DOCUMENT RESUME

ED 349 762 EC 301 508

AUTHOR	O'Connor, Rollanda E.; And Others
TITLE	Two Approaches to Reading Instruction for Children with Disabilities: Does Program Design Make a Difference?
PUB DATE	[92]
NOTE	21p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 20-24, 1992).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Beginning Reading; *Disabilities; *Instructional Effectiveness; Kindergarten; Kindergarten Children; Phonetics; *Phonics; Primary Education; Reading Achievement; *Reading Instruction; Teaching Methods
IDENTIFIERS	*Direct Instruction; *Distar Reading Program

ABSTRACT

This study tested the hypothesis that Distar Reading's demonstrated effects with disadvantaged children can be generalized to children with disabilities. The study compared the effects of two synthetic phonics reading programs, Direct Instruction "Reading Mastery" (which incorporated features of Distar Reading) and Addison Wesley's "Superkids." The two methods differed considerably in principles of instructional design and exemplified many of the unresolved conflicts in the phonics debate. The two methods were tested in a year-long intervention for 81 children who entered transitional kindergarten special education classes over a 4-year period. No significant achievement differences were evident for either instructional program, either at the end of the treatment year or at follow-up testing 1 year later. However, analysis focusing on children who progressed further in the two reading programs revealed that the Direct Instruction group registered larger reading gains. (Contains 37 references.) (JDD)

Two Approaches to Reading Instruction for Children with Disabilities: Does Program Design Make a Difference?

Rollanda E. O'Connor, Joseph R. Jenkins
Kevin N. Cole, and Paulette E. Mills
University of Washington

Abstract

This study examined the effects on reading achievement of variation in program design and tested the hypothesis that Distar's demonstrated effects with disadvantaged children can be generalized to children with disabilities. We compared the effects of two synthetic phonics reading programs, Direct Instruction *Reading Mastery* and Addison Wesley's *Superkids*, in a year-long intervention for 81 children who entered transitional kindergarten special education classes over a 4-year period. These programs differed considerably in principles of instructional design and exemplified many of the unresolved conflicts in the phonics debate. No significant achievement differences were evident for instructional program either at the end of the treatment year, or on follow-up testing 1 year later. However, an analysis focusing on children who progressed further in the two reading programs revealed that the DI group registered larger reading gains. Our discussion raises questions about design features in early reading programs and suggests another interpretation for the findings from Project Follow Through.

Practices and philosophies about beginning reading instruction vary widely, but there is strong research support for an early emphasis on letter-sound correspondences, especially for children at risk for reading failure (Anderson, Hiebert, Scott, & Wilkinson, 1985; Bond & Dykstra, 1967). Reading methods that include explicit, synthetic phonics instruction--isolated letter sounds and blending sounds into words--result in higher first grade achievement in word recognition and spelling (Adams, 1990) and these effects spread in second grade to comprehension, reading rate, and vocabulary (Chall, 1967). Researchers have investigated individual aspects of phonics instruction, the format, language and ordering of phonics activities (Carnine, 1976 & 1981; Williams & Ackerman, 1971), and used these studies as rationale for a theory of overall program design (Englemann & Carnine, 1982).

One of the most dramatic demonstrations of the effects of a specific reading program occurred as part of the national evaluation of the federally sponsored Project Follow Through, involving 20,000 disadvantaged children nation-wide and 22 different models. One model, Direct Instruction, employed Distar Reading (Engelmann & Bruner, 1974). The Abt Associates report (Stebbins, St. Pierre, Proper, Anderson & Cerva, 1977), in its analysis of Metropolitan Achievement Test reading scores, concluded "Only the children associated with the Direct Instruction Model appear to perform above the expectation determined by the progress of the non-Follow Through children" (p. 155).

Becker (1977) attributed the success of the Direct Instruction (DI) model in Project Follow Through to the design features of Distar Reading, which "utilize[d] advanced programming strategies which are consistent with current behavior theory, but which go beyond current research on task analysis and stimulus control" (Becker & Carnine, 1980, p. 433). The design of DI programs is founded on general case teaching, whereby children learn a small set of examples along with strategies for generalizing to a larger set.

Program Design

One aim in the present study was to examine the contribution of *program design* to the efficacy of beginning reading programs used as early intervention for young children with learning handicaps. We reasoned that the effects of program design ought to be most apparent in studies employing subjects who are just beginning the reading process, particularly children who are predicted to fail without careful instruction, specifically, those children who may have documented learning handicaps, or who are among the wider category of children at risk for learning failure. These children have little prior instructional experience to confound the effects of program design.

