

RUNNING HEAD: The Superkids Effectiveness

The Effectiveness of the Superkids on Student

Achievement and Teacher Outcomes

Jerome V. D'Agostino

Aryn Karpinski

The Ohio State University

Abstract

The Superkids is a comprehensive language arts and reading program for kindergarten and first grade students that emphasizes phonemic awareness, phonics, fluency, and comprehension skill development. Program effectiveness in terms of student achievement was ascertained by implementing a multilevel group-comparison design in which the mean language arts posttest scores from 21 treatment and 22 control classrooms were analyzed after adjusting for student and classroom prior differences. Teachers were asked to report on the effectiveness and time usage of their core reading programs as well. Students in Superkids classrooms outperformed their peers in comparison classrooms on measures of reading comprehension ($d=.24$), word analysis ($d=.41$), reading words ($d=.23$), and total reading ($d=.27$), but not on a vocabulary and a language test. Superkids was found to be easier and more satisfying to use, and more effective at teaching and motivating students than core programs implemented by control teachers.

The Effectiveness of the Superkids on Student Achievement and Teacher Outcomes

Reading proficiently is the most critical accomplishment for the young learner. Without the capability to read, students' development of all academic skills is greatly impeded. Though the percentage of 4th-grade students who fell below the basic performance level on the National Assessment of Educational Progress (NAEP) decreased from 1992 to 2007 from 38 to 33, still a third of all fourth grade students in the country are not functional readers (U.S. Department of Education, 2008). But it is far too late to wait until the third or fourth grade to address a child's reading problems (Pikulski, 1994; Spiegel, 1995). Numerous studies have documented that interventions for kindergarten or first grade students are critical to bolster their developing reading skills (National Reading Panel, 2000).

Superkids is a comprehensive reading and language arts program for beginning readers. The program was developed by Pleasant Rowland in 1978 based on her extensive experience as an elementary school teacher. Originally published by Addison-Wesley, the program was not marketed extensively but remained in use by many committed schools. The Rowland Reading Foundation was created in 2003 to update and disseminate the Superkids materials. Within the last five years, several public and private schools have adopted the program and implemented it successfully. In this study, we examined the effectiveness of the Superkids in terms of improving students' language arts achievement levels and implementation efficiency.

The Superkids Program

Congruent with research findings regarding the effective characteristics of early literacy programs (see National Reading Council, 2000; Pressley, Wharton-McDonald, Allington, Block, Morrow, et al., 2001; Snow, Burns, & Griffin, 1998), Superkids addresses thirteen critical strands taught in concert, including: (1) phonemic awareness; (2) phonics; (3) fluency; (4)

comprehension; (5) vocabulary; (6) listening and speaking; (7) handwriting; (8) spelling; (9) expressive writing; (10) early literacy; (11) grammar usage and mechanics; (12) structural analysis; and (13) study skills (see Rowland Reading Foundation, 2007, for a more detailed program description). The strands are taught within four levels consisting of 300 sequential lessons that span kindergarten and first grade. Levels 1 and 2 are primarily targeted for kindergarten students and Levels 3 and 4 usually are covered in the first grade. Meet the Superkids, Level 1, contains 85 lessons that focus on 13 letters, five short vowels, and 8 consonants. Students systematically learn how to read and write these letters of the alphabet, letter-sound associations for them, and several memory words. They learn how to blend sounds to read simple words. By the end of this level, students begin to understand how the written language they are learning relates to the spoken language they already know. During Level 2, The Superkids' Club, which consists of Lesson 86-145, students begin to learn letter-sound associations for the remaining 13 letters of the alphabet. Children start to use their new decoding abilities to read independently and to develop comprehension skills and strategies.

Lessons 146-237, contained within Level 3, Adventures of the Superkids, focus on more complex letter-sound relationships, such as digraphs and long-vowel sounds. Students continue to develop their decoding and encoding skills and read longer stories with comprehension and fluency emphasized. In Level 4, More Adventures of the Superkids, comprised of Lessons 238-300, students continue to build skills in all areas of language arts while learning more complex phonetics and how to identify words in context.

Certain instructional approaches and methods are integrated into the lessons to promote students' literacy development. To simultaneously teach reading and writing, letter-sound correspondences are organized in a strategic sequence based on letter shape and similarity.

