SIX READING RECOVERY STUDIES:

# Meeting the Criteria for Scientifically Based Research

by North American Trainers Group Research Committee

*An Evaluation of Reading Recovery*
CENTER, WHELDALL, FREEMAN, OUTHRED, AND McNAUGHT

*Phonological Processing Skills and the Reading Recovery Program*
IVERSEN AND TUNMER

*Reading Recovery: Helping At-Risk Children Learn to Read*
PINNELL

*Comparing Instructional Models for the Literacy Education of High Risk First Graders*
PINNELL, LYONS, DeFORD, BRYK, AND SELTZER

*Children's Achievement and Personal and Social Development in a First-Year Reading Recovery Program with Teachers-In-Training*
QUAY, STEELE, JOHNSON, AND HORTMAN

*Literacy Learning of At-Risk First-Grade Students in the Reading Recovery Early Invervention*
SCHWARTZ

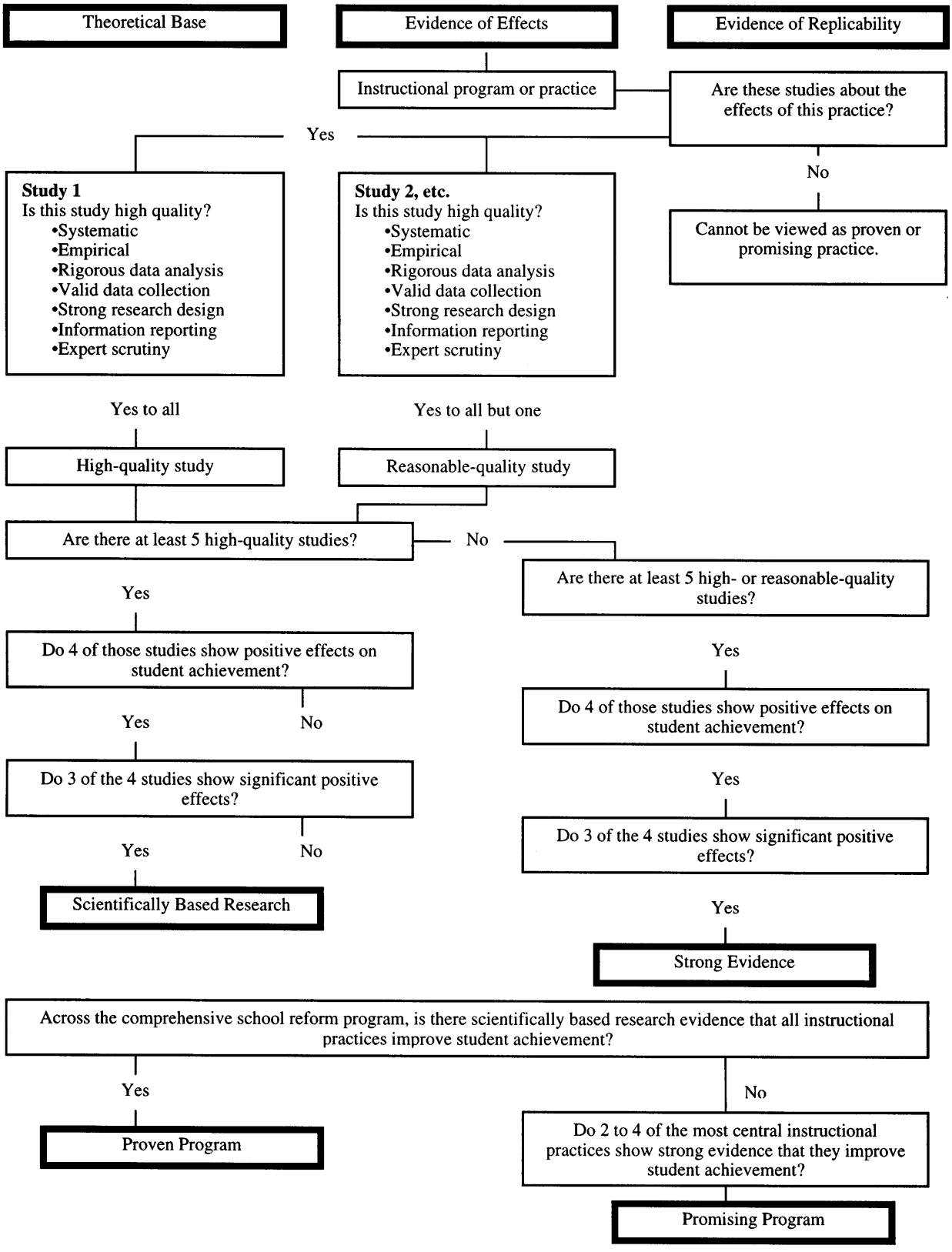
This report presents six studies that meet criteria for scientifically based reading research as defined by the United States Department of Education. Although research and evaluation on Reading Recovery has been conducted since its beginnings in the late 1970s, the research base has taken on new urgency in the U.S. since federal funding under Reading First has been tied to programs with proven effectiveness.