Both reading programs examined in our research used a synthetic phonics approach, but differed markedly in instructional design (Carnine, Silbert, & Kameenui, 1990). We included one program, DI's *Reading Mastery I*, because it is based on an explicit theory of instruction (Engelmann & Carnine, 1982) and because its predecessor, Distar, produced remarkably strong

achievement effects with economically disadvantaged youngsters. Our second program was Addison Wesley's *Superkids*. Like *Reading Mastery,* this program introduces letter sounds in isolation, teaches sound blending, and selects reading vocabulary that have regular decodable spellings. However, *Superkids* adopts an entirely different stance on certain other aspects of program design that Adams (1990) has referred to as unresolved dimensions of phonics instruction (e.g., the order of letter-sound introduction and use of letter names), and that Gersten and Carnine (1986) have identified as critical elements in effective instruction (e.g., explicit step-by-step strategies, student mastery, specified error corrections, and formative testing coupled with cumulative review). Below we illustrate the specific design differences in the two reading programs employed in this research.

Introduction of Letters. A basic premise of Engelmann and Carnine's (1982) theory of instruction is the principle of unambiguous communication. One expression of this principle is that the introduction of similar, potentially confusable stimuli should be separated. Using this principle, DI Reading maximally separates letters and sounds that are auditorily or visually similar (e.g., m and n, c and g, i and e) because clustered, they are difficult to discriminate. In contrast, *Superkids* clusters letters with similar visual and auditory features. For example the first three letters presented in *Superkids* (c, o, and g), are not only visually similar, but two of them (c and g) also have similar sounds (e.g., coat/goat). Taking a position in direct contrast to the separation principle, the program's author (Rowland, 1982) asserts that grouping letters that are similarly formed (i.e., 10 of the first 12 letters require circular formation) will facilitate learning.

Letter Names. Although letter name knowledge is one of the best predictors of later reading success, researchers have debated the value of teaching letter names as part of initial reading instruction (Adams, 1990; Hohn & Ehri, 1983; Jenkins, Bausell, & Jenkins, 1972). DI Reading uses only letter sounds through the first year of instruction, its designers arguing that letter sounds have higher utility for blending and reading than their names. In contrast, *Superkids* introduces and tests letter names alongside their sounds.

Explicit step-by-step strategies. DI Reading provides a strategy for each new skill. In teaching blending, children are taught to "sound it out--say it fast." Placing a finger under each sound, the teacher prompts the group to say the sounds slowly in a continuous fashion, then quickly underlines the word with his/her finger to prompt "saying it fast." In contrast, the author of *Superkids* expresses a more relaxed attitude toward strategic learning approaches. "No one has yet discovered that magic ingredient of beginning reading that makes all the parts snap together as a whole. I suspect that just time has a lot to do with it . . . . To some extent you must just keep casting your line over and over again--and we've tried to provide you with some interesting lures" (Rowland, Book 10, p. 1).

Student mastery of each step in the process. Whereas the teacher manual in DI's *Reading Mastery* directs teachers to repeat each task until the children are "firm," that is, they can perform the task without prompts, the *Superkids* manual states that mastery of each letter is not required, as all letters and sounds appear in subsequent letter books.

Specified Error Correction Procedures. DI Reading directs teachers to use specific correction procedures for various categories of reading errors. For example, the following correction procedure is offered for sound blending errors: "If children stop between the sounds at step *b*, stop them immediately. Tell the children what they did (You stopped between the sounds), repeat step *a* (model) and return to step *b* (test), until children are firm (Engelmann & Bruner, 1988, p. 23)."

In contrast, the *Superkids* manual is either vague ("If a child has difficulty, give a hint") or is altogether silent about specific procedures for correcting errors or assisting struggling students.

Formative Testing and Cumulative Review. Tests occur about every 5 teaching days in DI's *Reading Mastery* and are cumulative, i.e., they include items that test skills taught earlier in the program. The teacher is instructed to repeat or delete tasks and lessons, depending on student mastery of the material. *Superkids* provides less frequent tests, every two or three letterbooks (4 to 6 weeks), and the tests are not cumulative, e.g., the first three tests include only those sounds introduced in the just completed letterbooks. On the other hand, *Superkids* provides songs that stress words which begin with the initial sound introduced in current and previous letterbooks, and these songs may serve a review function.

Research on DI Reading in Special Education

We turn now to the use of DI Reading in special education programs. Eager to find methods for their hard-to-teach youngsters, many special educators hoped that DI's strong showing in Project Follow Through could be reproduced with their own special populations. Currently DI *Reading Mastery* is among the reading programs most widely used by special education teachers serving children with mild handicapping conditions.

Overall, empirical support for DI appears impressive (see Fabre's 1983 annotated collection and White's 1988 meta-analysis), but closer examination of the research literature reveals few studies testing the efficacy of DI Reading for young children with learning handicaps. These fall into two categories: analysis of outcomes for low IQ groups in Project Follow Through's database, and comparisons of the relative effectiveness of DI Reading vs. other programs.

