Journal of Educational Psychology
2000, Vol. 92, No. 4, 605–619

0022-0663/00/$5.00 DOI: 10.1037//0022-0663.92.4.605

How Effective Are One-to-One Tutoring Programs in Reading for Elementary Students at Risk for Reading Failure? A Meta-Analysis of the Intervention Research

Batya Elbaum
University of Miami

Sharon Vaughn
University of Texas at Austin

Marie Tejero Hughes and Sally Watson Moody
University of Miami

A meta-analysis of supplemental, adult-instructed one-to-one reading interventions for elementary students at risk for reading failure was conducted. Reading outcomes for 42 samples of students (N = 1,539) investigated in 29 studies reported between 1975 and 1998 had a mean weighted effect size of 0.41 when compared with controls. Interventions that used trained volunteers or college students were highly effective. For Reading Recovery interventions, effects for students identified as discontinued were substantial, whereas effects for students identified as not discontinued were not significantly different from zero. Two studies comparing one-to-one with small-group supplemental instruction showed no advantage for the one-to-one programs.

One-to-one instruction, provided as a supplement to classroom teaching, is generally considered to be the most effective way of increasing students' achievement. The effectiveness of one-to-one instruction has been validated by empirical research, especially for students who are considered at risk for school failure or have been identified as having reading or learning disabilities (Bloom, 1984; Jenkins, Mayhall, Peschka, & Jenkins, 1974; Juel, 1991; Wasik & Slavin, 1993). According to Adler (1998), more and more parents, dissatisfied with their children's academic progress, are hiring tutors to provide additional instruction to their children.

Classroom teachers identify adult-delivered one-to-one instruction as the ideal teaching practice but report that they are rarely able to implement it in their classrooms (Moody, Vaughn, & Schumm, 1997). Corroborating these teachers' reports is a study indicating that when one-to-one instruction is provided within the general education classroom, it is usually implemented for less than 1 min and serves largely to clarify information, answer questions, or check for understanding (McIntosh, Vaughn, Schumm, Haager, & Lee, 1993) rather than to provide systematic, remedial instruction. Even in special education classrooms, one-to-one instruction may occur in only a limited way (Vaughn, Moody, & Schumm, 1998).

Batya Elbaum, Department of Teaching and Learning and Department of Psychology, University of Miami; Sharon Vaughn, Department of Special Education, University of Texas at Austin; Marie Tejero Hughes and Sally Watson Moody, Department of Teaching and Learning, University of Miami.

This research was supported by U.S. Department of Education, Office of Special Education Programs Grant H023E5005-96.

Correspondence concerning this article should be addressed to Batya Elbaum, School of Education, P.O. Box 248065, Coral Gables, Florida 33124. Electronic mail may be sent to elbaum@miami.edu.

In the 1970s and early 1980s, many schools adopted schoolwide tutoring programs for students with academic difficulties. In a meta-analysis of tutoring outcomes for elementary and secondary students, Cohen, Kulik, and Kulik (1982) wrote the following:

> The tutoring programs offered in many elementary and secondary schools today differ in an important way from yesterday's tutorial programs. In most modern programs, children are tutored by peers or paraprofessionals rather than by regular school teachers or professional tutors. The use of peer and paraprofessional tutors has dramatically affected the availability of tutoring programs. No longer a luxury available only to an aristocratic elite, tutoring programs today are open to boys and girls in ordinary classrooms throughout the country. (p. 237)

Students who were tutored by their classmates or by older students made greater academic gains than did untutored students (Cohen et al., 1982; Mathes & Fuchs, 1994).

During the 1980s, concern mounted over the high percentage of students who were not reading at grade level. Students who did not acquire basic reading skills in the early grades were shown to be at risk not only for school failure but also for negative outcomes beyond the school years (Karweit & Wasik, 1992; Kennedy, Birman, & Demaline, 1986). Systematic, one-to-one instruction by trained adults was advanced as a way of ensuring that all children would learn to read in the first years of elementary school.

Subsequently, educational decision makers recognized the limited extent to which effective, systematic, individual instruction could be provided by teachers in the context of their classrooms. As a result, schools invested in additional personnel to provide one-to-one instruction to students experiencing the greatest difficulty in reading. Depending on the program, the one-to-one instruction was provided by teachers (e.g., Clay, 1985), by paraprofessionals and volunteers (e.g., Invernizzi, Juel, & Rosemary,

1997), or by college students (e.g., Butler, 1991). The huge investment of resources that some of these programs required was justified by the belief that students who were given intensive, one-to-one instruction by trained adults before second grade would attain average levels of performance in reading and would need no further remedial help.

Wasik and Slavin (1993) conducted a review of five adult-delivered, one-to-one instructional programs in reading for first graders with reading difficulties: Reading Recovery, Success for All, Prevention of Learning Disabilities, the Wallach Tutoring Program, and Programmed Tutorial Reading. They concluded that all five programs yielded significant positive effects, with larger effects for those programs that used certified teachers rather than paraprofessionals.

Since Wasik and Slavin's (1993) review, many issues have been raised concerning not only the cost-effectiveness but the efficacy of programs involving adult, one-to-one tutoring of at-risk readers (Shanahan, 1998). For example, Hiebert (1994), Shanahan and Barr (1995), and Grossen, Coulter, and Ruggles (1997) have reviewed the available evidence on Reading Recovery interventions and have concluded that numerous flaws in the methodology used by proponents of Reading Recovery to evaluate and report intervention outcomes have resulted in inflated claims as to what the intervention achieves. Other researchers have implemented multitreatment studies to assess the contribution of additional components to the standard Reading Recovery intervention (Iversen & Tunmer, 1993). Yet others have investigated whether one-to-one interventions provide greater benefits than small-group interventions (Evans, 1996).

In the light of the extreme importance of ensuring that as many children as possible acquire adequate literacy skills in the early years of schooling and given the debate over the efficacy of one-to-one reading interventions for children at risk for reading failure, we undertook a rigorous meta-analysis of the empirical findings related to adult-delivered, one-to-one instructional interventions in reading for elementary school children identified as being at risk for reading failure. Cognizant of the methodological pitfalls of many of the primary studies in this area (for discussions, see Center, Wheldall, & Freeman, 1992; Grossen et al., 1997; Hiebert, 1994; Shanahan, 1998; Shanahan & Barr, 1995; Wasik & Slavin, 1993), we applied stringent parameters not only for the inclusion of studies in the synthesis (cf. White, 1994) but also for the inclusion of individual effect size comparisons in the meta-analysis.

Our goal in conducting this meta-analysis was to answer the following questions:

How effective are adult-delivered, one-to-one instructional interventions in reading for children at risk for reading failure? Decisions concerning the adoption and implementation of educational interventions are often based on projections—or, lacking data, on conjectures—about expected outcomes. We wished to provide educators and policymakers with a reasonable estimate of the gains (as measured immediately after an intervention) that students at risk for reading failure are likely to achieve as a result of participating in a one-to-one reading intervention.

How do key features of the intervention relate to intervention outcomes? Variables related to the intervention can affect not only its effectiveness but also its cost. These variables include the expertise of the individuals who implement the program, the training they undergo before beginning the intervention, the frequency of tutoring sessions, and the total hours of instruction provided to each student.

To what extent are variables related to studies' research methodology associated with study outcomes? An accurate interpretation of intervention outcomes requires an examination of the relation between effect size variation and methodological variables. Among those we examined are the method of assigning students to treatment groups, whether the researchers implemented a fidelity of treatment check, and whether the outcomes were assessed by means of standardized or nonstandardized tests.

How do the outcomes of Reading Recovery compare with outcomes produced by other interventions? Reading Recovery is the most widespread teacher-implemented, one-to-one intervention currently in use in schools in the United States. In this meta-analysis, we compared Reading Recovery with other one-to-one interventions designed to prevent or remediate reading failure in young children.

How do the outcomes of one-to-one reading interventions compare with those of small-group interventions? One-to-one interventions place severe practical limits on the number of students that can receive supplemental instruction. Despite the popular belief that one-to-one instruction is more effective than instruction delivered to larger numbers of students, there is actually little systematic evidence to support this belief. Each additional student that can be accommodated in an instructional group represents a substantial reduction in the per-student cost of the intervention, or, alternatively, a substantial increase in the number of students that can be served (cf. Hiebert, 1994; Shanahan & Barr, 1995).

Adult-delivered, one-to-one reading interventions for students at risk have achieved widespread currency in the United States. The present meta-analysis was designed to provide researchers, educators, and policymakers with information that would inform options and improve academic outcomes for elementary students who experience severe difficulty in reading.

Method

The design of this meta-analysis followed best practices for research synthesis as described by Cooper and Hedges (1994) and applied analytic procedures described by Glass, McGaw, and Smith (1981) and Cooper and Hedges (1994).

Literature Search

Key terms related to one-to-one instruction in reading were identified on the basis of previous research and from database thesauruses (e.g., the *Thesaurus of ERIC Descriptors* [Educational Research Information Center, 1995]). These key terms were used to conduct multiple computer and hand searches of the literature. Criteria for the inclusion of studies in the research synthesis were as follows: (a) The study was published or available between 1975 and 1998; (b) study participants were elementary students identified as at risk for reading failure, scoring in the lowest 20–30 percentile on grade level reading assessments, or possessing learning disabilities; (c) outcomes of students who received one-to-one instruction in reading were compared with those of students who exhibited comparably low performance in reading but did not receive one-to-one instruction in reading; and (d) outcome data amenable to the calculation of an effect size (e.g., means and standard deviations or *t* tests) were reported. The criterion regarding students in the control group was especially important in that findings from studies using a higher performing comparison group

have been seriously confounded by the phenomenon of regression to the mean (for a discussion of this issue, see Shanahan & Barr, 1995).