Kindergarten students learn one sound for each letter, and each letter-sound connection is introduced, practiced, and reinforced over the course of five to seven lessons. Students have many opportunities to review and practice learned skills both individually and in small groups, and to apply their knowledge of letter-sound connections while reading and writing words, sentences, and stories. The program is designed to limit student word guessing by presenting text with controlled vocabulary in a hierarchical manner.

Fluency is developed in a multistep fashion. First it is modeled by a narrator on CDs and then by the teacher. This phase is followed by children repeating phrases and sentences from read-aloud stories. After connected text in story format is presented to students, they have the opportunity to learn discrete fluency skills, such as proper phrase reading and the observation of punctuation marks. Fluency lessons progress from explicit, direct instruction to oral reading practice monitored by the teacher in small groups or whole class. Oral reading also is promoted through choral activities, Readers' Theatre, and plays.

Children are introduced to characters (the Superkids) who continue through the instructional materials. Student books are enriched with illustrations designed to motivate discussion, provide common background, and introduce speaking and listening vocabulary. Fiction and nonfiction books with phonetically controlled vocabulary to support comprehension are introduced in the second level. Discrete comprehension skills, such as recognizing story elements, identifying main idea, drawing conclusions, and making inferences, are emphasized. Students are taught to use these skills to bolster their understanding and to rectify any misunderstandings.

Research on The Superkids

Recently, a number of impact evaluations have been conducted to examine the effectiveness of Superkids. In a study conducted in one large Texas school district, first-grade students from a school that had implemented Superkids outperformed comparison children at three comparable schools on the Stanford Achievement Test, 9th edition during the 2003-2004 school year (Institute for the Advancement of Research in Education, 2004). In the next school year, the Rowland Reading Foundation performed a study in three schools in which three Superkids classrooms were matched with similar classrooms that did not use the program. Superkids students scored significantly higher on phonological processing and graphophonemic tests relative to controls, with a standardized mean difference effect size of .57.

In another study conducted the following year, 2005-2006, the Center for Research in Educational Policy (2007) compared the scores on a battery of language arts tests from 152 treatment students in eight classrooms in two schools to the scores of 120 comparison students from seven classrooms in two schools. Treatment students outperformed control students on three of the 12 subtests that comprised the battery, with effect sizes ranging from .32 to .37. Comparison students, however, scored significantly higher on the DIBELS Phone Segmentation Fluency exam, with an effect size of -.36.

These prior studies were rather small, relied on convenience control groups, and were based on statistical analyses conducted at the student level. Borman and Dowling (2007) improved on the designs of these past studies by recruiting 23 kindergarten classrooms in various states to implement Superkids that were matched more precisely with control classrooms based on a computer algorithm. Using hierarchical linear modeling (HLM), they found that treatment classrooms scored significantly higher on all five subtests of the Stanford Achievement Test, 10th edition, with effect sizes ranging from .11 to .25. Though Superkids teachers spent less time per

day on language arts instruction, they spent more time per day using their core reading program than did comparison teachers. Treatment teachers' satisfaction levels with Superkids also were significantly higher than control teachers' satisfaction levels with their respective core language arts programs. The teacher report effects were quite sizable, ranging from .72 to 1.49.

Study Purpose

The primary purposes of the present study were to: (1) build on the extant Superkids research base by utilizing a slightly different multilevel quasi-experimental design; (2) examine if program effects generalized to another set of student outcomes, specifically the Iowa Test of Basic Skills (ITBS); and to (3) ascertain teachers' evaluations of Superkids and other language arts programs. Pretest data consisting of treatment and control kindergarten students' scores on a fall 2007 ITBS Word Analysis subtest and teacher and student background characteristics were collected in September 2007. Students were administered language arts subtests of the ITBS in the spring of 2008, which served as posttest measures. To adjust for any potential group differences, the fall Word Analysis scores and student and classroom characteristics were used to statistically equate the groups. HLM was employed to analyze the classroom nested data from this quasi-experimental design. Teachers were asked to report their satisfaction levels, impressions, and time usage of their core language arts programs at the end of the 2007-2008 school year.

Based on prior research findings, it was hypothesized that:

1. Superkids classrooms would have significantly higher average scores on the ITBS spring subtests than control classrooms after controlling for fall achievement levels and student and classroom characteristics, and

2. Superkids teachers would report higher levels of satisfaction, higher levels of effectiveness, and more time usage with their core program, but less overall time teaching language arts than control teachers.