Gersten, Becker, Heiry and White (1984) reanalyzed data from Project Follow Through focusing on program effects for students of differing abilities. After blocking students according to their entering scores on the Slossen Intelligence Test, Gersten et al. (1984) reported that even the lowest block of students (IQ scores below 70) made annual gains of 1 year in word recognition on the Wide Range Achievement Test. Favorable results with children low in cognitive abilities

suggest that DI Reading might be of benefit to special education populations, but without direct validation this can only be regarded as an interesting hypothesis.

We could locate just three experimental studies of DI Reading with young handicapped children (Appfel, Kelleher, Lilly and Richardson, 1975; Serwer, Shapiro and Shapiro, 1973; Stein and Goldman, 1980). Of these only Stein and Goldman (1980) found reliable effects favoring DI. Subjects in their year-long study were 63 six to eight-year-old handicapped children diagnosed as having minimal brain dysfunction but with IQ's in the normal range. Treatments were Distar and *Palo Alto*, another phonics-based program. The authors reported significant differences on the Peabody Individualized Achievement Test favoring Distar, which they attributed to differences in program design, specifically DI's phonetic decoding strategies and insistence on mastery of each task, both of which also distinguish DI's *Reading Mastery* from Addison Wesley's *Superkids*.

To summarize, advocacy for using DI Reading programs in special education is based on two interesting but essentially unvalidated hypotheses: (1) DI Reading's adherence to the principles of unambiguous communication and effective instruction makes it a better program for hard-to-teach youngsters; and (2) the positive effects observed in research with nondisabled groups can be generalized to special education populations. In this study we sought to test these hypotheses by comparing DI Reading with a program that differed in instructional design in ways seen as critical by DI theorists. Our research examined both immediate (end of kindergarten) and delayed (end of first grade) achievement outcomes.

Method

Subjects

Over the 4 years of this study, 81 6-year-old children participated in one of two treatments. They were enrolled in transitional kindergartens at the University of Washington's Experimental Education Unit. In Washington state, children qualify for special education by exhibiting a deficit of 2 standard deviations below the norm in one of the following areas, or 1.5 standard deviations below the norm in two areas: cognitive development, language development, gross motor skills, fine motor skills, or social-emotional development. Eligibility testing indicated that 85% of the subjects in this study exhibited delayed language, 49% delayed cognitive development, 64% fine motor delays, 59% gross motor delays, and 56% social-emotional delays. In addition to these categories, 25% also had a medical diagnosis such as cerebral palsy, Down syndrome or seizure disorder. Descriptive statistics which identify age, sex, general ability, and ethnicity for subjects in the two treatments are summarized in Table 1. The groups differed only in ethnicity (chi square with 2 d.f. = 8.10, $p < .05$).

Insert Table 1 about here

Treatments

Subjects attended a transitional kindergarten for 5.5 hours daily, 5 days a week, for 180 school days. Each class was taught by a head teacher with a master's degree in special education and an assistant teacher. Over the 4 years of implementation, each treatment experienced staff changes, two for the DI class and three for the *Superkids* class. The daily schedule, including therapy, playtime, other academic lessons and amount of teacher assistance, was similar for the two conditions. Reading lessons for both treatments lasted 30 minutes daily. Because in each treatment children were instructed in small homogeneous groups of two to four subjects, individual children covered varying amounts of content.

DI Reading. Subjects in this treatment progressed through as few as 50 lessons in *Reading Mastery I* through as much as the first 20 lessons of *Reading Mastery II.* They were taught 13-26 individual letter sounds and 1-5 digraphs, blending of sounds in regularly spelled words, and sentence and story reading.

Superkids. Subjects in this treatment received instruction in 13 to 26 letters (completing between 13 and 18 letterbooks) which introduced sounds in the initial, final, and medial positions, blending of short, regularly spelled words, sentences and stories, and writing and spelling of the reading vocabulary.

Fidelity of Implementation

Interviews with the head teachers in both treatments verified a similar amount of time each day spent in reading instruction and reliance on the teacher manuals and instructions for lesson presentation in both programs. Two of the three DI head teachers received teaching degrees from a program with specific emphasis in Direct Instruction (University of Oregon); a third received inservice training from that program. In addition, a consultant with extensive DI training experience monitored the fidelity with which teachers employed the procedures stipulated in the program.

Teachers of the *Superkids* program did not receive additional training, as it was not recommended by the program's publisher. To foster generalization, the teachers integrated sounds from the current letterbook with other activities outside of the scheduled reading time (usually via first sound matching), e.g., cooking with carrots during the c letterbook.

Measures

Throughout the course of the study we employed one measure of general ability and three measures of reading achievement. The McCarthy Scales of Children's Abilities (McCarthy, 1972) yield verbal, perceptual-performance, and quantitative scores which combine into a general cognitive index (GCI), with a mean of 100 and standard deviation of 16.

The Test of Early Reading (TERA) (Reid, Hresko & Hammill, 1981), norm-referenced for children aged four through 7 years, is an individually administered test which assesses a range of general knowledge of shapes, common symbols, letter names, matching, and word reading. We report raw scores on the TERA, as many children in the study (about one-third of our subjects over all 4 years) had pre- or posttest scores too low to convert to quotients.