Coding of Studies

An extensive code sheet was used to record and organize pertinent information from each of the identified studies. The four authors and a research assistant, all of whom had experience coding studies for a previous synthesis, each coded a portion of the studies. Batya Elbaum reviewed all code sheets for completeness and accuracy. In the very small number of cases in which an assigned code was questionable, Elbaum consulted with the coder to resolve ambiguities and reach a decision by consensus.

Calculation of Effect Sizes

Standardized effect sizes, computed as the difference between the mean posttest score of the intervention group minus the mean posttest score of the control or comparison group divided by the standard deviation of the control or comparison group, were calculated for all reading outcomes for which means and standard deviations were available ($n = 221$). When only a test statistic such as t or F was available ($n = 8$), we applied formulas provided by Rosenthal (1994). In the 12 cases in which a statistical test of the difference between groups was reported as nonsignificant and no other data were provided, we assumed an effect size of 0.

Data Screening

The initial data set consisted of 241 effect sizes from 32 studies described in 31 reports. Because some studies reported outcomes for more than one group of students receiving one-to-one tutoring, the data set included effect size comparisons for 45 independent samples of students ($M = 0.91$, $SD = 1.77$, $Mdn = 0.56$).

Through visual inspection of the means, standard deviations, and score ranges for all outcome measures, we determined that floor effects were present in some effect size comparisons. That is, there were instances in which the mean score for a group was very close to the bottom of the score range. When a floor effect is present in a control group, the restricted variation in scores results in an underestimate of the population variation on the measure and, consequently, in a spuriously inflated effect size. Our decision was to exclude such comparisons from the meta-analysis, using an algorithm we developed specifically for this purpose.[1] A total of 22 effect size comparisons involving floor effects were excluded from the analysis.

Procedure for Handling Outliers

In the present meta-analysis, we applied Tukey's definition of extreme values, namely, those that are 3 or more interquartile ranges below the first quartile or above the third quartile (Cooper, Charlton, Valentine, & Muhlenbruck, 2000; Tukey, 1977). Six unweighted effect sizes (hereafter denoted by *ES*), all at the positive end of the distribution, had values exceeding the Tukey boundaries. Four extreme values were from the study by Iversen and Tunmer (1993; $ES = 4.59$ and 3.57 for one sample and 4.28 and 3.80 for the second sample); one extreme value ($ES = 4.36$) was from the study by Ramaswami (1994), and one ($ES = 4.81$) was from the study by Graves (1986). These six extreme values were winsorized, that is, they were set at a uniform maximum value. We used the value of 3.45, equal to 3 interquartile ranges beyond Tukey's upper hinge. Winsorization makes it possible to make maximum use of the available data while limiting the impact of extreme values in a distribution.

Given the weighting of within-sample effect sizes in computing a population estimate (Cooper, 1989)—a procedure that gives greater weight to effect sizes based on larger samples—meta-analytic findings are highly influenced not only by effect size outliers but also by sample size outliers (cf. Cooper et al., 2000). In the present meta-analysis, the mean intervention sample size was 36. Only 3 independent samples of students out of 44 (see Table 1) had intervention sample sizes greater than 63; these sample sizes were 96, 170, and 266. Using a somewhat more stringent version of Tukey's criterion (2 instead of 3 interquartile ranges beyond the upper hinge), we winsorized these sample sizes to 80. The mean size of comparison groups was 44; 4 very high values (99, 138, 165, and 217) were similarly set to the value of 80. Sample size was therefore winsorized for a total of 6 independent samples of students, affecting 24 individual effect size comparisons.

Calculation of Mean Weighted Effect Sizes

Mean weighted effect sizes (for the weighting formula, see Cooper, 1999, p. 137) were calculated for various aggregations of effect sizes to examine particular substantive and methodological variables of interest. Following Cooper (1989), we used a shifting unit of analysis to ensure the independence of data in a given aggregation. Thus, effect sizes from multiple measures administered to the same group of students were averaged to yield a single effect size for that sample of students. For the analysis by outcome measure, each sample of students contributed only one effect size for each type of measure.

In meta-analysis, a weighted least squares analysis analogous to a multifactorial analysis of variance (Hedges & Olkin, 1985) is typically used to test the significance of main effects and interactions among the independent variables. However, the highly unbalanced distribution of cases across levels of the individual variables and the presence of empty cells in the multivariate design made it impossible to conduct an interpretable multivariate analysis. The significance of the independent variables was therefore tested in a series of single-factor homogeneity tests (Cooper & Hedges, 1994). The test statistic Q (subscripted Q_W to refer to the homogeneity of effect size comparisons within an aggregation and Q_B to refer to the homogeneity of categories of a moderator variable) is distributed as chi-square with degrees of freedom equal to $k - 1$, where k denotes the number of independent effect sizes in an aggregation, for Q_W, and one less than the number of categories of the variable, for Q_B. Values of Q are reported to be significant at $p < .05$.

---

[1] One indication of a floor effect in the control group is a large discrepancy in the standard deviations of the intervention and control groups. We calculated the ratio of the treatment standard deviation and control group standard deviation and found that the distribution of this ratio was $M = 1.07$, $SD = 0.87$, $Mdn = 0.93$. The mean value near 1 indicated that on average, the standard deviations for intervention and control groups were very similar. We determined that values exceeding 1.83 (1.5 times the interquartile range beyond the 75th percentile) would be considered extreme (this is the algorithm used, for example, by the SPSS [1999] Explore procedure). A total of 17 such values were identified. However, it is possible for floor effects to be present even if standard deviations are comparable. This occurs when the means of both the treatment and control groups are close to the bottom of the range of the measure. In such cases, the effect size is not meaningful because the measure used is apparently not sensitive enough to detect differences in outcomes at the level at which students are performing. To capture these cases, we identified all effect size comparisons in which the mean of the control group was in the bottom 15% of the range for the particular measure. Eleven effect size comparisons were so identified, including 6 that had also been identified as having extreme values with regard to the ratio of the groups' standard deviations. A total of 22 effect size comparisons, involving 11 samples of students from 8 studies, were therefore excluded from the analysis. The exclusion of these cases resulted in the elimination of one study (Towner & Davidson, 1998) that had been represented by a single effect size. The 22 excluded effect sizes had a mean of 3.75 and a standard deviation of 3.87; the 219 effect sizes retained in the analysis had a mean of 0.59 and a standard deviation of 0.99.

Table 1
Summary Information on Interventions Included in the Meta-Analysis by Sample

Study	Grade level	Intervention sample size	Instructor	Focus of instruction	*n* of within-sample *ES*	Mean within-sample *ES*
Arnold et al. (1977)	1	23	No information	V-P skills	2	0.00
Butler (1991)	4–6	20	College students	Mixed	4	0.75
Center, Wheldall, Freeman, Outhred, and McNaught (1995)[a]	1	22	Teachers	Mixed	4	1.08
Chapman, Tunmer, and Prochnow: Sample 1 (1998)[a]	1	26	Teachers	Mixed	8	−0.43
Chapman, Tunmer, and Prochnow: Sample 2 (1998)[a]	1	6	Teachers	Mixed	8	−1.32
Compton (1992)[a]	1	80[b]	College students	Mixed	1	1.83
Dorval, Wallach, and Wallach (1978)	1	20	Paraprofessionals	PA-phonics	1	0.68
Graves: Sample 1 (1986)	4–6	8	Teachers	Comprehension	2	1.85
Graves: Sample 2 (1986)	4–6	8	Teachers	Comprehension	2	3.34
Hagin, Silver, and Beecher (1978)	Range	63	Teachers	Mixed	3	0.49
Hedrick (1996)	1	80[b]	Teachers	Mixed	1	−0.01
Iversen and Tunmer: Sample 1 (1993)[a]	1	32	Teachers	Mixed	6	2.46
Iversen and Tunmer: Sample 2 (1993)[a]	1	32	Teachers	Mixed	6	2.53
Juel (1996)	2–3	6	College students	Mixed	1	3.15
Knapp and Winsor (1998)	2–3	8	Teachers	Underspecified	2	0.71
Lafave (1995)[a]	1	23	Teachers	Mixed	6	0.36
Mantzicopoulos, Morrison, Stone, and Setrakian: Sample 1 (1992)	1	59	Teachers	V-P skills	6	0.05
Mantzicopoulos, Morrison, Stone, and Setrakian: Sample 2 (1992)	1	52	Teachers	PA-phonics	6	0.09
McCarthy, Newby, and Recht (1995)	1	19	Teachers	Mixed	15	0.68
McGrady (1994)	4–6	35	No information	Mixed	1	−0.37
Morris, Shaw, and Perney (1990)	2–3	30	Volunteers	Mixed	6	0.52
Nielson (1991)	2–3	14	Volunteers	Mixed	1	0.38
Pinnell: Sample 1 (1988)[a]	1	80[b]	Teachers	Mixed	9	0.65
Pinnell: Sample 2 (1988)[a]	1	37	Teachers	Mixed	9	0.58
Pinnell, Lyons, DeFord, Bryk, and Seltzer: Sample 1 (1994)[a]	1	31	Teachers	Mixed	4	0.74
Pinnell, Lyons, DeFord, Bryk, and Seltzer: Sample 2 (1994)[a]	1	38	Teachers	Mixed	4	0.13
Pinnell, Lyons, DeFord, Bryk, and Seltzer: Sample 3 (1994)	1	30	Teachers	Mixed	4	−0.05
Ramaswami: Sample 1 (1994)[a]	1	12	Teachers	Mixed	5	2.17
Ramaswami: Sample 2 (1994)[a]	1	5	Teachers	Mixed	5	1.20
Ramaswami: Sample 3 (1994)[a]	1	19	Teachers	Mixed	6	0.35
Ramaswami: Sample 4 (1994)[a]	1	13	Teachers	Mixed	6	−0.37
Ramey: Sample 1 (1991)	2–3	18	Volunteers	Underspecified	1	0.00
Ramey: Sample 2 (1991)	4–6	59	Volunteers	Underspecified	1	−0.25
Saginaw Public Schools (1992)[a]	1	35	Teachers	Mixed	6	0.92
Torgeson et al.: Sample 1 (1998)	2–3	33	Teachers	PA-phonics	8	0.68
Torgeson et al.: Sample 2 (1998)	2–3	36	Teachers	Mixed	8	0.16
Torgeson et al.: Sample 3 (1998)	2–3	37	Teachers	Underspecified	8	0.05
Vadasy, Jenkins, Antil, Wayne, and O'Connor: Sample 1 (1997)	1	6	Volunteers	Mixed	11	0.85
Vadasy, Jenkins, Antil, Wayne, and O'Connor: Sample 2 (1997)	1	14	Volunteers	Mixed	11	0.06
Vadasy, Jenkins, and Pool (1998)	1	23	Volunteers	Mixed	9	0.98
Wallach and Wallach (1976)	1	36	Volunteers	PA-phonics	6	0.67
Weeks (1992)[a]	1	20	Teachers	Mixed	3	−0.35

Note. *ES* = unweighted effect size; V-P = visual–perceptual skills; PA-phonics = phonemic awareness–phonics.
[a] Reading Recovery intervention. [b] Winsorized sample size.