In 2002, the United States Department of Education published a Quality of Research Decision Tree as a guide to help states and local schools evaluate proposed reading programs. The decision tree diagram required decision-makers to examine evidence based on (1) the theoretical base, (2) evidence of effects, and (3) evidence of replicability.

**The Theoretical Base**
For Reading Recovery, the theoretical base is primarily represented in the writing of Marie Clay (1991, 2001, in press). Another relevant review of Reading Recovery theory and practice is *Changing Futures: The Influence of Reading Recovery in the United States* (Schmitt, Askew, Fountas, Lyons, & Pinnell, 2005).

**Evidence of Effects**
The major emphasis of the Quality of Research Decision Tree research is devoted to determining the effectiveness of programs for student achievement. The six studies presented here, all of which are published in peer-reviewed journals, demonstrate strong effects of the Reading Recovery intervention in carefully controlled experimental studies. Four of these research studies were conducted by researchers associated with Reading Recovery, and two by researchers who have been critical of Reading Recovery. All six studies support the evaluation data collected and reported annually by NDEC. This extensive database reflects the valuable work done by Reading Recovery teachers, teacher leaders, university trainers, site coordinators, and district administrators to maintain a high quality network of professionals who work together to support the literacy learning of our most at-risk children.

**Evidence of Replicability**
The evidence of replicability is provided in the reports of the Reading Recovery National Data Evaluation Center (NDEC) that document 20 years of work with more than 1.5 million at-risk first-grade children. National results for Reading Recovery students are published as a technical report on the NDEC Web site at http://www.ndec.us.




*Note:* The Quality of Research Decision Tree appears in a 2002 United States Department of Education publication, *Scientifically Based Research and the Comprehensive School Reform (CSR) Program,* Appendix 3, pp. 17–18.

## An Evaluation of Reading Recovery

**Center, Y., Wheldall, K., Freeman, L., Outhred, L., & McNaught, M. (1995). *Reading Research Quarterly, 30*(2), 240–263.**

**1. Systematic and Empirical**

This is a high-quality experimental design using random assignment of subjects to Reading Recovery or a no-intervention control group in 10 schools. A comparison group of low-progress students from five matched schools that had not implemented Reading Recovery was also followed. All groups were assessed on a variety of reading-related measures just prior to the intervention (pre-test), after 15 weeks (post-test), after another 15 weeks (short-term maintenance), and again 12 months after the post-test (medium-term maintenance).

**2. Rigorous Data Analysis**

A series of multivariate and univariate analysis of variance procedures were used to compare Reading Recovery students and control group students at pre-test, post-test, short-term maintenance, and medium-term testing. Control students were compared with comparison students at pre-test, post-test, and short-term maintenance.

**3. Valid Data Collection**

Across-testing-period results were reported for Clay's book level test, Burt Word Reading Test, Neale Analysis of Reading Ability, Passage Reading Test, Waddington Diagnostic Spelling Test, Phonemic Awareness Test, Cloze Test, and Word Attack Skills Test. References and reliabilities for these measures were included in the article. Testing at post-test and each maintenance period was conducted by a research assistant not involved with Reading Recovery. Similarly the researchers who conducted this study had no connection or commitment to the program. They conducted an independent, critical evaluation.