The reading portions of the California Achievement Test (CAT) (CTB/McGraw-Hill, 1985), level 10, is a prereading, or readiness, test and gives scores for visual recognition, sound recognition, vocabulary, comprehension, and total reading. The CAT provides fall and spring norms translated to Normal Curve Equivalents (NCEs).

The reading and spelling portions of the Peabody Individual Achievement Test (PIAT) (Dunn & Markwardt, 1970), measure reading recognition, reading comprehension, and spelling. Standard scores are based on a distribution with a mean of 100 and a standard deviation of 15.

Procedures

Research assistants administered the McCarthy Scales individually to all subjects at the beginning of the kindergarten year. They also administered the CAT and the TERA individually as pre- and posttests for each treatment year. Only 55 of the 81 subjects received pre and post CATs, as this test was not introduced until the second year of the study.

Each year as children enrolled in kindergarten, we randomly assigned them to one of two classrooms (14 students each) using either DI Reading or *Superkids*. We excluded children from the study who did not complete a full treatment year because of late enrollment or early departure, leaving 38 in the *Superkids* condition and 43 in DI over the 4 combined years.

Following the treatment year (kindergarten), the children entered first grade in the public school system, either in regular or special education classrooms. We were able to locate 45 of the original sample. Research assistants administered the PIAT to them at the end of their first grade year.

Results

Pretests. As a preliminary analysis, we compared the pretest status of the two treatment groups on GCI and reading scores, using one-way analyses of variance (ANOVA). These tests yielded only one significant difference, the comprehension subtest of the CAT, which favored the *Superkids* group ($F(1,53) = 5.24$, $p < .05$. Tables 2 (TERA) and 3 (CAT) provide descriptive statistics. Although the two groups' McCarthy GCI means did not differ significantly prior to treatment, they were five points apart in favor of the *Superkids* condition. Also, though only one pretreatment difference between groups was significant on any of the reading subtests, all of the reading pretest means were slightly higher for children in the *Superkids* treatment. In testing for treatment effects, we decided to employ analyses of covariance, adjusting post test scores for GCI and the relevant pretest score.

Insert Tables 2 and 3 about here

End of kindergarten results. ANCOVAS yielded no significant differences between the two treatments on any measure, i.e., TERA, $F(1,77)=.06$, ns.; CAT total reading, $F(1,50)=1.73$, ns.; vocabulary, $F(1,50)=1.65$, ns.; sound recognition, $F(1,50)=1.18$, ns.; visual recognition, $F(1,50)=1.57$, ns.; or comprehension, $F(1,50)=0.04$, ns. The regressed adjusted means appear in Tables 2 and 3.

Noting that content coverage varied widely among the subjects in each treatment, we entertained the possibility that completion of a minimum number of lessons within a program might be necessary before program outcomes differed. In a post hoc analysis, we rank ordered children within each treatment, using the point they had reached at the end of the year to mark their progress in the reading curriculum, then split each group at the median of its progress in the curricula (letterbook 13 in *Superkids*; lesson 140 in DI). The "advanced progress" DI subjects significantly outperformed the "limited progress" DI group on CAT total reading ($t = 2.23$, $p < .05$), visual recognition ($t = 4.61$, $p < .001$) and comprehension ($t = 2.66$, $p < .05$) posttests, and on the TERA posttest ($t = 4.52$, $p < .001$). In contrast, limited and advanced progress groups within the *Superkids* treatment did not differ significantly on any of the reading measures. Interpretation of the difference between the advanced and limited progress subjects within DI is clouded by the eight point difference between the two groups on McCarthy GCI scores. Although this difference in GCI was not statistically significant, we cannot be certain that curriculum progress was not confounded with general ability.

A ten point difference in general ability also invalidates comparisons between the two limited progress groups (GCI means of 65.8 and 75.9 for limited progress DI and *Superkids*, respectively) violating the assumption of homogeneous regression coefficients.

We did compute ANCOVAs (adjusting post test scores for GCI and the relevant pretest score) on reading outcomes for the two advanced progress groups, whose GCI means were comparable. Table 4 shows adjusted and unadjusted means and standard deviations for these children. The only significant difference on end-of-kindergarten measures occurred on the sound recognition subtest of the CAT ($F = 5.960$; $p > .05$), favoring DI.

Insert Table 4 about here

One year follow-up. We administered the PIAT in the spring of first grade, a year after children's participation in the treatments. Several of the children moved out of the area, decreasing the size of

our sample to 26 for *DI Reading* and 19 for *Superkids*. Pretest scores for these smaller groups did not differ significantly. Table 5 gives Piat post and adjusted (for GCI) post scores.

