, 3, or 5 times per week.
7. *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 seven substantive features in eight randomly selected studies (38% of the sample). Interrater agreement for the substantive features ranged from 75% to 100%. Average agreement across all seven substantive variables was 86%.

Characteristics of the Sample

Of the 23 references in the Appendix, which represent 21 separate investigations, there are 4 dissertations, 11 unpublished studies, and 8 journal articles. Among the published papers, 3 appeared in *Exceptional Children*, 2 in *American Educational Research Journal*, and 1 each in *Teaching Exceptional Children*, *American Journal of Mental Deficiency*, and *Journal of Precision Teaching*. A total of 3,835 subjects participated in these studies, with 83% of the investigations employing handicapped subjects. Of these handicapped pupils, 98% were mildly to moderately handicapped and 2% were severely handicapped. The grade level of these subjects ranged from preschool through high school, with a median grade level of 3.8. Among the 21 investigations, 8 (38%) focused solely on the academic area of reading, 4 (19%) on reading and math, 3 (14%) only on math, and

1 (5%) each on (a) high school content areas, (b) preschool skills, (c) spelling, (d) math and spelling, (e) reading, math, and spelling, and (f) writing, math, and spelling.

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 weighted average UES was .70 (v = .0005, z = 31.18, $p < .001$). In terms of the percentage of nonoverlap between experimental and control group distributions, U_3 (Cohen, 1977), a UES of .70 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 110.50, or from the 50th to 76th percentile. A test for the homogeneity of effect size yielded a statistically significant value, $\chi^2(95) = 1188.19$ (see Hedges, 1984). Therefore, additional analyses were pursued to examine the relations between study features and magnitude of effect sizes.

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 Study Features and Effect Size Magnitude

To examine the relation between study features and magnitude of effect size, Hedges's (1984) analogue to analysis of variance was employed. With conventional analysis of variance conducted on effect sizes, problems exist because it is possible for systematic variance to be pooled into the estimate of error variance. Moreover, violation of the homoscedasticity assumption is severe in research synthesis, and there is little reason to believe that the usual robustness of the F test will prevail (see Hedges, 1984). The use of Hedges's analogue to analysis of variance avoids these conceptual and statistical problems.

Methodological Features. Table 1 displays weighted 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 analysis (see Hedges, 1982) indicated that the weighted mean UES associated with reports published in journals was statistically significantly greater than the weighted average UES associated with unpublished studies.

Substantive Features. Table 1 also shows UESs by substantive features. There were three statistically significant effects. First, the data-evaluation variable yielded a significant chi-square value, with UESs greater for the use of data-evaluation rules (average = .91) than for teacher judgment (mean = .42). This indicates that the upper 50% of the experimental group distribution, wherein evaluation rules are employed, exceeds approximately 82% of the control group distribution. Second, the factor behavior modification resulted in a significant difference: UESs were greater when behavior modification procedures were incorporated as part of the experimental treatment (mean = 1.12 vs. .51). Therefore, the upper 50% of the experimental group distribution of effect sizes associated with the incorporation of behavior modification exceeds approximately 87% of the control group distribution. Third, the data display variable produced a significant chi-square statistic: UESs associated with graphed data (average = .70) were higher than those of recorded data (mean = .26). Consequently, the upper 50% of the experimental group that incorporates graphic

TABLE 1
Weighted Means, z Values,[a] and Chi-Square Statistics as Analogues to Analyses of Variance on UESs by Methodological and Substantive Features

Feature	Weighted mean	z value	N[b]	χ^2	df
		Methodological			
Publication Type				216.27[d]	2
Journal	1.18	24.11	20		
Dissertation	.89	19.55	25		
Unpublished	.40	13.33	51		
Publication Year				5.16	2
Before 1975	.72	5.71	6[c]		
Between 1975 and 1979	.57	15.41	40		
Between 1980 and 1984	.72	24.80	50		
Quality of Study				4.78	2
Good	.88	19.56	19		
Fair	.67	24.44	69		
Poor	.52	7.65	8[c]		
		Substantive			
Behavior Modification				53.01[d]	1
With behavior modification	1.12	23.72	30		
Without behavior modification	.51	19.62	66		
Data Display				16.47[d]	1
Graphed	.70	30.43	89		
Recorded	.26	2.55	7[c]		
Data Evaluation				63.33[d]	1
By rule	.91	29.35	49		
By judgment	.42	12.73	47		
Grade Level				5.62	2
N–3	.50	12.44	31		
4–6	.68	18.51	30		
7–12	.87	22.36	35		
Handicapped Status				3.82	1
Handicapped	.73	27.60	74		
Nonhandicapped	.63	12.29	22		
Measurement Frequency				5.82	2
Twice per week	.85	17.36	12		
Three times per week	.41	5.55	16		
Daily	.69	25.46	68		
Treatment Duration				1.42	2
Fewer than 3 weeks	.50	1.95	2[c]		
3–10 weeks	.50	6.49	16		
More than 10 weeks	.70	30.43	78		