Method

Participants

In the spring of 2007, Rowland Reading Foundation staff asked potential public and parochial schools in various regions of the country to apply to be part of a pilot program. Of the 51 schools accepted as pilots, 11 were chosen to participate in the study based on the need to have a sample that reflected the nation's demographic composition. Once treatment schools were selected, a control school was recruited for each treatment site based on similar demographic characteristics. The 11 control schools were offered grants of \$500-1,500 per school, depending on the number of participating teachers.

There were 43 kindergarten classrooms (31 public, 12 Catholic), 21 (15 public, 6 Catholic) treatment and 22 (16 public, 6 Catholic) control, nested in the 22 (12 public, 10 Catholic) sampled schools. The sites were located in urban, suburban, and rural districts throughout the country, including Avalon, Blue Ridge, Corpus Christi, Fort Worth, and Stockdale, Texas, Jacksonville, Florida, Columbus, Kansas, Uhrichsville, Ohio, Tucson, Arizona, Baltimore, Maryland, and Kansas City, Missouri.

During late September 2007, participating students were administered the Word Analysis subtest (Level 5, Form A) of the ITBS as a pretest measure of their language arts skills. The great majority of kindergarten students in study classrooms, 829 in all (426 students representing the treatment and 403 representing the control), took the fall pretest. In the spring of 2008, students were asked to take a battery of language arts ITBS exams as posttest measures. Not all students

who took the pretest had posttest scores. The final sample reduced to 750 total students, with 382 treatment and 368 control students. The loss of students from the initial to final sample was about nine percent.

Table 1 provides initial and final sample demographic characteristics, including ethnicity, eligibility for the federal free or reduced-price lunch program (FRL), English-language learner status (ELL), and minority status (African-American, Hispanic, Native-American, and Pacific Islander students). Average fall pretest scores by group also are provided in the table. These figures allow for the comparison of the final treatment and control student groups and to assess any potential effects due to student pretest to posttest attrition. By comparing the figures across variables for the initial and final total samples, it is evident that the percentages of students representing each ethnic group, FRL, ELL, and minority statuses are very comparable. The average fall pretest scores are equivalent as well. Apparently, student attrition was not systematic and did not lead to a biased final sample relative to the sample of students who took the pretest.

The figures for the final control and treatment groups from Table 1 reveal some differences between the two groups. A slightly larger percentage of treatment students were of African-American descent, but a slightly greater proportion of control students were FRL status. Perhaps the most marked difference between the two groups was in terms of pretest score—the treatment group scored, on average, two NCE points higher on the fall Word Analysis test compared to the control group, though the difference was not statistically significant, $t(748) = 1.65, p > .05$. Nonetheless, these findings evinced the need to adjust for group differences during the statistical analyses.

No teachers or classrooms dropped from the study during the 2007-2008 school year. Table 2 provides treatment and control group characteristics at the classroom level. Percentages

of teachers who had Master's degrees and certification in reading are provided. Average years teaching, including the present year, average class sizes, average classroom mean NCE pretest scores, and mean classroom percentages of FRL, ELL, and minority students are presented as well. As can be seen from the table, only a minority of teachers overall had reading certification, but a larger percentage of treatment teachers were certified in reading compared to the control group. Further, a slightly larger percentage of treatment teachers held Master's degrees, and treatment teachers, on average, had about one more year of teaching experience. Treatment classrooms also had slightly larger enrollments, greater proportions of minority students, and higher pretest scores, on average, than control classrooms. As was the case at the student level, control classrooms had larger percentages of FRL eligible students, on average, compared to treatment classrooms. The two groups were comparable, however, in terms of percentages of ELL students, on average. The data presented in Table 2 revealed that treatment and control classrooms were similar demographically, but not identical.

Measures

Student Achievement. During late September 2007, participating students were administered the Word Analysis subtest (Level 5, Form A) of the ITBS. Though Word Analysis likely did not fully represent the ensemble of language arts skills, multiple ITBS subtests were not administered in the fall to encourage study participation. Word Analysis was chosen as the sole pretest measure because it reflected the primary language arts skills learned by kindergarten students in the beginning of the school year. During the spring of 2008, participating students were asked to take the Vocabulary, Reading Comprehension, Word Analysis, Reading Words, and Language subtests of the ITBS, Level 5, as posttest measures. Students' Total Reading

scores, which were comprised of their Reading Comprehension and Reading Words subtests, also were used as an outcome.