Results

Disaggregation of Studies by Research Design

Twenty-nine of the 31 studies that met our search criteria contrasted one or more groups of students who participated in a supplemental one-to-one instructional intervention in reading with a group of students who did not receive any one-to-one instruction. Although students in some of the control groups received supplemental academic support through federally funded after-school programs, the programs did not include any systematic one-to-one tutoring in reading. Two studies, comprising three effect size comparisons, contrasted outcomes for students participating in a one-to-one reading intervention with outcomes for students participating in a small-group reading intervention. Because the effect sizes of these latter studies must be interpreted differently than effect sizes that contrast a one-to-one instructional intervention with a control group, we analyzed the treatment–comparison studies separately from the treatment–control studies.

The Treatment–Control Database

This database consisted of effect sizes from 29 treatment–control studies, described in 28 separate reports. Sources were 14

published articles, 5 doctoral dissertations, 4 technical reports, 1 conference presentation, 1 book, and 3 manuscripts prepared for submission to professional journals. Two studies were conducted in the late 1970s, 5 in the 1980s, and the remainder in the 1990s. The students in 2 studies (3 samples of students) had learning disabilities; students in all other studies (39 samples) had no identified disability but were identified as at risk for reading difficulties. The preponderance of students represented in the present synthesis were first graders. Included in the synthesis were 28 samples of first graders (n = 1,164), 8 samples of students in Grades 2 or 3 (n = 182), 5 samples of students in Grades 4–6 (n = 130), and 1 sample of students ranging from 1st through 4th grade (n = 63). Summary information on the interventions is provided in Table 1; a description of intervention procedures is provided in Table 2.

The treatment–control studies yielded a total of 216 individual effect size comparisons. When the individual effect sizes were aggregated by independent sample, so that each sample of students contributed a single, averaged effect size for that group of students, the distribution of the 42 effect sizes was somewhat positively skewed (skewness = 2.85), M = 0.67, SD = 0.98, Mdn = 0.55. Six effect sizes were markedly negative (−1.32 to −0.25); 10 were very small or close to 0 (−0.12 to 0.16); 14 were moderately large (0.35 to 0.75); 5 were large (0.85 to 1.20); and 7 were very large (1.83 to 3.34).[2] The mean weighted effect size was 0.41. The various disaggregations of effect sizes, with their accompanying weighted means and homogeneity statistics, are presented in Table 3.

Qualifications of Instructors, Training, and Treatment Fidelity

The interventions included in the meta-analysis were conducted by individuals with varying qualifications for teaching. The significant homogeneity test associated with the instructor variable, $Q_B(4)$ = 77.05, indicated that the variation in effect sizes was significantly associated with the qualifications of the instructor. The tutors whose students made the greatest gains as a result of one-to-one instruction were college students (d = 1.65, k = 3). The mean weighted effect sizes for teachers and community volunteers were 0.36 (k = 28) and 0.26 (k = 8), respectively. The effect size for the single sample of students taught by paraprofessionals was 0.68. Two studies provided no information on the individuals carrying out the intervention. The mean weighted effect size associated with these studies was −0.20.

Six studies investigated outcomes of one-to-one interventions in which community volunteers served as the tutors. Five of the 6 studies described the training that was provided to tutors; 1 study did not specify whether the tutors received any training. The mean weighted effect size associated with studies that described the tutors' training was 0.59 (k = 6), compared with −0.17 (k = 2) for the study that did not indicate whether the volunteer tutors received training.

Because treatment fidelity is a particular concern with regard to volunteers, Vadasy, Jenkins, Antil, Wayne, and O'Connor (1997) explicitly contrasted the outcomes of two subgroups of volunteer tutors: those who maintained treatment fidelity (e.g., came to each tutoring session, tutored for the full amount of time) and those who did not. The effect size associated with consistent tutors was 0.85; that associated with inconsistent tutors was 0.06. If volunteers who performed inconsistently were excluded from the aggregation of volunteers who were known to have received training before the intervention, the resulting mean weighted effect size for volunteers was 0.67 (k = 5).

Students' Grade Level

The significant homogeneity statistic for this variable, $Q_B(3)$ = 9.27, indicated that students' grade level was significantly associated with the variation in effect sizes. Mean effects for all except the oldest students were in the moderate range (d = 0.37–0.49). The mean effect of one-to-one instruction for students in Grades 4–6 was not significantly different from 0. Note that as indicated by the significant Q value associated with the effect size estimate for older students, $Q_W(4)$ = 37.60, the five within-sample effect sizes constituting this aggregation were quite disparate. Effect sizes from the studies by Butler (1991; ES = 0.75) and Graves (1986; ES = 1.85 and 3.34) were very high, whereas those from the studies by Ramey (1991; ES = −0.25) and McGrady (1984; ES = −0.37) were negative (see footnote 2 for a further treatment of negative effect sizes). The variation in effects for first graders is examined in a subsequent analysis (see below).

Focus of the Intervention

Interventions were coded in terms of their primary instructional focus. The categories used for this variable were (a) decoding–word recognition, (b) comprehension, (c) mixed (a combination of

[2] The effect sizes in the present meta-analysis include six with moderate to large negative values. The occurrence of negative values is problematic in that it is difficult to imagine how implementation of a one-to-one intervention could lower outcomes for participating students. Two of the six samples of students in this category were students who failed to successfully complete the Reading Recovery program. The largest negative effect size was for the small sample (n = 6) of students in the study by Chapman et al. (1998) who did not successfully complete the program and were "referred on for additional remedial reading assistance" (p. 7). The other negative effect size for a not-discontinued sample was from Ramaswami (1994). The negative effect sizes for two other samples of students might be due to the nonequivalence of treatment and comparison groups at pretest. Weeks (1992) reported that the Reading Recovery group performed worse at pretest on several reading measures than did the Supported Control group. Fall to spring gains were approximately equivalent for the two groups. Thus, had initial differences been statistically controlled, the effect size for the sample would have been close to zero. McGrady (1984) reported that students in the study were not randomly assigned and that the treatment and control groups differed significantly at pretest. The intervention, according to the author, narrowed the gap between the groups. Two samples of students receiving one-to-one instruction, from studies by Ramey (1991) and Chapman et al. (1998), did not perform as well as students in the comparison group. In the study by Ramey, students in the tutoring program were compared with students receiving instruction in a Reading Resource Specialist model, which is not described in the report. In the study by Chapman et al. (1998), students participating in Reading Recovery did not do as well as students in the comparison group. According to the authors, students in the Reading Recovery group did not demonstrate average grade-level performance either immediately after the intervention or 6 or 12 months later.