**4. Strong Research Design**

As mentioned above, this was an experimental design with random assignment.

**5. Information Reporting and Expert Scrutiny**

The article was published in *Reading Research Quarterly,* the research journal of the International Reading Association. This is considered one of the top-quality journals in the field of reading education and research with a strong editorial process and low acceptance rate for manuscripts submitted.

**6. Significant Positive Effects**

At post-test, after 15 weeks of intervention, the "Reading Recovery students significantly outperformed control students on all tests measuring words read in context and in isolation, but not on some tests of metalinguistic skills" (p. 252). At the end of first grade (short-term maintenance), the Reading Recovery students continued to score higher than the control group on all measures. The authors conclude "that the Reading Recovery group continued to perform significantly better than control students on all tests measuring word reading in context and on a phonemic awareness measure. However, on tests measuring phonological recoding and syntactic awareness, not specifically addressed by the program, the differences just failed to reach significance" (p. 252).

At medium-term maintenance, a year after the intervention period, the Reading Recovery group continued to score higher than both the control group and the comparison group on all measures (see Table 7, p. 254). The authors reported a MANOVA value of $p = .0268$ for the comparison of the Reading Recovery and control group at this point in time (p. 253). This was significant according to the alpha = .05 criteria established in the beginning of their analysis section (p. 250). The reduced size of effects in this final comparison may be due to the fact that 15 of the original 31 control students had in fact entered Reading Recovery for intervention support prior to this testing and therefore the control group was reduced to 16 by the removal of students making the least progress.

In summary, this is a high-quality, independent evaluation of Reading Recovery showing highly significant and long-lasting effects of the intervention.

## Phonological Processing Skills and the Reading Recovery Program

**Iversen, S., & Tunmer, W. E. (1993). *Journal of Educational Psychology, 85*(1), 112–126.**

**1. Systematic and Empirical**

This is an experimental study looking at the progress of three matched groups of at-risk first-grade students across the year. Groups of 32 students were assigned using quasi-random procedures to treatments labeled standard Reading Recovery, modified Reading Recovery, or standard intervention (small-group Title I). Students completed a battery of tests at the beginning and end of the school year and a midyear point corresponding to the discontinuing point for the Reading Recovery subjects. In addition, average classroom students were tested from the same classrooms as the Reading Recovery students at the discontinuing point.

**2. Rigorous Data Analysis**

Analysis of variance procedures were used to compare each of the treatment and comparison groups at each of the testing periods. Additional questions were investigated using correlation and path analysis procedures.

**3. Valid Data Collection**

Test measures included the six tasks from Clay's Diagnostic Survey, the Dolch Word Recognition Test, and measures of phoneme segmentation, phoneme deletion, and phonological recoding. Testing was done by trained teachers involved in the instruction of each child, but care was taken to ensure that teachers were not aware of which of the children they taught were the focus of the study. The first author had been previously trained as a Reading Recovery teacher leader, but was not involved in ongoing professional development with that program at the time the study was conducted. The second author was a university researcher who had engaged in independent, critical evaluation of Reading Recovery.

**4. Strong Research Design**

As described above, this was an experimental study using matching and random assignment to the extent possible for research in school settings. A total of 23 schools and 26 Reading Recovery teachers during their first year of training worked with 64 children from 34 classrooms for the two Reading Recovery groups. The standard small group intervention involved 32 children from seven schools taught by seven reading specialists. The Reading Recovery teachers participated in the yearlong Reading Recovery training in two separate groups with the modified Reading Recovery group of teachers trained to include a component in the Reading Recovery lesson "to make children more aware that words with common sounds often share spelling patterns" (p. 117).

**5. Information Reporting and Expert Scrutiny**

The article was published in a high-quality peer-reviewed publication, *The Journal of Educational Psychology,* a research publication of the American Psychological Association.

**6. Significant Positive Effects**

The three treatment groups were essentially equal and low on all measures at pre-test. At discontinuation of the Reading Recovery intervention, both the standard and modified Reading Recovery groups scored significantly higher on all outcome measures than the standard intervention group. For measures like Text Reading Level, with no ceiling effect, these differences were extremely large (over eight standard deviations on this measure and over two standard deviations for the Dolch Word Recognition Test). In addition, both Reading Recovery groups had test profiles very similar to average students from their classroom settings.