Insert Table 5 about here

An examination of post test performance with one-way ANCOVAs using GCI as a covariate revealed no significant treatment differences either on reading recognition, $\underline{F}(1,42) = 2.62$, ns., or comprehension, $\underline{F}(1,42) = 0.01$, ns. On the spelling subtest, however, the DI group performed significantly higher than *Superkids*, $\underline{F}(1,42) = 4.07$, $\underline{p}<.05$. The associated effect size for spelling was .58. Although not statistically significant, the effect size for reading recognition was .50, favoring DI.

Insert Table 6 about here

Table 6 gives the adjusted and unadjusted PIAT means and standard deviations for children in the two advanced progress groups (n = 14 for DI; n = 6 for *Superkids*). Using the adjusted means, the DI advanced progress group scored significantly higher than *Superkids* on PIAT spelling ($\underline{F}(1,20) = 5.581$, $\underline{p} < .05$) and reading recognition ($\underline{F}(1,20) = 5.702$, $\underline{p} < .05$), but not on reading comprehension ($\underline{F}(1,20) = 2.40$, ns.). Effect sizes, all favoring DI were 0.99 for reading recognition, 0.70 for comprehension and 0.98 for spelling.

Discussion

Our results could be viewed as two discrepant sets of findings: (1) no treatment effects between intact groups at the end of kindergarten or first grade, and (2) significant long term treatment effects for subgroups of children who proceeded farther (above the class median) through their reading curriculum. Each set offers different implications for further research and educational practice. Below, we discuss each set of findings.

Findings For the Entire Sample

Children in *both* reading treatments improved in the skills measured during the intervention year. Yet despite pronounced differences in program philosophy and design, the two reading programs yielded similar reading achievement. We entertained three ideas that might be of assistance in interpreting the lack of a predicted advantage for DI Reading: statistical power, test sensitivity, and program design.

Statistical power. White's (1988) meta-analysis of Direct Instruction programs found an average advantage for DI reading of .85 (a large effect size). In a meta-analysis of early intervention research, Castro and Mastropieri (1986) reported higher effect sizes in studies with longer, intense treatments (ranging from .62 to .71 standard deviations for interventions of more

than 2 hours per week and lasting a total of at least 50 hours). The treatments in our study lasted an entire school year, or approximately 90 hours. We conducted a power analysis of our results based on Cohen's (1988) recommendations. For an effect size comparable to that of White's and 40 subjects per treatment, we could anticipate a power of 97%, which is substantial for finding treatment differences in educational research. However, the effect size on CAT total reading was only .21, and the adjusted means on the TERA did not even favor the DI treatment. Because our study qualified as a long and intense treatment by Castro and Mastropieri's standards, and employed enough subjects to have detected an effect size comparable to those reported in other DI studies, an explanation for the lack of treatment effects must lie elsewhere.

Test sensitivity. Interpretation of the present results must be conditioned on the degree to which our tests were sensitive to the treatments. Test sensitivity is grounded in the match between test and curriculum (Jenkins & Pany, 1978). We performed a careful analysis of curricula and tests, but could not detect any bias favoring either reading program. However, neither the CAT nor TERA target letter-sound knowledge or reading regular words, suggesting that they may not have served as highly sensitive, near transfer measures. Although these two tests are commonly used to assess kindergarten reading achievement, we are inclined to consider them to be rather global transfer measures for our two reading programs, tapping a broad array of reading related skills, e.g., word reading, listening comprehension, oral vocabulary, letter naming, and word to word matching. (See Appendix A for the task requirements of each outcome measure and a discussion of floor and ceiling effects.)

However we must underscore one point. Lack of treatment differences cannot be simply dismissed because of concern about test sensitivity. The tests (MAT and WRAT) that demonstrated a strong advantage for DI reading in the original Follow Through evaluation were similarly flawed.

Program design. The theory underlying the design of Direct Instruction programs emphasizes a logical analysis of communication (Engelmann & Carnine, 1982), and the approach itself represents a systematic application of cognitive and behavioral theory to instruction (Butterfield, Slocum, & Nelson, in press). The program is supported by research on features like optimal example sequences, separation of similar features, cumulative introduction of sounds and mastery-based progress, and is consistent with conclusions and recommendations derived from observational and correlational studies that make up the effective teaching literature (Brophy & Evertson, 1974; Rosenshine, 1983). But, given the results of our study, we must entertain the possibility that unspecified features within phonics programs, other than those emphasized by DI theory, have as much impact on learning. Even though our two programs differed on many design features, these differences may not have been of sufficient importance to produce different learning outcomes. Program features of DI are designed to teach the general case, however, the

justification for teaching phonics *at all* (Adams, 1990) is to facilitate generalization. Perhaps, in the beginning stages of learning to read, phonics instruction *is* teaching the general case, minimizing the fundamental differences between the two programs examined in this research. As a related matter, it may be worth noting that the DI model was the only Follow Through model to use a synthetic phonics approach, thus confounding program design and phonics content. That researchers observed treatment effects when phonics and non-phonics reading programs were compared (Follow Through), but did not observe them when two phonics programs were compared (the present study) may indicate that phonics is the critical element, and other design features are less important.