Table 2
Description of Interventions

Study	Intervention procedures
	Treatment–control studies[a]
Arnold et al. (1977)	*Channel-specific perceptual stimulation (ES = 0.00):* Students were tutored for 30 min twice a week for 6 months. The tutoring was based on stimulation of the specific deficits noted in the students' perceptual profile. Techniques developed by Silver and Hagin (1976) were used to train out perceptual deficits in visual, auditory, kinesthetic, haptic, and body-image modalities. *Control group:* Regular classroom instruction.
Butler (1991)	*Reading Assistance Tutorial (R.A.T.) Pack (ES = 0.75):* Children were tutored by university students for 20 min 3 times a week for 8 weeks. The R.A.T. Pack is a psycholinguistic and social semiotic approach to literacy that aims at developing students' phonological processing strategies, linguistic awareness, and sight word vocabulary. Functional language use is promoted through sentence construction, cloze passages, puzzles, games, and creative manipulations of the surface features of language. The R.A.T. Pack consists of 12 books varying in difficulty from early sounds through vocabulary development and comprehension. *Control group:* Regular classroom instruction.
Center et al. (1995)	*Reading Recovery (ES = 1.08):* Students were tutored for 30 min daily for an average of 15 weeks. Each tutoring session consisted of the following components: (a) rereading of two or more familiar books, (b) independent reading of the previous day's new book while the teacher took a running record (miscue analysis), (c) letter identification (if needed), (d) writing a story the child composed, with emphasis on hearing sounds in words, (e) reassembling a cut-up story, (f) introducing a new book, and (g) reading the new book. *Control group:* Regular classroom instruction plus any support in reading typically available at the school. Remedial assistance consisted of up to 2 hr per week of additional instruction.
Chapman et al. (1998)	*Reading Recovery[b] (successfully discontinued; ES = −0.43). Reading Recovery (not discontinued and referred on for remedial services; ES = −1.32). Control group:* Regular classroom instruction.
Compton (1992)	*Reading Connection (ES = 1.83):* Students were tutored 30 min a day, 4 days a week, for 14 weeks by university students using the Reading Recovery[b] method. *Control group:* Regular classroom instruction plus 30 min of instruction daily through Chapter 1 for 14 weeks.
Dorval, Wallach, and Wallach (1978)	*Wallach and Wallach (ES = 0.68):* Students were tutored 30 min daily for 28 weeks using the Wallach and Wallach program designed to teach phonemic awareness. *Control group:* Regular classroom instruction.
Graves (1986)	*Direct instruction* (ES = 1.85): Over eight individual tutoring sessions, students were taught to find the main idea of stories using techniques described by Carnine and Silbert (1979). *Direct instruction plus metacomprehension (ES = 3.34):* The metacomprehension training, based on Loper (1980), emphasized self-monitoring as a way of recording one's progress during instructional tasks. *Control group:* Students were prompted to read stories and answer questions at the end about the main ideas.
Hagin, Silver, and Beecher (1978)	*TEACH (ES = 0.49):* Students were tutored three to four times per week using a perceptual stimulation approach aimed at developing the accuracy of students' perceptions within single modalities and across modalities. The learning tasks proceeded through three stages: recognition, copying, and recall. *Control group:* Regular classroom instruction.
Hedrick (1996)	*ICARE (ES = −0.01):* Students were tutored 30 min daily for one semester. The intervention was based on Reading Recovery[b] with the addition of a phonemic awareness component. The ICARE Program also required parents to read to the student for 15 min each night. *Control group:* Regular classroom instruction plus any other available support services.
Iversen and Tunmer (1993)	*Standard Reading Recovery[b] (ES = 2.46):* Students were tutored for 30 min four times per week for an average of 10.5 weeks. *Modified Reading Recovery (ES = 2.53):* Students were tutored for 30 min four times per week for an average of 14.3 weeks. The lessons involved the seven standard Reading Recovery activities; however, explicit instruction in the letter–phoneme patterns took the place of the letter-identification segment when the children demonstrated that they could identify at least 35 of the 54 alphabetic characters. The explicit training in phonological skills involved asking the students to manipulate magnetic letters to make, break, and build new words having similar visual and phonological elements. The teacher chose suitable words from one of the books the student had read earlier in the lesson or from a list of frequently occurring words. *Control group:* Students received whatever additional support (generally funded by Chapter 1) was normally offered to at-risk readers at their schools. Typical support consisted of out-of-class, small-group (6–7 students) instruction four times a week.
Juel (1996)	*Literacy tutoring (ES = 3.15):* Students were tutored by university student athletes for 45 min two times a week for 1 school year. Tutoring sessions consisted of three or four of the following activities: (a) reading children's literature, (b) writing, (c) introducing high-frequency words from the basal readers, (d) journal (the tutor wrote words, and the child copied), (e) alphabet book (the child selected words to add for each letter), (f) hearing word sounds (phonemic awareness), and (g) letter–sound activities. *Control group:* Students received mentoring from the same university students in regularly scheduled weekly meetings. Mentoring included reading to the students (but not other tutoring activities) or reading and talking with the student in the school library or on the playground.

Table 2 (*continued*)

Study	Intervention procedures
Knapp and Winsor (1998)	*Cognitive apprenticeship in reading (ES = 0.71):* Students were tutored three times a week for 10 weeks by adult volunteers. The program was based on the cognitive apprenticeship model explicated by Collins, Brown, and Holum (1991). The student and tutor first read a book of the student's choice, alternately reading aloud to each other and commenting on what was read. The tutor modeled reading strategies and fluent reading, helped with the decoding of difficult words, and offered questions and explanations to clarify text meaning. Students selected personally interesting books and were allowed to discontinue reading any books they found uninteresting or too difficult. *Control group:* Regular classroom instruction.
Lafave (1995)	*Reading Recovery[b] (ES = 0.36):* Students were tutored for 30 min daily for 5 months. *Control group:* Regular classroom instruction plus additional instruction through Chapter 1 consisting of small-group (3–5 students) instruction 5 days a week for 5 months. Chapter 1 sessions included instruction in reading aloud, letter–sound relationships, word families, and writing stories.
Mantzicopoulos, Morrison, Stone, and Setrakian (1992)	*TEACH (ES = 0.05):* Students were tutored for 30 min twice a week for 25 weeks. The intervention consisted of 55 teaching activities organized into 5 clusters: visual (e.g., visual discrimination, visual sequencing), visual–motor, auditory (e.g., recognition of rhyming words, ordering, blending), body image (left-to-right progression of reading), and intermodal (e.g., matching sounds to visual symbols). *Phonetic tutoring (ES = 0.09):* Students received 20 min of reading drill and 10 min of spelling drill per session. Students were expected to read lists of words as quickly and accurately as possible within a set time period, with the tutor keeping count of the number of words read correctly. *Control group:* Regular classroom instruction.
McCarthy, Newby, and Recht (1995)	*Early Intervention Program (ES = 0.68):* Students were tutored for 30 min daily using a program based on Reading Recovery.[b] Each tutoring session involved three 10-min segments: (a) students reread books they had covered in previous lessons, (b) students wrote a message of their own composition in standard spelling, with explicit instruction from the tutor in sound segmentation and relations with the alphabetic code, (c) tutors presented new reading material using a guided-reading format. Phonological training involved two strategies: the Elkonin "boxes" strategy (the student slowly articulated the sounds in a word sequentially while manipulating corresponding counters) and the "stretch it out" strategy (the student slowly articulated sounds in a word while choosing the appropriate alphabetic symbols to represent sounds). *Control group:* Regular classroom instruction.
McGrady (1984)	*Programed [sic] tutoring (ES = −0.37):* Students were tutored for 15 min a day for 1 school year using the Houghton Mifflin Tutorials, a set of materials designed to be used as a supplement to classroom teaching based on the Houghton Mifflin Reading Series. Activities included oral reading, comprehension, and word attack. *Control group:* Regular classroom instruction plus available Chapter 1 services.
Morris, Shaw, and Perney (1990)	*Tutoring program (ES = 0.52):* Students were tutored 1 hr a day, 4 days a week, for 8 months by adult volunteers. The premise of the tutoring approach was that children who are having difficulty learning to read need the semantic and syntactic support offered by good stories written in natural (as opposed to formulaic) language and that children should be led to automatize basic one-syllable spelling patterns as a means of building word knowledge. Tutoring involved (a) 15–20 min of easy contextualized reading at the student's instructional level, (b) 10–12 min of word study, (c) 15 min of writing, (d) 10–15 min easy reading in trade books, (e) 5–10 min reading the student a good piece of literature, for example, a fairy tale, fable, short picture book, or chapter from a longer book. *Control group:* Regular classroom instruction.
Nielson (1991)	*Tutoring (ES = 0.38):* Students were tutored over a period of 9 months by adult volunteers. Students did oral reading and were drilled on vocabulary items that were missed during the reading; tutors also provided some instruction in improving reading comprehension. Tutors maintained a log of voluntary home reading, oral in-school reading, and flash and card drill. Students were given a point for each activity and for each sentence read correctly the first time. At the end of each month, points could be exchanged for prizes such as pencils, erasers, balls, books, and school logo shirts. *Control group:* Regular classroom instruction.
Pinnell (1988)	*Reading Recovery[b] (ES = 0.65):* Students received the standard Reading Recovery intervention and were in classrooms whose teachers were trained in Reading Recovery. The students therefore received group as well as individual instruction using the Reading Recovery approach. *Reading Recovery (ES = 0.58):* Students received the standard Reading Recovery intervention; however, their regular classroom teachers were not trained in Reading Recovery. *Control group:* Regular classroom instruction plus a compensatory program involving skill-oriented and drill activities conducted by a paraprofessional.
Pinnell et al. (1994)	*Reading Recovery[b] (ES = 0.74):* Students were tutored for 30 min daily for 5 months. *Reading Success (ES = 0.13):* Students were tutored for 30 min daily for 5 months using a program modeled on Reading Recovery and taught by certified teachers who received a condensed 2-week version of the Reading Recovery training. *Direct Instruction Skills Plan (ES = −0.05):* Students were tutored for 30 min daily for 5 months using a program focusing on systematic instruction in skills considered to be basic to the performance of reading tasks. For each child, tutorial sessions were linked to the classroom instruction the child was receiving. Lessons included work on letters and sounds, words, and text-level strategies such as sequencing, filling in the blanks, and answering questions, as well as reading extended texts. *Reading and Writing Group (ES compared with one-to-one interventions = 0.12):* Students received Reading Recovery-based tutoring in a small-group format. The teachers, who had been trained as Reading Recovery teachers, could modify Reading Recovery procedures to adjust to group instruction and could adopt any techniques they believed to be consistent with the theoretical base developed during their training. *Control group:* Regular classroom instruction plus any existing Chapter 1 services for first graders.