The two Reading Recovery groups looked similar on most measures, including the phonemic measures included to assess the effectiveness of the modified Reading Recovery approach. In fact, the standard Reading Recovery group scored significantly higher than the modified Reading Recovery group on the phoneme deletion measure. A major advantage for the modified Reading Recovery group was that students reached the discontinuing point in fewer lessons (41.75 versus 57.31). A modification in the Reading Recovery lessons structure similar to that described in this study has been incorporated in the standard program, not as a result of this study, but as a concurrent adjustment based on literacy research (see Clay, 1993, pp. 43–51).

At the end of the year both the standard and modified Reading Recovery groups appeared very similar with a sight advantage for the modified group on the text reading measure (19.56 vs. 18.38).

In summary, this was a high-quality study of two variations on Reading Recovery procedures versus small group Title I instruction by certified teachers with master's degrees in reading. The results clearly showed a very large advantage for students in the one-to-one tutoring setting using Reading Recovery procedures. The results also showed that both sets of Reading Recovery procedures supported the learning of phonemic awareness knowledge and the application of that knowledge to text reading and writing.

## Reading Recovery: Helping At-Risk Children Learn to Read

**Pinnell, G. S. (1989). *The Elementary School Journal, 90,* 161–181.**

**1. Systematic and Empirical**

This study utilized a high-quality experimental design involving 21 teachers, all of whom were in their training year for Reading Recovery. Children were from six urban schools with high proportions of low-income students. Reading Recovery children (n = 55) were the lowest students in the program classrooms (taught by a Reading Recovery-trained teacher), and comparison children (n = 55) were the lowest children in comparison classrooms (taught by a teacher not trained in Reading Recovery). Measures were Text Reading Level

and other measures (Observation Survey and the Stanford Achievement Test).

**2. Rigorous Data Analysis**

Means and standard deviations were calculated for all measures. Multivariate analysis (Hotelling's T2) was used to determine significance in comparisons.

**3. Valid Data Collection**

Data were collected at four points: October, mid-year, end of year, and end of the year following treatment. "Blind" testers administered measures individually, except for the standardized test, which was administered to children in small groups.

**4. Strong Research Design**

This is an experimental study with valid comparison groups.

**5. Information Reporting and Expert Scrutiny**

This article was published in a refereed journal, *The Elementary School Journal.* It received the International Reading Association's Albert J. Harris Award for research on reading difficulties.

**6. Significant Positive Effects**

Means and standard deviations were calculated on all measures for four groups: (1) Reading Recovery children in program classrooms, (2) Reading Recovery children in regular classrooms, (3) comparison children, and (4) random sample children (May scores only). Multivariate analysis (Hotelling's T2) indicated significant differences between Reading Recovery children from regular classrooms and comparison children from regular classrooms. The univariate t tests revealed that Reading Recovery children from regular classrooms performed better ($p < .05$) than comparison children on seven of the nine dependent measures. (Ceiling effects were observed for Letter Identification and the Word Test.) Reading Recovery children from program classrooms performed better ($p < .05$) than comparison children on all measures. The two groups of Reading Recovery children (program and regular classrooms) achieved similar results on all measures (whether taught in a classroom by a Reading Recovery teacher or not). A year later, results of the follow-up study of children indicated that Reading Recovery children scored significantly higher ($p < .05$) on all measures than did comparison children.

In summary, this is a high-quality study. It was conducted by a university team who was engaged in implementing and testing Reading Recovery during its first year of operation in the United States. Results were independently audited by a team of researchers led by Dr. Richard Anderson of the Center for the Study of Reading, University of Illinois, and reported to the Ohio Department of Education.