Covariates. As indicated from the student and classroom characteristic comparisons between the control and treatment groups (Tables 1 and 2), it was imperative to control for differences on variables that potentially could have influenced students' posttest achievement scores. All variables reported in Tables 1 and 2 at the student and classroom levels were considered as potential covariates. Students' FRL, ELL, and minority statuses, and their fall ITBS Word Analysis scores were treated as possible individual-level control variables. Taken together, these measures reflected students' degree of socioeconomic privilege and prior language arts achievement prior to the onset of the treatment. Classroom and teacher characteristics also had to be considered as possible covariates to control for potential environmental and teacher skill effects. Teachers' years of experience, Master's degree and reading certification statuses, the proportions of minority, ELL, and FRL students per classroom, average classroom fall Word Analysis scores, and class size all were considered as potential control variables.

Teacher report variables. Teachers were interviewed in spring, 2008 to ascertain their satisfaction with and usage of their reading and language arts curricular materials. They were asked to rate on a five- point scale the effectiveness of their current programs in terms of: (1) satisfaction with the current core reading program; (2) effectiveness of the current core reading program at teaching students how to read; (3) effectiveness of the current core reading program at motivating students to learn; and (4) ease of use of the current core reading program. Teachers also were asked to report the amount of minutes per day they typically spent teaching language arts, including reading and writing, and to report the amount of minutes per day they spent using

their core language arts programs (for the treatment group, the core program was Superkids). A variable that indicated the proportion of total language arts instructional time spent using the core program was computed by dividing the latter variable by the former one.

Procedures

Treatment teachers were asked to implement Levels 1 and 2 (Meet the Superkids and Superkids' Club) during the 2007-2008 school year. They were provided with a guide that stipulated seven essential (e.g., teach all the steps and the Daily Routines in each lesson; teach only one sound for each letter as represented in the lessons) and six recommended (e.g., reinforce and enrich learning with Ten-Minute Tuck-in activities; differentiate instruction in small groups as appropriate) best practices.

In the spring, 2008, the treatment teachers reported that, on average, 136 ($SD = 18.38$) of the possible 145 lessons that comprise Levels 1 and 2 were taught to their students. Fourteen of the 21 treatment teachers reported teaching at least 139 lessons, and five of those teachers reported covering all 145 lessons. One teacher claimed to teach only 63 of the lessons. The second lowest reported number of lessons taught by a teacher was 127.

Control teachers implemented the language arts programs that were adopted by their respective schools. Among the 21 of 22 control teachers who provided a description of their curricular materials, the Open Court series was used in four classrooms and the Scholastic, Fountas and Pinnell series was implemented in three additional rooms. The Sunform, Guided Reading program was used in another three rooms, while Saxon Phonics also was used in three rooms. The McGraw Hill series was adopted in two rooms, and the Houghton Mifflin Reading Program, MacMillan textbook, Scott Foresman textbook, SRE, University of Columbia's

Literacy program, and the Reading A to Z internet-based program were implemented in one classroom each.

Statistical Analyses

For the analysis of student outcomes, hierarchical linear modeling (HLM) was employed. The six ITBS posttest measures served as outcomes, and prior group differences were controlled by including various covariates. To examine teacher report mean differences among treatment and control groups, t-tests were conducted on each outcome.

Results

Student achievement findings. Preliminary analyses of student test scores were conducted by comparing treatment and comparison group means. Table 3 presents the average normal curve equivalent (NCE) fall Word Analysis and spring subtest scores of the entire final sample and by group. The table also provides the standard deviations of each NCE mean, the means converted to percentile scores, and the sample sizes for which the means were based. Note that treatment students, on average, were slightly more proficient during the fall than were the control students (the fall effect size was .11), but on most of the spring tests, the treatment-control gap widened rather considerably. The effect size in spring for Reading Words was .28, the effect for Word Analysis was .45, for Reading Comprehension it was .31, and for Total Reading it was .32. On Vocabulary and Language, however, the gap was virtually non-existent.

These comparisons are potentially misleading, though, because prior group differences were not accounted for and comparing overall student-level means assumes that all students received uniform instructional experiences in the treatment and control conditions, respectively. For these reasons, HLM was employed to handle the nesting of students within classrooms, and covariates were used to adjust for prior treatment and comparison differences.