Table 2 (*continued*)

Study	Intervention procedures
Ramaswami (1994)	1991–1992 cohort: *Reading Recovery*[b] *(successfully discontinued; ES = 2.17). Reading Recovery (not discontinued; ES = 1.20). Control group:* Regular classroom instruction plus compensatory instruction for students who qualified. 1992–1993 cohort: *Reading Recovery*[b] *(successfully discontinued; ES = 0.35). Reading Recovery (not discontinued; ES = −0.37). Control group:* Regular classroom instruction plus compensatory instruction for students who qualified.
Ramey (1991)	*HOSTS (Helping One Student to Succeed; Grades 2 and 3; ES = 0.00):* Students were tutored by community volunteers for 1 year. *HOSTS (Grades 4–5; ES = −0.25):* Same. *Control group:* Students received compensatory education for 1 year through a traditional pull-out approach.
Saginaw Public Schools (1992)	*Reading Recovery*[b] *(ES = 0.92). Control group:* Regular classroom instruction.
Torgesen et al. (1998)	Students in all tutoring conditions were tutored for 20 min, four times a week for 2½ years beginning in the second semester of kindergarten. *Phonological Awareness Plus Synthesis Phonics (ES = 0.68):* Tutors provided implicit instruction in phonemic awareness by leading students to discover and label auditory gestures associated with each phoneme. Then students were engaged in activities to build skills in tracking sounds in words and to represent sounds with letters. Students learned to spell syllables with letters and then to read syllables by blending separate phonemes together. Students then read short stories containing the words they could decode. During second grade, children received direct fluency-building practice and were taught strategies for multisyllabic words. *Embedded Phonics (ES = 0.16):* Tutoring consisted of (a) learning to recognize small groups of whole words by using word-level drill and word games, (b) instruction in letter–sound correspondences in the context of the sight words being learned, (c) writing the words in sentences, and (d) reading the sentences that were written. Stimulation of phonological awareness was done during writing activities in which students were asked to identify the sounds in words before writing them. Most grapheme–phoneme correspondences were taught in the context of word reading and writing activities. Basal readers were also used. *Regular classroom support (ES = 0.05):* Students received tutoring in the activities and skills taught in their regular classroom reading programs. The activities varied from phonics-oriented activities to sight word drill to writing in journals. *Control group:* Regular classroom instruction.
Vadasy et al. (1997)	*Tutoring by community volunteers (high treatment fidelity; ES = 0.85):* Students received 30 min of tutoring 4 days a week for up to 23 weeks. Each lesson included six to eight activities selected from the following: (a) letter sounds and beginning sound instruction, (b) rhyming, (c) auditory blending, (d) segmenting, (e) spelling and analogy use, (f) story reading, and (g) writing. *Tutoring by community volunteers (low treatment fidelity; ES = 0.06):* Same. *Control group:* Regular classroom instruction.
Vadasy, Jenkins, and Pool (1998)	*Tutoring (ES = 0.98):* Students were tutored for 30 min, 4 days a week, for the school year. Tutoring included instruction in phonological skills, letter–sound correspondence, explicit decoding, rime analysis, writing, spelling, and reading phonetically controlled text. *Control group:* Regular classroom instruction plus Title 1 services where available.
Wallach and Wallach (1976)	*Wallach and Wallach tutorial program (ES = 0.67):* Students were tutored for 30 min, five times a week, for 30 weeks by community volunteers. First, students were taught to recognize sounds at the start of words, to recognize the shape of letters, and to connect letter shapes with sounds. Second, students gained skill at recognizing and manipulating the sounds in words and blending sounds in the context of short, regularly spelled words. Third, students practiced applying previously acquired skills using regular classroom reading materials. *Control group:* Regular classroom instruction.
Weeks (1992)	*Reading Recovery*[b] *(ES = −0.35):* Students were tutored 5 days a week for a maximum of 12 weeks. *Control group:* Students received regular classroom instruction from teachers who were participating in a 7-month in-service program similar to that of the Reading Recovery teacher training and emphasizing whole language literacy instruction.
	Treatment–comparison studies[c]
Acalin (1995)	*Reading Recovery*[b] *(ES = −0.12):* Students were tutored 30 min daily for 1 school year. *Comparison group:* Project READ, based on the Orton–Gillingham method emphasizing phonics. Students received instruction in groups of 2 to 5, 30 min daily, for 1 school year. Lessons progressed from a focus on phonology to comprehension and then to writing, with an emphasis on instruction in vowel and consonant sounds, blends, and word syllabication. Each lesson introduced five to ten new words, with sentence complexity increasing over the sequences of lessons.
Evans (1996)	*Reading Recovery*[b] *(ES = 0.05):* Students received tutoring for 16 weeks. *Comparison group:* Students received small-group (4 students) instruction in the regular classroom for 30 min daily, for 16 weeks. The instruction was based on the principles of Reading Recovery and included the following components: (a) independent reading, (b) shared reading, (c) shared journal, and (d) introduction to the new text.

Note. *ES* = unweighted effect size.

[a] Studies compared students who received supplementary one-to-one tutoring in reading with students who received no supplementary one-to-one or other systematic instructional intervention in reading but may have received additional support through Chapter 1 or Title 1 programs. [b] For a description of the standard Reading Recovery intervention, see table entry under Center et al. (1995). [c] Studies compared students who received one-to-one tutoring in reading with students who received a systematic, small-group reading intervention.

decoding, word recognition, and comprehension), (d) phonemic awareness–phonics, (e) visual–perceptual skills, and (f) underspecified (not sufficiently well described to be coded). The majority of interventions were coded as mixed; these interventions accounted for 30 samples of students, including the 16 samples of students that received Reading Recovery instruction. The distribution of samples across the remaining categories was reading comprehension (two samples), phonemic awareness–phonics (four samples), visual–perceptual skills (two samples), and underspecified (four samples). The significant homogeneity statistic, $Q_B(4) = 42.44$, indicated that focus of instruction was reliably associated with the variation in effect sizes. The focus associated with the largest effect ($d = 2.41$) was reading comprehension; this effect was derived from two interventions that used direct instruction to improve the comprehension of upper elementary students with learning disabilities. Interventions that had a mixed focus or a focus on phonemic awareness–phonics yielded mean weighted effect sizes in the moderate range ($d = 0.50$ and 0.43, respectively). Interventions that focused on visual–perceptual skills and those that were not adequately described in reports had mean weighted effect sizes close to 0.

Outcome Measures

To investigate whether the aspect of reading measured by the outcome measure was significantly associated with the variation in effect sizes, individual effect sizes were aggregated by measure type within independent samples. Thus, each sample of students contributed one effect size for each type of measure used to assess outcomes. The significant homogeneity statistic, $Q_B(9) = 54.81$, indicated that the variation in effect sizes was significantly associated with the aspect of reading or language that was measured after the intervention. Modest effects were found with measures of reading comprehension ($d = 0.28$) and spelling ($d = 0.14$). Measures of decoding, oral reading of words, oral reading of passages, composites based on subtests of different skills, and writing yielded moderate effects ($d = 0.41$–0.54). The single listening comprehension outcome in the corpus had an effect size of 0.68. Writing vocabulary, as measured by the Clay Writing Vocabulary Test, produced the largest effects ($d = 0.94$). The only mean effect size of negative valence was for measures of phonemic awareness ($d = -0.29$).

Within the category of oral reading of passages, outcomes measured by the Text Reading Level measure (Clay, 1985) accounted for half of the 18 effect size comparisons. This measure has been criticized as having poor psychometric properties, in that growth between levels is much smaller at the lower end of the scale than at the higher end (cf. Iversen & Tunmer, 1993). When effect sizes produced by the Text Reading Level measure (Clay, 1985; $d = 0.64$) were contrasted with effect sizes produced by other measures of oral reading of passages ($d = 0.30$), the difference was statistically significant, $Q_B(1) = 7.23$.

Standardized measures generally yield smaller effects than nonstandardized measures, because the latter are typically more closely aligned with particular interventions. We coded each outcome measure as standardized or not standardized; measures coded as standardized had to use a standard set of stimulus materials, a standard administration procedure, and a standard scoring procedure and be supported by norming information. In 12 studies ($k = 14$), outcomes were assessed by means of standardized measures only; in 5 studies ($k = 7$), outcomes were assessed by means of nonstandardized measures only; 12 studies ($k = 21$) used both types of measures. To conduct the most stringent test of differences owing to whether the outcome was assessed by means of a standardized measure, we compared effect sizes associated with standardized and nonstandardized measures for the 21 samples of students (11 Reading Recovery and 10 other interventions) for which both types of measures were used. The difference between the mean weighted effect sizes calculated for the two types of measures was not statistically significant, $Q_B(1) = 1.09$. Reading Recovery and other intervention samples were also considered separately, to examine whether the difference between effect size estimates based on standardized and nonstandardized measures was significant for either group alone. For samples receiving interventions other than Reading Recovery, the mean weighted effect sizes for standardized and nonstandardized measures were almost identical ($d = 0.46$ vs. 0.42, respectively). For Reading Recovery samples, the mean weighted effect size for standardized measures was less than that for nonstandardized measures ($d = 0.60$ vs. 0.80, respectively), but the difference was not statistically reliable.

Intervention Intensity

Intervention intensity was examined in two ways: by duration, coded as the number of weeks over which the intervention was carried out, and total instructional time, coded as the number of hours of instruction provided to each student. Information on the duration of the intervention was available for 30 samples of students; information on total instructional time was available for 27 samples. The interventions ranged in duration from 8 to 90 weeks and in total instructional time from 8 to 150 hr. Duration of the intervention was reliably associated with the variation in effect sizes, $Q_B(1) = 7.9$; interventions lasting up to 20 weeks had a mean weighted effect size of 0.65, compared with 0.37 for those lasting longer than 20 weeks. Total instructional time, however, was not reliably associated with effect size variation, $Q_B(1) = 0.35$.

We further examined the relation between intervention duration and intensity. The mean instructional time for interventions lasting up to 20 weeks was 63 hr; the mean time for interventions lasting longer than 20 weeks was 61 hr. Duration and total instructional time did not significantly covary ($r = .116$, *ns*). This finding suggested that the same amount of instructional time, delivered more intensively, tends to have more powerful effects.

Methodological Variables

Two methodological variables were examined for their potential impact on study outcomes. Homogeneity tests revealed that the use of a check on the fidelity of treatment was not significantly associated with the variation in effect sizes, $Q_B(1) = 0.42$. In contrast, the method that researchers used to assign students to treatments was reliably associated with such variation, $Q_B(1) = 20.24$, so that studies that used random assignment or matching yielded significantly higher effect sizes ($d = 0.56$) than studies that used other procedures (e.g., teacher judgment, convenience; $d = 0.17$).