## Comparing Instructional Models for the Literacy Education of High Risk First Graders

**Pinnell, G. S., Lyons, C. A., DeFord, D. E., Bryk, A. S., & Seltzer, M. (1993). *Reading Research Quarterly, 29,* 8–39.**

**1. Systematic and Empirical**

This study utilized a high-quality experimental design using random assignment of subjects to four treatment groups, each of which had its own control group. The four treatment groups were (1) Reading Recovery (individual tutoring), (2) a Reading Recovery-like intervention (individual tutoring by a teacher trained in an alternative and short setting), (3) a Reading Recovery-like small group intervention, and (4) a basic skills small group intervention. Measures used were the Gates-MacGinitie Reading Test, the Woodcock Reading Mastery Test, and Text Reading Level and Hearing and Recording Sounds in Words (Clay's Observation Survey). The design employed for the study was a split plots design replicated over a series of blocks (in this case, districts). One school in each of the selected districts already had Reading Recovery. This school was designated as the Reading Recovery treatment site for the district. Three additional schools were identified and randomly assigned to one of the three alternative treatments. Within each school, a pool of the 10 lowest-scoring students was identified. Four students were randomly assigned to the treatment at the school, and the remaining students constituted a randomized comparison group. Each school was in essence a small randomized trial for one treatment. A total of 403 students (238 male and 165 female) representing two rural, two suburban, and six urban school districts participated in the study. Data were collected at the beginning of the year, mid-year, end of year, and beginning of the following year.

**2. Rigorous Data Analysis**

The Hierarchical Linear Model was employed to analyze the data. It consisted of a student-level and a school-level model. In the student-level model, an indicator variable captured the assignment of students within each school to either the comparison group or the specific treatment present in that school. Two student-level covariates were included as predictors for each outcome variable analysis. This model allowed researchers to estimate a separate treatment effect for

each school in which the treatment was administered. Each of these treatment effect estimates was adjusted for any observed differences on the two covariates between the treatment and comparison groups in each school. In the follow-up studies, using the May and October (of Year 2) outcome data, a third student-level predictor was included to indicate post-experiment exposure to Reading Recovery. The coefficient associated with this variable estimates the effect of Reading Recovery on these students. Thus, there were two estimates of the long-term effects of Reading Recovery: one from the formal experimental group as designed in the study, and a second from this auxiliary quasi-experimental group.

**3. Valid Data Collection**

In this large field trial, data were collected simultaneously in 40 different school sites. Four of the forty schools were found not to have correctly identified the lowest students; while these data did not change the results, these schools were withdrawn from the study. Three others were also dropped because of invalid test administration. The overall experimental sample consisted of 324 students in 33 schools. Data were collected by research associates who were "blind" as to the treatment received by children.

**4. Strong Research Design**

This field study involved large numbers of children, random assignment to treatment and control groups, and one full year of time.

**5. Information Reporting and Expert Scrutiny**

This research was supported by a grant from the John D. and Catherine T. MacArthur Foundation. A national advisory board supervised the research and took an active role in every phase of the research project. This board included Isabel Beck, University of Pittsburgh; Gerald Bracey, Washington, D.C.; Shirley Brice-Heath, Stanford University; Robert Slavin, Johns Hopkins University; Dorothy Strickland, Rutgers University; and Richard Venezky, University of Delaware, who served as chair. Rebecca Barr, National Louis University, and Jana Mason, University of Illinois, acted as consultants. Jeanne Chall, Harvard Graduate School of Education, acted as advisor for the research and reviewed all data and publications of the research. The data were independently analyzed by Anthony Bryk and Michael Seltzer, University of Chicago.

**6. Significant Positive Effects**

Reading Recovery (individual tutoring with trained teachers) was the only group for which the mean treatment effect was significant on all four measures (Hearing and Recording Sounds in Words, 2nd form; Text Reading Level; Gates-MacGinitie; and Woodcock) at the conclusion of the field experiment. The sustained effects of this treatment were measured the following fall, with significant mean treatment effects on Text Reading and small effects on Hearing and Recording Sounds in Words, a possible result of a ceiling effect. The quasi-experimental group that emerged from district decisions to move the lowest-achieving children from the comparison groups into Reading Recovery (i.e. Post Study Reading Recovery) were also significantly different on Hearing and Recording Sounds in Words (Fall of Year 2) and in terms of the end-of-year gain on the Gates-MacGinitie. Reading Recovery emerged as the most powerful of the interventions tested from the beginning of Year 1 through the beginning of Year 2 of the study.