Before conducting the final HLM analyses, initial models were generated to determine which student and classroom level variables worked best to control for prior group differences. Students' minority, FRL, and ELL status, and their fall pretest scores were considered, as were classroom proportions of minority, FRL, and ELL students, classroom average pretest scores, and class size. Also considered at the classroom level were teacher characteristics such as holding a Master's degree, reading certification, and years of service. As was found in the Borman and Dowling (2007) study, the teacher characteristics did not significantly predict classroom average spring test scores once the other variables were included. Neither did class size. Consequently, the teacher characteristic variables and class size were not used as covariates for the final HLM analyses.

It is informative to examine the correlations among variables before conducting HLM analyses. Tables 4 and 5 include the correlations at the individual and classroom levels, respectively. The variable Superkids was a dummy variable coded "0" for comparison student or classroom and "1" for treatment student or classroom. In Table 4, the student level correlations, it is evident that fall Word Analysis pretest scores were correlated from about .40 to .60 with the spring outcome measures, revealing the effectiveness of the fall subtest as an indicator of prior achievement. Indeed, the fall scores are as correlated with some of the spring outcomes as the spring scores are intercorrelated. The correlations among outcomes ranged from about .30 to .60, except for Reading Comprehension and Reading Words with Total Reading (which are highly correlated because the two subtests comprise the composite). The lack of relation between the outcomes perhaps indicates the diversity of skills that define language arts and reading, and evinced the need to have multiple outcome indicators. It also can be seen from Table 4 that Superkids was correlated moderately with the outcomes except for Vocabulary and Language.

Table 5, the classroom level variable correlations, reveals that classrooms with larger proportions of minority students also tended to contain larger portions of FRL students, and were slightly more likely to contain larger proportions of ELL students. Classroom fall Word Analysis averages were most correlated with proportion of FRL students—classrooms with greater proportions of FRL students tended to have lower average fall scores.

A separate HLM analysis was conducted for each ITBS outcome. All student level and classroom level covariates were included in the models regardless of whether the variables were statistically significant in all cases. The fall Word Analysis and all outcome test scores were converted to linear z-scores. By standardizing the outcomes in this manner, and by grand mean centering all covariates, the HLM Superkids coefficient closely approximated the Hedges *g* effect size for multilevel data analyses. As a check, the actual Hedges *g* values also were computed by dividing the Superkids variable coefficient from each model by the student-level pooled variance of each outcome (see Institute of Educational Sciences, 2008). In all models, the Superkids coefficient values equaled the Hedges *g* statistics.

Table 6 presents the final HLM results by the six outcomes. The unstandardized coefficient and standard error of the coefficient (se) for each variable are included. Significant t-ratios are denoted with asterisks. The t-ratios are computed by dividing the coefficients by the respective standard errors. A t-ratio value greater than about ± 2.00 indicates that the coefficient was significantly greater than zero ($\alpha < .05$), and thus, the variable significantly predicted the outcome. The top portion of the table provides the results for the classroom-level variables, whereas the bottom portion presents the results for the student-level covariates. In the top portion, the intercept for the mean classroom outcome (Intercept, γ_{00}) simply indicates if the average classroom mean is greater than zero. Because the outcomes were standardized, the

average classroom means were expected to be near zero. Thus, none were statistically significant.

In Table 6, it can be seen that classrooms with greater percentages of minority students had lower classroom average Vocabulary scores. Also, classrooms with greater average fall Word Analysis scores had greater average spring Vocabulary scores. No other level-two covariate was significant. All four student-level covariates were significant predictors of students' spring vocabulary scores—FRL, minority, and ELL students tended to have lower Vocabulary scores and students with larger fall Word Analysis scores tended to have larger Vocabulary scores. These results are indicated by the intercepts of the average slopes for each student-level covariate in the bottom half of the table. The slopes in this case are the correlations between each covariate and the outcome at the individual level.

The coefficient of .02 for Superkids in Table 6 reveals that the average adjusted Vocabulary mean of Superkids classrooms was .02 standard deviation units larger than the average comparison classroom. The classroom means have been adjusted based on the student-level covariates. The Superkids variable was not significant, indicating no difference in the adjusted classroom vocabulary classroom means between the intervention and control groups.

Table 6 also indicates that fall student-level Word Analysis was the only significant covariate for Reading Comprehension. The Superkids coefficient of .24 was significant, $p < .05$, revealing that Superkids classrooms had about a quarter standard deviation larger average adjusted mean than comparison classrooms. The model for Word Analysis, presented in Table 6, indicates that several covariates were significant, including student-level ELL, FRL, and fall Word Analysis, and classroom-level percent minority and percent ELL. The Superkids variable

also was significant—treatment classrooms outperformed comparison classrooms, on average, by .41 of a standard deviation.