Table 3
Meta-Analysis

Aggregation	*k*	*ES*	*d*	95% CI for *d* Lower	95% CI for *d* Upper	Q_w	Q_B
Samples compared with a control	42	0.67	0.41	0.32	0.49	248.80*	
Instructor							77.05*
Teachers	28	0.67	0.36	0.26	0.46	136.91*	
Community volunteers	8	0.40	0.26	0.07	0.45	20.34*	
College students	3	1.91	1.65	1.34	1.96	13.41*	
Paraprofessionals	1	0.68	0.68	0.04	1.32	0.00	
No information	2	−0.19	−0.20	−0.55	0.15	1.09	
Training of volunteers							5.20*
Samples tutored by volunteers	8	0.40	0.26	0.07	0.45	20.34*	
Volunteers were trained	6	0.58	0.59	0.33	0.84	4.63	
Training not reported	2	−0.13	−0.17	−0.46	0.12	0.59	
Grade level							9.27*
1	28	0.59	0.46	0.36	0.56	178.16*	
2–3	8	0.71	0.37	0.16	0.57	23.77*	
4–6	5	1.06	0.06	−0.19	0.30	37.60*	
Range	1	0.49	0.49	0.14	0.85	0.00	
Focus of instruction							42.44*
Mixed	30	0.67	0.50	0.40	0.60	194.77*	
Reading comprehension	2	2.59	2.41	1.48	3.33	2.34	
Phonemic awareness–phonics	4	0.53	0.44	0.21	0.68	5.64	
Visual–perceptual skills	2	0.03	0.03	−0.26	0.33	0.02	
Underspecified	4	0.13	−0.07	−0.30	0.17	3.59	
Type of outcome measure							54.81*
All measure types	133	0.62	0.43	0.38	0.48	703.36*	
Reading comprehension	19	0.67	0.28	0.15	0.41	72.35*	
Oral reading of words	20	0.69	0.54	0.40	0.67	110.66*	
Decoding	29	0.56	0.41	0.31	0.51	114.53*	
Oral reading of passages	18	0.48	0.49	0.37	0.61	90.04*	
Composite reading	19	0.72	0.48	0.37	0.59	148.17*	
Spelling	11	0.10	0.14	−0.02	0.29	29.87*	
Writing	4	0.44	0.45	0.19	0.71	0.87	
Listening comprehension	1	0.68	0.68	0.04	1.32	0.00	
Writing Vocabulary (Clay, 1985)	10	1.55	0.94	0.73	1.14	77.24*	
Standardized versus nonstandardized measures							
All interventions	42	0.81	0.58	0.49	0.66	183.10*	1.09
Standardized	21	0.74	0.53	0.41	0.65	75.35*	
Nonstandardized	21	0.89	0.62	0.50	0.74	106.66*	
Reading Recovery	22	0.96	0.70	0.58	0.82	107.68*	2.81
Standardized	11	0.83	0.60	0.43	0.77	46.03*	
Nonstandardized	11	1.09	0.80	0.63	0.97	58.84*	
Other interventions	20	0.65	0.44	0.32	0.56	66.68*	0.06
Standardized	10	0.64	0.46	0.28	0.63	27.98*	
Nonstandardized	10	0.67	0.42	0.25	0.60	38.64*	
Intervention duration							7.90*
Duration reported	30	0.66	0.48	0.38	0.58	177.91*	
Up to 20 weeks	14	0.77	0.65	0.50	0.80	123.59*	
Over 20 weeks	16	0.57	0.37	0.24	0.49	46.42*	
Total instructional time							0.35
Studies reporting total time	27	0.67	0.35	0.24	0.46	130.40*	
Less than 50 hr	15	1.01	0.38	0.24	0.52	102.57*	
50–150 hr	12	0.25	0.32	0.15	0.48	27.48*	
Fidelity of treatment check							0.42
Reported	17	0.55	0.37	0.23	0.51	80.02*	
Not reported	25	0.75	0.43	0.32	0.53	168.36*	
Method of assignment to groups							20.24*
More stringent	25	0.84	0.56	0.45	0.66	140.14*	
Less stringent	17	0.41	0.17	0.04	0.30	88.42*	

Table 3 (*continued*)

Aggregation	*k*	*ES*	*d*	95% CI for *d* Lower	95% CI for *d* Upper	Q_w	Q_a
Reading Recovery student status							13.79*
All	16	0.63	0.66	0.52	0.80	121.58*	
Discontinued	5	1.11	0.71	0.40	1.03	38.29*	
Not discontinued	4	−0.04	0.00	−0.40	0.39	14.07*	
Combination	2	0.61	0.62	0.32	0.92	0.06	
Not reported	5	0.69	0.83	0.63	1.03	55.37*	
Interventions using teachers							3.59
All	28	0.67	0.36	0.26	0.46	136.91*	
Reading Recovery	15	0.55	0.47	0.32	0.62	76.50*	
Other	13	0.81	0.28	0.14	0.41	56.82*	
Interventions for first graders							12.74*
All	26	0.63	0.49	0.39	0.59	173.46*	
Reading Recovery	16	0.63	0.66	0.52	0.80	121.58*	
Other (matched)	10	0.64	0.29	0.14	0.44	39.14*	

Note. For standardized vs. nonstandardized measures section, aggregations include only samples of students assessed using both standardized and nonstandardized measures. *k* = number of independent samples in an aggregation; *ES* = unweighted effect size; *d* = weighted effect size; CI = confidence interval; Q_w = within-groups homogeneity statistic; Q_B = between-groups homogeneity statistic.
* $p < .05$.

Reading Recovery

We conducted several additional analyses to more accurately interpret the effects of Reading Recovery interventions. Because evaluations of Reading Recovery have not been consistent in terms of reporting outcome data for all students who received the intervention, the effect sizes deriving from Reading Recovery studies were disaggregated according to whether the students for whom data were reported were identified as discontinued, not discontinued, a combination of the two, or status not reported. Students who were determined to have completed the Reading Recovery program successfully were identified as discontinued. Those students who did not successfully complete the program were termed not discontinued. As indicated by the significant homogeneity test, $Q_B(3) = 13.79$, students' Reading Recovery status was reliably associated with the magnitude of the effect size estimate. The mean weighted effect size for all samples of Reading Recovery-tutored students was 0.66. Mean weighted effect sizes for the subgroups were as follows: discontinued students ($d = 0.71$, $k = 5$); students who were not discontinued ($d = 0.00$, $k = 4$); discontinued and not discontinued students combined (relative percentages not reported) ($d = 0.62$, $k = 2$); students whose status was not reported ($d = 0.83$, $k = 5$).

It is well known that standard Reading Recovery interventions use trained, certified teachers. This was true for 15 out of 16 samples in the present meta-analysis; in one case, college students were trained to implement the Reading Recovery intervention. Teachers likewise served as the instructors for 13 samples of students receiving other interventions. When only interventions delivered by teachers were considered, the mean weighted effect size for Reading Recovery interventions ($d = 0.47$) was higher than that for other interventions ($d = 0.28$), but the difference was not statistically reliable, $Q_B(1) = 3.59$ (confidence intervals for the respective aggregations were 0.32–0.62 and 0.14–0.41).

Using the available data, we examined how the effects of Reading Recovery compared with those of other one-to-one interventions for first graders. The meta-analysis included 16 samples of first graders taught by Reading Recovery and 12 samples of first graders who received other interventions. To more closely match the qualifications of the instructors used in Reading Recovery programs, we excluded 2 samples from the set of other interventions: (a) the sample from a study by Arnold et al. (1977) that did not provide any information concerning the instructors and the sample from the study by Vadasy et al. (1997) whose college-student tutors were identified as having been very inconsistent in delivering the treatment. Instructors for the remaining 10 samples of students that received an intervention other than Reading Recovery were teachers (6 samples), paraprofessionals (1 sample), or trained volunteers (3 samples). The mean weighted effect size for the Reading Recovery interventions ($d = 0.66$) was significantly higher than that for the other matched interventions, ($d = 0.29$), $Q_B(1) = 12.74$. In interpreting this finding, it should be recalled that some Reading Recovery studies did not include posttest data on all students who began the intervention.

Reading Recovery Versus Small-Group Interventions

As stated earlier, two studies that met our criteria for inclusion in the synthesis contrasted a one-to-one with a small-group intervention. In both cases, the one-to-one intervention was Reading Recovery. The mean within-study effect sizes for these studies were −0.12 (Acalin, 1995) and 0.05 (Evans, 1996), indicating that the small-group interventions (one of which was based on Reading Recovery) achieved outcomes comparable to those of standard, one-to-one Reading Recovery while serving 3–4 times the number of students per instructor.

Discussion

The purpose of this meta-analysis was to assess the effectiveness of supplemental, adult-delivered, one-to-one interventions for elementary school children with low reading skills. On average,

students who received one-to-one instruction performed at a level 2/5 of a standard deviation higher than the average level of the comparison group, corresponding to a move from the 50th to the 65th percentile on a standardized measure. For students experiencing severe difficulty in reading, an effect of this magnitude would not likely be sufficient to raise performance to within the average range for a grade level or to eliminate the need for sustained support beyond the time frame of the intervention. The benefit might, however, be great enough to allow these students to keep up with classroom instruction and to avoid academic failure.

The meta-analysis revealed that college students and trained, reliable community volunteers were able to provide significant help to struggling readers. This finding suggests that it may be possible to reduce the cost of providing effective, supplemental, one-to-one instruction to students at risk for reading failure (Hiebert, 1994). Alternatively, the number of children to whom schools provide supplemental, one-to-one tutoring can be greatly increased by having trained tutors work under the supervision of a qualified teacher or reading specialist (Dromsky & Gambrell, 1999; Howard, 1999; Morrow & Woo, 1999).

In considering the positive findings for trained college students and community volunteers, it should be recalled that the instruction provided by these tutors was not a substitute for, but rather an adjunct to, classroom instruction provided by certified teachers. Thus, the findings should not be interpreted to suggest that individuals who have had little or no formal teacher training can equal the instructional performance of highly skilled professionals under circumstances other than those of a well-designed and well-monitored intervention in a circumscribed domain.

Intervention duration was not significantly associated with outcomes, a finding also reported by Swanson (1999) in a recent meta-analysis of reading research for students with learning disabilities. One study that contrasted a standard Reading Recovery program with a modified Reading Recovery program (Iversen & Tunmer, 1993) reported that students in the modified program were discontinued after an average of 41.75 lessons, compared with 57.31 lessons for students in the standard program. The effect size for students in the modified program was comparable to that of students in the standard program, suggesting that it is possible to achieve the same outcomes in a much shorter period of time by modifying the content of instruction. This finding suggests that efficiency, or the amount of progress over time, may be a useful variable to consider in conducting future studies.