## Children's Achievement and Personal and Social Development in a First-Year Reading Recovery Program with Teachers-In-Training

**Quay, L. C., Steele, D. C., Johnson, C. I., & Hortman, W. (2001).** ***Literacy Teaching and Learning: An International Journal of Early Reading and Writing, 5,*** **7–25.**

**1. Systematic and Empirical**

This is an experimental study looking at the progress of two equivalent groups of at-risk first graders across the year. The two groups were assigned using quasi-random procedures. In each of 34 schools, one classroom was randomly designated the class from which the Reading Recovery children would be served. Another classroom was randomly designated for selection of the control group. Children with the lowest scores on the Observation Survey were assigned to the two groups. Both groups were administered the Observation Survey and the Iowa Test of Basic Skills (ITBS) in the fall and in the spring. The Gates-MacGinitie Reading Test and the Classroom Teacher Assessment of Student Progress were administered in the spring.

**2. Rigorous Data Analysis**

A multivariate analysis of variance was used to determine the equivalency of the two groups on the fall ITBS. In the spring, a series of multivariate and univariate analysis procedures were used to compare Reading Recovery students and control group students on the ITBS, the Gates-MacGinitie Reading Test, the Observation Survey, and Classroom Teacher Assessment

of Student Progress. Retention rates for the two groups were compared using a chi square.

**3. Valid Data Collection**

The ITBS was administered to all first graders as part of the schools' testing program. The Gates-MacGinitie Reading Test was administered to both groups concurrently. The Observation Survey was administered by teachers with a high level of training.

**4. Strong Research Design**

A multivariate analysis was used to demonstrate that the two groups did not differ on any of the ITBS scales in the fall, confirming equivalence on reading achievement. Analyses were conducted again in the spring after accounting for attrition, and fall scores for remaining subjects were still equivalent for both groups. Random assignment of classrooms in the same schools provided randomization to the extent possible for this type of research. Reading Recovery children were in the regular classroom except for the 30 minutes of Reading Recovery. The control group were not served in Reading Recovery but had access to other programs, with 66% participating in daily literacy groups conducted by the Reading Recovery teachers.

**5. Information Reporting and Expert Scrutiny**

The article was published in a refereed journal, *Literacy Teaching and Learning: An International Journal of Early Reading and Writing.* The study was funded by a foundation that required a concurrent external evaluation. While one author was associated with Reading Recovery, the others were independent (one a widely published research professor; one a statistician; and one a district administrator with experience in measurement, research, statistics, and program evaluation).

**6. Significant Positive Effects**

The two groups were low performing and essentially equivalent at pre-test. At the end of the year, multivariate and univariate analyses of variance indicated that the Reading Recovery children were significantly superior to the control group children on three valid and reliable standard measures: (a) four of the six subtests on the ITBS, (b) all of the subtests of the Gates-MacGinitie, and (c) all of the tasks of the Observation Survey.

The Reading Recovery children were also significantly superior to the control group on all nine measures of the Classroom Teacher Assessment of Student Progress, an instrument developed and used extensively in large-scale evaluations and demonstrating high test-retest reliability. A chi square indicated that a significantly higher percentage of Reading Recovery children than control group children were promoted at the end of Grade 1.

These robust results were obtained with Reading Recovery teachers-in-training, indicating that in-training teachers can be effective in producing results.

In summary, this is a high-quality study comparing two equivalent groups of low-performing first graders, with one group receiving Reading Recovery. Results show a clear advantage for the Reading Recovery children; they performed significantly better on standard measures, and their classroom teachers perceived them to be significantly better in four academic areas and five personal or social attributes. A significant difference was also noted in retention rates, translating to an economic advantage as well.

## Literacy Learning of At-Risk First-Grade Students in The Reading Recovery Early Intervention

**Schwartz, R. M. (2005). *Journal of Educational Psychology, 97*(2), 257-267.**

**1. Systematic and Empirical**

This is an experimental study using random assignment to treatment and control conditions. Thirty-seven Reading Recovery teachers from different schools in 14 states submitted the names of two at-risk first-grade students to a Web-based program for random assignment to first- or second-round Reading Recovery service, and submitted data on those students across the school year that allowed comparison of at-risk students with and without intervention services. In addition data was collected on a low average and a high average student from the same classroom as the two at-risk students. These students (n = 148) were assessed on a variety of literacy measures at the beginning of the school year, at the transition from first to second round Reading Recovery service and at the end of the year.