One Year Follow Up. Several researchers have suggested that the effects of early intervention are delayed, and sometimes missed, because of outcome testing that occurs before the full benefit of the intervention is known. In a study of early intervention in phonemic manipulation skills, Lundberg, Frost, and Petersen (1988) found delayed (but no immediate) effects on reading achievement.

We wondered whether the kind of reading program delivered to children in a transitional kindergarten would affect reading achievement in more traditional basal programs in first grade. When we reexamined them at the end of first grade, we still could not detect statistically significant reading differences between the DI and *Superkids* groups, although the DI group performed better in spelling. Two factors combine to make these follow-up results difficult to interpret. The first is lack of observed treatment differences between groups as a whole at the end of kindergarten, which could have been due either to the absence of true treatment effects, or to shortcomings in the measures employed to test for treatment differences. The second is the follow-up results themselves. Statistical tests did not yield significant treatment differences on either reading subtest, but we cannot summarily dismiss the statistically significant effect on spelling, especially considering this difference emerged one year after treatment ended. Despite this hint of a delayed treatment effect, we are left with an inescapable fact: the predicted advantage for DI failed to materialize.

Treatment Effects for Children who Made Advanced Progress in the Curriculum

In a post hoc analysis comparing children who made above average progress in their kindergarten curricula, we found significant differences favoring the DI group on the CAT sound recognition subtest (end of kindergarten) and the PIAT reading recognition and spelling subtests (end of first grade). The reading comprehension subtest, though not significant, also favored DI with an effect size of 0.70. As noted in Table 6, the effect sizes for these differences were substantial.

These results suggest that for children who make greater progress in the curriculum, program design may make a difference in both short and long-term reading outcomes. Three considerations

detract from this finding, however. First, classrooms for kindergarten children with disabilities are typically heterogeneous; children represent a range of handicapping conditions, language and learning skills and family backgrounds. With instruments currently available and with our current knowledge base, it would be difficult, if not impossible, to determine in advance how much progress each child will make in reading, and then choose a reading program to fit that prediction. Second, subdividing our groups according to their progress in the curricula did yield a different pattern of results, but not without a price. The effect of selecting children from each treatment to satisfy a post hoc classification eliminated the experimental advantage that was established through the initial random assignment. Third, the finding is based on an extraordinarily limited sample size (i.e., only 6 subjects in the *Superkids* advanced group). Thus, we advise caution in interpreting this result. All things considered, we are more disposed to use the analyses that include all of the children to evaluate the merits of the two programs.

Lingering Questions

Returning to the two hypotheses tested in this research, i.e., that the instructional design used in DI programs produces better reading achievement and that findings from research with disadvantaged groups can be generalized to children with disabilities, we are left without a clear answer. Arguments supporting the design of DI programs are compelling (Gersten, Woodward, & Darch, 1986), yet it is disturbing to discover the paucity of experiments that examine the relative effectiveness of DI and non-DI reading programs for young children with handicaps. If the design of reading programs makes a difference for anyone, it should make a difference for these children. But before we can confidently specify the features of an appropriate and effective educational program for young children with disabilities, we will need to examine immediate and delayed program effects on the learning of children with specific characteristics.

Whether the relative efficacy of DI Reading for young children with disabilities is limited to "relatively higher performers," whether a one-year treatment period is insufficient to provide young children a foothold in reading, whether past estimates of DI reading's superiority were due to its use of synthetic phonics rather than its specific design features, or whether the measures we employed were not adequate for detecting real treatment differences can be ascertained only by further study. At the very least, our research should alert proponents of any instructional approach to exercise restraint in advocating for specific programs solely on the basis of design features and their presumed benefits.

References

Adams, M. J. (1990). *Beginning to Read: Thinking and learning about print.* Cambridge, MA: the MIT Press.

Anderson, R. C., Hiebert, E. H., Scott, J. A., & Wilkinson, I. A. (1985). *Becoming a Nation of Readers.* Washington, D.C.: The National Institute of Education.

Appfel. J. A., Kelleher, J., Lilly, M. S., & Richardson, R. (1975). Developmental reading for moderately retarded children. *Education and Training of the Mentally Retarded, 10,* 229-235.

Becker, W. C. (1977). Teaching reading and language to the disadvantaged: What we have learned from field research. *Harvard Educational Review, 47,* 518-543.

Becker, W., & Carnine, D. W. (1980). Direct instruction: An effective approach to educational intervention with the disadvantaged and low performers. In B. J. Lahey & A. E. Kazdin (Eds.) *Advances in Clinical Child Psychology, Vol. 3.* Plenum Publishing Corporation.

Bond, G., & Dykstra, R., (1967). The cooperative research program in first-grade reading instruction. *Reading Research Quarterly, 2,* 5-142.