[a]A significant z value indicates that the weighted mean is reliably different from zero. All are significant beyond the .001 probability level except those of Recorded Data Display ($p < .025$) and Fewer than 3 Weeks Treatment Duration ($p < .05$).
[b]N represents number of UESs, not number of studies.
[c]The small *N* in these categories may result in unstable estimates of UES.
[d]$p < .001$.

data displays exceeds approximately 76% of the control group distribution.

DISCUSSION

The purpose of this meta-analysis was to determine the effects of systematic formative evaluation of educational programs on academic achievement. Results indicated that the use of systematic formative evaluation procedures, within a group of studies that employed predominantly mildly handicapped subjects, significantly increased students' school achievement, both statistically and practically. The weighted mean effect size of .70 was reliably different from zero. It suggests that one can expect handicapped students whose individualized educational programs are monitored systematically and developed formatively over time to achieve, on average, .7 standard deviation units higher than students whose programs are not systematically monitored and developed formatively. Moreover, the weighted mean effect size, calculated on only those associated with handicapped pupils, was a comparable .73.

Effect size findings were robust over two of three 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 derived 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 certain substantive variables. Specifically, systematic formative evaluation was similarly effective regardless of students' age, treatment duration, the frequency with which measurements were taken, and handicapped status. Nevertheless, three substantive features did produce 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. On average, with behavior modification, systematic formative evaluation might be expected to boost the typical achievement outcome more than 1 standard deviation over control group outcomes. 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 systematic 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. On average, systematic formative evaluation that incorporates evaluation rules might be expected to raise the typical achievement outcome .9 standard deviation unit. 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.

Finally, the method by which data were displayed produced a significant finding. When data were graphed, effect sizes were higher than when data simply were recorded. With graphing, systematic formative evaluation boosts the

average achievement outcome score almost .8 standard deviation units over control group outcomes. This finding may reflect one or a combination of at least two possible explanations. First, a graphed data display may allow teachers to analyze student performance trends more accurately and frequently; second, graphs may facilitate more frequent performance feedback directly to pupils.

Therefore, findings of the current study indicate that the use of systematic formative evaluation procedures within special education reliably increases academic achievement, and that effects may be enhanced when teachers also employ behavior modification, data-evaluation rules, and graphed data displays. Therefore, although some special education practitioners may object to systematic formative evaluation because of its time-consuming nature (Wesson et al., 1984), the magnitude of effect size associated with this methodology suggests that systematic formative evaluation may be worth additional teacher time. Moreover, recent developments in computer software, which facilitate the collection, storage, graphing, and analysis of student performance data (see Fuchs, Deno, & Mirkin, 1983; Hasselbring, 1985; Hasselbring & Hamlett, 1983), indicate that technology may improve the feasibility of implementing this methodology for specifying effective individualized instructional programs.

The apparent effectiveness of systematic formative evaluation suggests that, given an adequate methodology, special education practitioners can inductively formulate successful individualized educational programs (IEPs). This conclusion contrasts with a body of literature indicating that ATI approaches to individualization, wherein different instructional programs are *deductively* 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 water."

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 prevalences 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. *Educ. Researcher, 13*(6), 4-16.

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

Corrigan, D. C. (1978). Public Law 94-142: A matter of human rights; A call for change in schools and colleges of education. In J. K. Grosenick & M. C. Reynolds (Eds.), *Teacher education: Renegotiating roles for mainstreaming*. Reston, VA: Council for Exceptional Children.

Council for Exceptional Children Delegate Assembly. (1983). Code of ethics and standards for professional practice. *Exceptional Children, 50*, 205-209.

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

Fuchs, D., & Fuchs, L. S. (1986). Test procedure bias: A meta-analysis of examiner familiarity effects. *Review of Educational Research, 56*, 243-262.