**2. Rigorous Data Analysis**

Repeated measures analysis of variance with follow-up main effect or simple effect comparisons were conducted. Analyses among groups at the transition period are of primary importance because this provided a comparison of the learning of randomly assigned groups of at-risk students with and without intervention services and a comparison to the progress of average students from the same classrooms.

**3. Valid Data Collection**

Measures include six tasks from Clay's *An Observation Survey of Early Literacy Achievement*. In

addition, at the transition period and at year-end, students were assessed on the Yopp-Singer Phonemic Segmentation task, a sound deletion task, the Degrees of Reading Power Test and the Slosson Oral Reading Test.

**4. Strong Research Design**

This was an experimental design with random assignment of at-risk students to first round intervention services or a comparison group that did not receive intervention service until after the transition period testing. The design also controlled for classroom literacy instruction by selecting all participants from the same classroom within each school.

**5. Information Reporting and Expert Scrutiny**

The article was published in a high quality peer reviewed publication, *The Journal of Educational Psychology*, a research publication of the American Psychological Association.

**6. Significant Positive Effects**

The at-risk students who received Reading Recovery in the first half of the year performed significantly better at the end of their intervention period than at-risk students assigned to receive intervention services later in the year. This is most apparent in the large effect sizes for Text Reading Level (d = 2.02), the Ohio Word test (d = 1.38), Concepts About Print (d = 1.10), Writing Vocabulary (d = 0.90), Hearing and Recording Sounds in Word (d = 1.06), and the Slosson Oral Reading Test-Revised (d = 0.94). Comparisons of the at-risk intervention group with the high-average and low-average classroom groups at the transition period showed the at-risk students had closed the achievement gap with their average peers. A further efficiency analysis showed selection procedures were effective in identifying students in need of early intervention services and that the Reading Recovery intervention could reduce the number of children who appear to need long-term literacy support from 17% to 5% of the first-grade cohort.

## References


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

Clay, M. M. (1991). *Becoming literate: The construction of inner control*. Portsmouth, NH: Heinemann.

Clay, M. M. (1993). *Reading Recovery: A guidebook for teachers in training*. Portsmouth, NH: Heinemann.

Clay, M. M. (2001). *Change over time in children's literacy development*. Portsmouth, NH: Heinemann.

Clay, M. M. (in press). *Literacy lessons designed for individuals part one: Why? when? and how?* Portsmouth, NH: Heinemann.

Clay, M. M. (in press). *Literacy lessons designed for individuals part two: Teaching procedures.* Portsmouth, NH: Heinemann.

Gómez-Bellengé, F. X., Rodgers, E. M., & Fullerton, S. K. (2003). *Reading Recovery and Descubriendo la Lectura: National report, 2001–2002*. Columbus, OH: National Data Evaluation Center.

Pinnell, G. S. (2000). *Reading Recovery: An analysis of a research-based intervention*. Columbus, OH: Reading Recovery Council of North America.

Reading Recovery Council of North America (2003). *Results 2001–2002*. Columbus, OH: Author.

Schmitt, M. B., Askew, B. J., Fountas, I. C., Lyons, C. A., & Pinnell, G. S. (2005). *Changing Futures: The Influence of Reading Recovery in the United States*. Worthington, OH: Reading Recovery Council of North America.

United States Department of Education (2002). *Scientifically based research and the Comprehensive School Reform (CSR) program* (pp. 17–18). Washington, DC: Government Printing Office.


**About the North American Trainers Group**

Within North America, Reading Recovery is supported by more than 20 university training centers where teacher leaders are prepared. Faculty at each of these centers guide the work of affiliated teacher leaders who are located in local school districts and consortia or regional training sites.

This paper was prepared by the Research Committee of the North American Trainers Group.


Reading Recovery® Council of North America

400 W. Wilson Bridge Road, Suite 250 • Worthington, OH 43085 • 614-310-READ (7323) • www.readingrecovery.org