Brophy, J., & Evertson, C. (1974). Process-product correlations in the Texas teacher effectiveness study: Final report (Research Report 74-4). Austin, TX: University of Texas, R & D Center for Teacher Education (ERIC Document Reproduction Service No. ED 099 345).

Butterfield, E., Slocum, T., & Nelson, G. (1992). Cognitive and behavioral analyses of teaching and transfer: Are they different? In D. Detterman & R. Sternberg (Eds.) *Transfer on Trial.* Ablex Pubs.

Carnine, D. W., (1976). Similar sounds separation and cumulative introduction in learning letter sound correspondences. *Journal of Educational Research, 69,* 368-72.

Carnine, D. W., (1981). Reducing training problems associated with visually and auditorily similar correspondences. *Journal of Learning Disabilities, 14,* 276-79.

Carnine, D., Silbert, J., & Kameenui, E. (1990). *Direct Instruction Reading.* Columbus: Charles E. Merrill Publishing Co.

California Achievement Test (1985), Form E, Level 10, Monterey, CA: CTB/McGraw-Hill.

Castro, G., & Mastropieri, M. (1986). The efficacy of early intervention programs: A meta-analysis. *Exceptional Children, 52,* 5, 417-424.

Chall, J. (1967). *Learning to read: The great debate.* New York: McGraw-Hill.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Dunn, L., & Markwardt, F. (1970). *Peabody Individual Achievement Test.* Circle Pines, MN: American Guidance Service.

Engelmann, S., & Bruner, E. (1974). *Reading mastery I: Distar reading.* Chicago: Science Research Associates.

Engelmann, S. & Bruner, E. (1988). *Reading Mastery I: Teachers Guide.* Chicago: Science Research Associates.

Engelmann, S., & Carnine, D. (1982). *Theory of instruction: Principles and applications.* New York: Irvington Publishers, Inc.

Fabre, T. (1983) The application of Direct Instruction in special education: An annotated bibliography. Unpublished manuscript, University of Oregon, Department of Special Education, Eugene.

Gersten, R., Woodward, J., & Darch, C. (1986). Direct instruction: A research-based approach to curriculum design and teaching. *Exceptional Children, 53,* 1, 17-31.

Gersten, R. M., Becker, W. C., Heiry, T. J., & White, W. A. T. (1984). Entry IQ and yearly academic growth of children in Direct Instruction programs: A longitudinal study of low SES children. *Educational Evaluation and Policy Analysis.* 6(2), 109-121.

Gersten, R. & Carnine, D. W., (1986). Direct instruction in reading comprehension. *Educational Leadership, 43,* 7, 70-77.

Hohn, W. E. & Ehri, L. C. (1983). Do alphabet letters help prereaders acquire phonemic segmentation skill. *Journal of Educational Psychology.*

Jenkins, J. R., Bausell, R. B., & Jenkins, L. M. (1972). Comparisons of letter name and letter sound training as transfer variables. *American Educational Research Journal*, 9(1), 75-86.

Jenkins, J. R. & Pany, D. (1978). Standardized achievement tests: How useful for special education? *Exceptional Children*, *44*, 448-453.

Lewkowicz, N. (1980) Phonemic awareness training: What to teach and how to teach it. *Journal of Educational Psychology, 72*, 5, 686-700.

Lundberg, I., Frost, J., & Petersen, O. (1988). Effects of an extensive program fro stimulating phonological awareness in preschool children. *Reading Research Quarterly*, *23*, 3, 263-284.

McCarthy, R. (1972). *McCarthy Scales of Children's Abilities.* New York: The Psychological Corporation.

Reid, D., Hresko, W., & Hammill, D. (1981) *Test of Early Reading* (TERA) Austin, TX: Pro-Ed.

Rosenshine, B. (1983). Teaching functions in instructional programs. *The Elementary School Journal, 83*, 4, 335-351.

Rowland, P. (1982). *Meet the superkids.* Menlo Park, CA: Addison Wesley Publishing Company, Inc.

Serwer, B. L., Shapiro, B. J., & Shapiro, P. P. (1973). The comparative effectiveness of four methods of instruction on the achievement of children with specific learning disabilities. *Journal of Special Education*, *7*, 241-49.

Stebbins, L. B., St. Pierre, R. G., Proper, E. C., Anderson, R. B., & Cerva, T. R. (1977). *Education as experimentation: A planned variation model (vol. IV-A),* An evaluation of Project Follow Through. Cambridge, MA: Abt Associates.

Stein, C., & Goldman, J. (1980). Beginning reading instruction with minimal brain dysfunction. *Journal of Learning Disabilities*, *13*, 4, 219-222.

White, W. A. T. (1988). A meta-analysis of the effects of direct instruction in special education. *Educational Treatment of Children*, *11*, 364-374.

Williams, J., & Ackerman, M. (1971). Simultaneous and successive discrimination of similar letters. *Journal of Educational Psychology*, *62*, 132-137.