Two studies included in the meta-analysis compared a one-to-one intervention with a small-group intervention. When Reading Recovery was compared with Project READ (Acalin, 1995), a program in which students are taught in groups of 2 to 5 students, the overall effect size was −0.12, indicating no advantage for the one-to-one intervention over the small-group intervention. Evans (1996) compared Reading Recovery with a small-group intervention based on Reading Recovery principles; the resulting effect size, 0.05, showed no advantage for the one-to-one intervention. Further evidence of the comparability of well-designed small-group interventions to one-to-one interventions is available from the study by Pinnell, Lyons, DeFord, Bryk, and Seltzer (1994), which compared three different one-to-one interventions (standard Reading Recovery, Reading Recovery in which teachers received condensed training, and direct instruction by certified teachers) with a small-group intervention as well as with a control group. The effect sizes used in the meta-analysis were based on comparisons of the one-to-one interventions with the control groups. However, we also calculated the effect sizes resulting from a comparison of the same one-to-one interventions and the small-group intervention, which was based on Reading Recovery and taught by Reading Recovery-trained teachers. The mean effect size for comparisons of the three one-to-one interventions with the small-group intervention was −0.12, again indicating no advantage for one-to-one over small-group instruction. When only the Reading Recovery interventions were compared with the control group, thus holding constant all the important parameters of instruction (teacher qualifications, content of instruction, procedures, materials, duration of instruction) except group size, the mean weighted effect size was 0.04.

Together, the findings from the available small-group comparisons indicate that when highly qualified teachers implement a well-designed intervention, the academic benefit to students is the same, whether students are taught individually or in a group of 2 to 6 students. With regard to academic benefits, this finding is in line with those of other recent meta-analyses indicating the effectiveness of small-group instruction across all subject areas (Lou et al., 1996) as well as in the specific area of reading (Elbaum, Vaughn, Hughes, & Moody, 1999; Swanson, 1999).

The meta-analysis reveals that students who participate in Reading Recovery perform at a level that is 2/3 of a standard deviation higher than the average level of comparable students who do not participate in the program. However, the meta-analysis also underscores a number of concerns regarding the interpretation of reported outcomes for Reading Recovery-tutored students. We highlight three of these concerns: the disparity in outcomes for discontinued and not discontinued students, the uneven representation of discontinued and not discontinued students at posttest, and the possible overestimation of intervention effects owing to the low reliability of typically used outcome measures.

One-to-one instruction through Reading Recovery yields different effects for different groups of students. Whereas the effects for discontinued students were substantial (d = 0.71), the effects for students who were not discontinued, that is, who did not complete the program successfully, were nil (d = 0.00). Additionally, the primary studies did not consistently report data for both discontinued and not discontinued students, and studies that reported results for discontinued students only did not always report how many students were dropped from the intervention for failure to make adequate progress. Sometimes this number was substantial. For example, in one study that reported this information (Saginaw Public Schools, 1992), 20 of 55 students who began the Reading Recovery intervention were dropped. In the study by Ramaswami (1994), 18 of 30 students were dropped. This practice represents a particularly pernicious form of participant attrition in which the researchers selectively remove participants from a study based precisely on the participants' failure to respond adequately to the treatment. Comparisons of selectively reduced treatment groups with control groups that remain intact are consequently extremely suspect.

Concerns have been raised regarding the psychometric properties of some of the measures used to assess Reading Recovery outcomes (Chapman, Tunmer, & Prochnow, 1999). For example, Askew, Fountas, Lyons, Pinnell, and Schmitt (1998) stated that "the text reading measure is not an equal interval scale; that is,

there are smaller differences in the beginning levels than at upper levels. For beginning readers, it is necessary to look at the reader's progress in more detail" (p. 10). Indeed, in our meta-analysis, this measure yielded a higher effect size estimate than any other measure. The Text Reading Level measure (Clay, 1985) is additionally suspect in that Chapman, Tunmer, and Prochnow (1998) reported a dramatic discrepancy between Text Reading Level outcomes measured by Reading Recovery teachers and outcomes for the same students measured by classroom teachers who had not participated in the intervention (Reading Recovery teacher assessment $M = 16.6$, $SD = 2.6$; classroom teacher assessment $M = 9.0$, $SD = 4.7$), $t(21) = 7.15$, $p < .001$.

In the present meta-analysis, intervention effects based on standardized reading tests were compared with those based on researcher-developed measures. For Reading Recovery samples, the effect size estimate based on nonstandardized measures was larger than that based on standardized measures ($d = 0.80$ vs. 0.60), although the difference was not statistically reliable. The Writing Vocabulary measure (Clay, 1985) yielded an effect size estimate ($d = 0.94$) whose confidence interval did not overlap with that of any other measure type. Thus, effect size estimates based on nonstandardized measures, particularly the Text Reading Level measure (Clay, 1985), may overestimate intervention effects.

Overall, the findings of this meta-analysis do not provide support for the superiority of Reading Recovery over other one-to-one reading interventions. Typically, about 30% of students who begin Reading Recovery do not complete the program and do not perform significantly better than control students (Pinnell, 1988). As indicated in this meta-analysis, results reported for students who do complete the program may be inflated due to the selective attrition of students from some treatment groups and the use of measures that may bias results in favor of Reading Recovery students. Thus, it is particularly disturbing that sweeping endorsements of Reading Recovery still appear in the literature. For example, in a recent issue of *Educational Leadership*, the official publication of the Association for Supervision and Curriculum Development, Daniels, Zemelman, and Bizar (1999) stated that "Reading Recovery has been phenomenally successful" (p. 35).

The findings of the present meta-analysis should be considered in the light of several limitations. A number of studies of one-to-one reading interventions could not be included in the meta-analysis because they did not fulfill all the criteria for inclusion. Among these were various Reading Recovery evaluations that compared at-risk students who participated in Reading Recovery with average-achieving students who did not. Other studies of one-to-one interventions, such as that of Vellutino et al. (1996), could not be included because effect sizes could not be computed from the information presented.

With some notable exceptions (Juel, 1996; Torgesen et al., 1998; Vadasy et al., 1997), the studies included in the meta-analysis did not provide a sufficiently detailed description of the intervention for us to reliably code features other than the general focus of the instruction. When examined in these admittedly general terms, interventions focusing on visual–perceptual skills had no measurable effect; interventions focusing on phonemic awareness–phonics or having a mixed (some would say *balanced*) focus had moderate effects; and interventions that focused on reading comprehension (for older elementary students only) had large effects. For a detailed investigation of the components of instruction that are associated with reading outcomes for students with learning disabilities, we refer readers to the recent meta-analysis by Swanson (1999).

A final limitation of the meta-analysis concerns the analytic approach imposed by the data. Given the uneven distribution of studies across combinations of the important moderator variables, it was not possible to determine the effects of each variable (e.g., focus of instruction) across all categories of the other variables (e.g., type of instructor).

In sum, the findings of this meta-analysis support the argument that well-designed, reliably implemented, one-to-one interventions can make a significant contribution to improved reading outcomes for many students whose poor reading skills place them at risk for academic failure. Based on these findings, we recommend that schools give serious consideration to one-to-one reading interventions that use trained college students and volunteers and to intensive small-group interventions.

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Acalin, T. A. (1995). *A comparison of Reading Recovery to Project READ.* Unpublished doctoral dissertation, California State University, Fullerton.

Adler, J. (1998, March 30). The tutor age. *Newsweek,* pp. 46–56.

*Arnold, L. E., Barnebey, N., McManus, J., Smeltzer, D. J., Conrad, A., Winer, G., & Desgranges, L. (1977). Prevention by specific perceptual remediation for vulnerable first-graders. *Archives of General Psychiatry, 34,* 1279–1294.

Askew, B. J., Fountas, I. C., Lyons, C. A., Pinnell, G. S., & Schmitt, M. C. (1998). *Reading Recovery review: Understandings, outcomes, & implications.* Columbus, OH: Reading Recovery Council of North America.

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

*Butler, S. R. (1991). Reading program—Remedial, integrated and innovative. *Annals of Dyslexia, 41,* 119–127.

Carnine, D., & Silbert, J. (1979). *Direct instruction reading.* Columbus, OH: Charles E. Merrill.

Center, Y., Wheldall, K., & Freeman, L. (1992). Evaluating the effectiveness of Reading Recovery: A critique. *Educational Psychology, 12,* 263–274.

*Center, Y., Wheldall, K., Freeman, L., Outhred, L., & McNaught, M. (1995). An evaluation of Reading Recovery. *Reading Research Quarterly, 30,* 240–263.

*Chapman, J. W., Tunmer, W. E., & Prochnow, J. E. (1998, April). *Reading recovery in relation to language factors, reading self-perceptions, classroom behavior difficulties and literacy achievement: A longitudinal study.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Chapman, J. W., Tunmer, W. E., & Prochnow, J. E. (1999). *Success in Reading Recovery depends on the development of phonological processing skills* (revised research report for Phase Three of Contract ER35/299/5). Palmerston North, New Zealand: Massey University.

Clay, M. M. (1985). *The early detection of reading difficulties.* Portsmouth, New Hampshire: Heinemann.

Cohen, P., Kulik, J. A., & Kulik, C. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal, 19,* 237–248.

Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American Educator, 15*(3), 6–11, 38–46.

*Compton, G. L. (1992). *The reading connection: A leadership initiative designed to change the delivery of educational services to at-risk children.* Unpublished doctoral dissertation, Western Michigan University, Kalamazoo.

Cooper, H. M. (1989). *Integrating research: A guide for literature reviews* (2nd ed.). Newbury Park, CA: Sage.

Cooper, H. M. (1999). *Synthesizing research: A guide for literature reviews* (3rd ed.). Newbury Park, CA: Sage.