Fuchs, D., Fuchs, L. S., Power, M. H., & Dailey, A. M. (1985). Bias in the assessment of handicapped children. *American Educational Research Journal, 22*, 185-198.

Fuchs, L. S., Deno, S. L., & Mirkin, P. K. (1983). Data-based program modification: A continuous evaluation system with computer technology to facilitate implementation. *Journal of Special Education Technology, 6*(2), 50-57.

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.

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 special education* (Vol. 1, pp. 33-48). Philadelphia: JSE Press.

Hasselbring, T. S. (1985). *Toward the development of expert assessment systems*. Working paper. (Available from T. Hasselbring, Box 328, Peabody College, Vanderbilt University, Nashville, TN 37203.)

Hasselbring, T. S., & Hamlett, C. L. (1983). *AIMSTAR: A data management and decision-making system*. Portland, OR: ASIEP Education Co.

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

Hedges, L. (1982). Fitting categorical models to effect sizes from a series of experiments. *Journal of Educational Statistics, 7*, 119-137.

Hedges, L. (1984). Advances in statistical methods for meta-analysis. In W. H. Yeaton & P. N. Wortman (Eds.), *Issues in data synthesis*. New Directions for Program Evaluation, 24. San Francisco: Jossey-Bass.

Kavale, K. (1981). Functions in the Illinois Test of Psycholinguistic Abilties (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.

Lysakowski, R. S., & Walberg, H. J. (1982). Instructional effects of cues, participation, and corrective feedback: A quantitative synthesis. *American Educational Research Journal, 19*, 559-578.

Mann, P. H., Suiter, P. A., & McClung, R. (1979). *Handbook in diagnostic-prescriptive teaching* (2nd ed.). Boston: Allyn & Bacon.

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.

Schenck, S. J. (1980). The diagnostic-instructional link in individualized education programs. *Journal of Special Education, 14*, 337-345.

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

Spears, H. (1973). Kappans ponder typical school procedures. *Phi Delta Kappan, 54*, 615-618.

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)

Turnbull, A., Strickland, B. B., & Brantley, J. C. (1978). *Developing and implementing individualized educational programs*. Columbus: Merrill.

Wang, M., & Reynolds, M. C. (1985). Avoiding the "Catch 22" in special education reform. *Exceptional Children, 51*, 497-502.

Wesson, C., King, R. P., & Deno, S. L. (1984). Direct and frequent measurement of student performance: If it's good for us, why don't we do it? *Learning Disability Quarterly, 7*, 45-48.

West, T. L., & Bates, P. (1977). The implications of mainstreaming for teacher education. In P. Bates, T. L. West, & R. B. Schmerl (Eds.), *Mainstreaming: Problems, potentials, and perspectives*. Minneapolis: University of Minnesota.

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

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

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

APPENDIX
Reports Included in the Meta-Analysis

Note: One paper authored by Haring (1971) and two additional reports by Haring and Krug (1975a, 1975b) described aspects of the same investigation. Only nonredundant effect sizes were extracted from these reports and, when analyses required that effect sizes be grouped by investigation, 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 this Appendix due to the separate listings of the Haring and Haring and Krug papers.

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. *Exceptnl. Childrn.*, 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 teaching effectiveness and student performance. *American Educational Research Journal*, 15, 125-140.

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

Frumess, S. C. (1983). *A comparison of management groups involving the use of the standard behavior chart and setting performance aims.* Unpublished doctoral dissertation, University of Houston.

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.

Haring, N. G. (1971). *Investigation of systematic instructional procedures to facilitate academic achievement in mentally retarded disadvantaged children. Final report.* (ERIC Document Reproduction Service 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 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 children. *Journal of Precision Teaching*, 3, 85-91.

Mirkin, P. K. (1978). *A comparison of the effects of three formative evaluation strategies and contingent consequences on reading performance.* Unpublished doctoral dissertation, University of MN.

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 multiple handicapped preschoolers and Head-start program.* (ERIC Document Reproduction Service No. ED 112 [illegible])

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 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 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)

Portions of this article were presented at the annual meeting of the American Educational Research Association in Chicago, April 1985. Requests for reprints should be sent to Lynn S. Fuchs, Box 328, Department of Special Education, Peabody College, Vanderbilt University, Nashville, TN 37203.

Manuscript received May 1985; final revision accepted May 1985.