Table 1

Subject Characteristics for the Two Treatment Groups

		Gender		Ethnicity[a]			McCarthy GCI		Age	
	n	M	F	C	AA	Oth	M	SD	M	SD
DI Reading	43	30	13	17	25	1	71.8	15.0	6.2	0.37
Superkids	38	25	13	20	12	6	76.7	16.0	6.3	0.40

[a]Ethnicity: C = Caucasian, AA = African American, Oth = All Others.

Table 2

Mean NCEs and Standard Deviations on the Test of Early Reading Ability

	Direct Instruction Reading		Superkids	
	Mean	(SD)	Mean	(SD)
Pre	10.12	(10.7)	10.76	(8.3)
Post	16.21	(10.2)	17.92	(8.8)
Adjusted[a]	16.82	(8.9)	17.23	(5.4)

[a] Scores were adjusted for pretest and the general cognitive index (GCI) from the McCarthy Scales of Children's Abilities.

Note: For DI and Superkids, n's were 43 and 38 respectively.

Table 3

Mean NCEs and Standard Deviations on the California Achievement Test

		DI Reading		Superkids	
		Mean	(SD)	Mean	(SD)
Sound	Pre	33.1	(16.9)	34.9	(16.8)
Recognition	Post	33.9	(19.0)	31.7	(19.2)
	Adjusted[a]	35.1	(14.2)	30.3	(17.1)
Visual	Pre	36.0	(20.4)	45.1	(20.0)
Recognition	Post	34.4	(21.9)	43.0	(15.2)
	Adjusted	37.1	(14.7)	40.0	(13.1)
Vocabulary	Pre	32.4	(11.7)	34.9	(17.0)
	Post	34.6	(13.4)	34.3	(15.4)
	Adjusted	36.0	(10.1)	32.8	(10.3)
Comprehension	Pre	30.7	(16.8)	41.5	(18.1)
	Post	34.3	(15.9)	39.2	(18.5)
	Adjusted	37.4	(10.6)	35.8	(12.5)
Total Reading	Pre	29.5	(12.9)	36.2	(16.5)
	Post	32.5	(13.7)	34.3	(14.4)
	Adjusted	34.8	(9.4)	31.6	(8.1)

Note: n's were 29 (DI) and 26 (Superkids) on all subtests.

[a] Scores were adjusted for pretest and the general cognitive index (GCI) from the McCarthy Scales of Children's Abilities.

Table 4

End-of-Kindergarten Post Test Means and Standard Deviations for Advanced Progress Groups

	DI Reading			Superkids		
	Mean	(SD)		Mean	(SD)	
Age	6.2	(.35)		6.2	(.27)	
McCarthy GCI	73.7	(10.7)		75.9	(14.0)	
	Mean	(SD)	Adj[a]	Mean	(SD)	Adj
California Achievement Test[b]						
Sound Recognition	41.8	(14.2)	41.5	28.8	(13.2)	29.0
Visual Recognition	51.7	(16.8)	50.3	41.4	(16.9)	42.6
Vocabulary	39.0	(13.4)	39.4	37.8	(13.4)	37.5
Comprehension	43.1	(13.6)	43.1	39.2	(18.8)	39.2
Total Reading	39.0	(11.9)	39.3	36.7	(13.5)	36.4
Test of Early Reading Ability[c]	21.8	(9.5)	20.8	16.5	(4.4)	17.7

a Scores were adjusted for the GCI from the McCarthy Scales of Children's Abilities and the relevant pretest score.

b Normal Curve Equivalent scores for reading subtests of the California Achievement Test, level 10. n's for the CAT, administered in the last three years of the study, were 12 (DI) and 13 (Superkids).

c n's for the TERA, administered all four years, were 21 (DI) and 17 (Superkids).

Table 5

Means and Standard Deviations on the Peabody Individual Achievement Test

	Direct Instruction Reading			Superkids			
	Mean	(SD)	Adj[a]	Mean	(SD)	Adj	Effect size
Reading Recognition	94.50	(11.8)	94.34	90.05	(10.8)	89.05	0.50
Reading Comprehension	91.65	(11.3)	91.94	92.0	(9.2)	91.61	0.08
Spelling	95.35	(10.7)	95.21	90.11	(9.3)	89.30	0.58

Note: n's were 26 (DI) and 19 (*Superkids*) on all subtests.

Table 6

Means and Standard Deviations on Peabody Individual Achievement Test for Advanced Progress Group

		DI Reading			Superkids			
		Mean	(SD)	Adj	Mean	(SD)	Adj	Effect size
PIAT	Reading Recognition	102.0	(10.5)	102.0	92.2	(5.1)	92.3	1.26
	Comprehension	98.5	(10.7)	98.5	91.3	(7.9)	91.4	0.77
	Spelling	100.9	(11.5)	100.8	89.5	(7.6)	89.6	1.19

[a] Scores were adjusted for GCI on the McCarthy Scales of Children's Abilities.

N's for the follow-up sample were 14 (DI) and 6 (Superkids).