Cooper, H., Charlton, K., Valentine, J. C., & Muhlenbruck, L. (2000). *Making the most of summer school: A meta-analytic and narrative review.* Malden, MA: Blackwell.

Cooper, H., & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis.* New York: Russell Sage Foundation.

Daniels, H., Zemelman, S., & Bizar, M. (1999). Whole language works: Sixty years of research. *Educational Leadership, 57,* 32–37.

*Dorval, B., Wallach, L., & Wallach, M. A. (1978). Field evaluation of a tutorial reading program emphasizing phoneme identification skills. *The Reading Teacher, 31,* 784–790.

Dromsky, A. J., & Gambrell, L. B. (1999, April). *America Reads at the University of Maryland: Practices, evaluation and implications.* Panel presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada.

Educational Resources Information Center. (1995). Thesaurus of ERIC descriptors. Washington, DC: Author.

Elbaum, B., Vaughn, S., Hughes, M. T., & Moody, S. W. (1999). Grouping practices and reading outcomes for students with disabilities. *Exceptional Children, 65,* 399–415.

*Evans, T. L. P. (1996). *I can read deze books: A qualitative comparison of the Reading Recovery program and a small-group reading intervention.* Unpublished doctoral dissertation, Auburn University, Auburn, Alabama.

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

*Graves, A. W. (1986). Effects of direct instruction and metacomprehension training on finding main ideas. *Learning Disabilities Research 2,* 90–100.

Grossen, B., Coulter, G., & Ruggles, B. (1997). *Reading Recovery: An evaluation of benefits and costs* [On-line]. Available: http://darkwing.uoregon.edu/~bgrossen/rr.htm

*Hagin, R. A., Silver, A. A., & Beecher, R. (1978). TEACH: Learning tasks for the prevention of learning disabilities. *Journal of Learning Disabilities, 11,* 54–57.

Hedges, L., & Olkin, I. (1985). Statistical methods for meta-analysis. Orlando, FL: Academic Press.

*Hedrick, D. E. (1996). *An administrative review of an early reading intervention.* Unpublished doctoral dissertation, University of North Carolina at Greensboro.

Hiebert, E. H. (1994). Reading Recovery in the United States: What difference does it make to an age cohort? *Educational Researcher, 23*(9), 15–25.

Howard, J. K. (1999, April). *Project VALUE: Valid, Adult-Mediated Literacy Instruction which is Useful and Efficient.* Panel presented at the annual meeting of the Educational Research Association, Montreal, Quebec, Canada.

Invernizzi, M., Juel, C., & Rosemary, C. A. (1997). A community volunteer tutorial that works. *The Reading Teacher, 50,* 304–311.

*Iversen, S., & Tunmer, W. E. (1993). Phonological processing skills and the Reading Recovery program. *Journal of Educational Psychology, 85,* 112–126.

Jenkins, J. R., Mayhall, W. F., Peschka, C. M., & Jenkins, J. M. (1974). Comparing small group and tutorial instruction in resource rooms. *Exceptional Children, 40,* 245–251.

Juel, C. (1991). Cross-age tutoring between student-athletes and at-risk children. *The Reading Teacher, 45,* 178–186.

*Juel, C. (1996). What makes literacy tutoring effective? *Reading Research Quarterly, 31,* 268–289.

Karweit, N. L., & Wasik, B. A. (1992). *A review of the effects of extra-year kindergarten programs and transitional first grades* (Report No. CDS-R-41). Washington, DC: Office of Educational Research and Improvement. (ERIC Document Reproduction Service No. ED 357 894)

Kennedy, M. M., Birman, B. F., & Demaline, R. E. (1986). *The effectiveness of Chapter I services.* Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

*Knapp, N. F., & Winsor, A. P. (1998). A reading apprenticeship for delayed primary readers. *Reading Research and Instruction, 38,* 13–29.

*Lafave, C. E. (1995). *Impact of Reading Recovery on phonemic awareness.* Unpublished doctoral dissertation, University of Toledo, Toledo, Spain.

Loper, A. (1980). Metacognitive development: Implications for cognitive training. *Exceptional Education Quarterly, 1,* 1–8.

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research, 66,* 423–458.

*Mantzicopoulos, P., Morrison, D., Stone, E., & Setrakian, W. (1992). Use of the search/teach tutoring approach with middle-class students at risk for reading failure. *The Elementary School Journal, 92,* 573–586.

Mathes, P. G., & Fuchs, L. S. (1994). The efficacy of peer tutoring in reading for students with mild disabilities: A best-evidence synthesis. *School Psychology Review, 23*(1), 59–80.

*McCarthy, P., Newby, R. F., & Recht, D. R. (1995). An early intervention program for first grade children at-risk for reading disability. *Reading research and Instruction, 34,* 273–294.

*McGrady, D. S. (1984). *The effect of programmed tutoring upon the reading comprehension of fourth-grade students enrolled in a Chapter 1 reading program.* Paper presented at the annual meeting of the Association for Educational Communications and Technology, Dallas, TX. (ERIC Document Reproduction Service No. ED 243 430)

McIntosh, R., Vaughn, S., Schumm, J., Haager, D., & Lee, O. (1993). Observations of students with learning disabilities in general education classrooms. *Exceptional Children, 60,* 249–261.

Moody, S. W., Vaughn, S., & Schumm, J. S. (1997). Instructional grouping for reading: Teachers' views. *Remedial and Special Education, 18,* 347–356.

*Morris, D., Shaw, B., & Perney, J. (1990). Helping low readers in Grades 2 and 3: An after-school volunteer tutoring program. *The Elementary School Journal, 91,* 133–150.

Morrow, L. M., & Woo, D. G. (1999, April). *The effects of an American Reads tutoring program on literacy achievement and an evaluation of the attitudes of teachers, tutors, and children toward the program.* Panel presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada.

*Nielson, B. B. (1991). *Effects of parent and volunteer tutoring on reading achievement of third grade at-risk students.* Unpublished doctoral dissertation, Brigham Young University, Provo, UT.

*Pinnell, G. S. (1988). *Success of children at risk in a program that combines writing and reading.* Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement. (ERIC Document Reproduction Service No. ED 292 061)

*Pinnell, G. S., Lyons, C. A., DeFord, D. E., Bryk, A. S., & Seltzer, M. (1994). Comparing instructional models for the literacy education of high-risk first graders. *Reading Research Quarterly, 29,* 9–39.

*Ramaswami, S. (1994). *The differential impact of Reading Recovery on achievement of first graders in the Newark School District.* Newark, NJ: Newark Board of Education, Office of Planning, Evaluation and Testing. (ERIC Document Reproduction Service No. ED 374 180)

*Ramey, M. (1991). *Compensatory education sustained gains from spring 1988 to spring 1990.* Seattle, WA: Seattle Public Schools. (ERIC Document Reproduction Service No. ED 338 774)

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231–244). New York: Russell Sage Foundation.

*Saginaw Public Schools, Department of Evaluation Services. (1992). *Compensatory education product evaluation: Reading Recovery Program 1991–92.* Saginaw: Author. (ERIC Document Reproduction Service No. ED 350 587)

Shanahan, T. (1998). On the effectiveness and limitations of tutoring in reading. In P. D. Pearson & A. Iran-Nejad (Eds.), *Review of research in education* (Vol. 23, pp. 217–234). Washington, DC: American Educational Research Association.

Shanahan, T., & Barr, R. (1995). Reading Recovery: An independent evaluation of the effects of an early instructional intervention for at-risk learners. *Reading Research Quarterly, 30,* 958–995.

Silver, A. A., & Hagin, R. A. (1976). *SEARCH and TEACH.* New York: Walker Educational Publishing.

Slavin, R. E., & Madden, N. A. (1989). What works for students at risk: A research synthesis. *Educational Leadership, 46*(5), 4–13.

SPSS, Version 10.0 [Computer software]. (1999). Chicago: SPSS, Inc.

Swanson, H. L. (1999). Reading research for students with LD: A meta-analysis of intervention outcomes. *Journal of Learning Disabilities, 32,* 504–532.

*Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Rose, E., Lindamood, P., Conway, T., & Garvan, C. (1998). *Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction.* Unpublished manuscript.

Towner, J., & Davidson, M. (1998, April). *The quality of evidence in monitoring academic progress in reading: Comparisons of Reading Recovery, Special Education, Title 1, and general education students in Grades 1 through 3 using curriculum-based measures.* Panel presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Tukey, J. W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

*Vadasy, P. F., Jenkins, J. R., Antil, L. R., Wayne, S. K., & O'Connor, R. E. (1997). The effectiveness of one-to-one tutoring by community tutors for at-risk beginning readers. *Learning Disability Quarterly, 20,* 126–139.

*Vadasy, P. F., Jenkins, J. R., & Pool, K. (1998). *Effects of a first-grade tutoring program in phonological and early reading skills.* Unpublished manuscript.

Vaughn, S., Moody, S., & Schumm, J. S. (1998). Broken promises: Reading instruction in the resource room. *Exceptional Children, 64,* 211–226.

Vellutino, F., Scanlon, D. M., Sipay, E., Small, S., Pratt, A., Chen, R., & Denckla, M. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88,* 601–638.

*Wallach, M. A., & Wallach, L. (1976). *Teaching all children to read.* Chicago: University of Chicago Press.

Wasik, B. A., & Slavin, R. E. (1993). Preventing early reading failure with one-to-one tutoring: A review of five programs. *Reading Research Quarterly, 28,* 178–200.

*Weeks, D. (1992). *A study of the implementation of Reading Recovery in Scarborough: 1990–1991.* Unpublished doctoral dissertation, University of Toronto, Toronto, Ontario, Canada.

White, H. D. (1994). Scientific communication and literature retrieval. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 41–56). New York: Russell Sage Foundation.

Received November 29, 1999
Revision received April 29, 2000
Accepted April 29, 2000 ■