Table 6 reveals a .23 effect for Superkids classrooms on Reading Words after controlling for the covariates, mainly classroom-level percent ELL, and student-level FRL and fall Word Analysis. The Superkids effect of .27 on Total Reading was significant, $p < .05$, (Table 6) after adjusting for the covariates, especially student-level FRL and fall Word Analysis. Table 6 indicates that Superkids was not a predictor of adjusted average Language classroom means. The student-level FRL, minority, ELL, and fall Word Analysis variables all were significant covariates of Language scores.

Summary information for all six HLM models is presented in Table 7. The first row in the table provides the average classroom mean reliability values. Reliability in this case primarily is driven by classroom size, and values of .70 or above commonly indicate adequate reliability. The reliability values ranged from .73 to .87. The next row in the table provides the proportion of total variance for each outcome that was explained by students' classroom membership. Larger values indicate greater diversity of classroom means, or that classroom membership made more of a difference in terms of students' outcome scores. The values range from .18 to .33, revealing that about one fifth to one third of the total outcome variances were accounted for by the classroom in which each student was enrolled. These values are large in relation to many other multilevel studies of educational programs.

For any outcome, however, a large proportion of the classroom mean variance likely was explained by the student-level covariates. The third row reveals the degree of between-classroom variability that was accounted for by the student-level demographics and fall achievement levels. The first cell entry in the third row indicates the proportion of total variance explained by

classroom membership after adjusting for the level-one covariates. The second cell entry indexes the percent of between classroom variance reduced by the inclusion of the student-level variables. It can be seen that between classroom variance was reduced across the models by about 20 to 40 percent. Nonetheless, the inclusion of the student variables did not account for all or most of the classroom differences, as indicated by the proportion of between classroom variance (i.e., 60 to 80 percent) that remained after the inclusion of those indicators.

The values in the fourth row indicate the proportion of adjusted between classroom variance (i.e., between classroom variance remaining after the level-one covariates were included) explained by the Superkids variable without the level-two covariates in the model. Essentially, the fourth row r^2 values are indicators of the treatment effect alone. For Vocabulary, the value was zero, but varied from two to 16 percent for the other outcomes. The fifth row of Table 7 provides the proportion of adjusted between classroom variance explained by Superkids and the level-two covariates added as predictors in each model. It is evident that including the level-two covariates increased the accounted for variance for most outcomes. The last two rows of the table provide the standardized mean difference effect size, or Hedges g values, per outcome, and the accompanying percentile standing in the student level distribution of the average treatment classroom mean. For example, the average Superkids classroom hypothetically would have performed at the 66th percentile level on Word Analysis in the student distribution, which is indicative of a sizable group difference.

Teacher-report findings. The means and standard deviation by group and for each teacher reported variable are provided in Table 8. The t-test results, standardized mean difference effect sizes, and percentile standing of the average treatment teacher in the control group distribution also are presented in the table. It is evident from the table that treatment teachers reported

significantly less time teaching language arts, including reading and writing, compared to the control group, but the treatment teachers also reported spending significantly more minutes per day using their core program (i.e., Superkids) relative to the control teachers. Thus, the average proportion of total language arts time spent using the core program was significantly larger for treatment teachers compared to their counterparts.

The Superkids teachers, on average, reported significantly greater satisfaction with their core program than did control teachers. According to the teachers, Superkids was significantly more effective at teaching students how to read and at motivating students to learn than core programs implemented in the control classrooms, as evinced by the t-test results presented in Table 8. Treatment teachers also had a significantly higher average score on the question pertaining to ease of use of their core program than did control teachers.

Discussion

The primary purpose of this quasi-experiment was to examine the effectiveness of Superkids when implemented by regular classroom teachers and while comparing program students' outcomes to those of students in classrooms in which other language arts programs had been implemented. Though students were not assigned randomly to classrooms, and classrooms were not assigned at random to treatment or comparison conditions, all efforts were made to statistically adjust outcomes based on prior student and classroom achievement and characteristics. This process increases the credibility of the causal claims.

Though about 9 percent of the initial sample was lost at posttest, attrition did not seem to be systematic. The student demographics of the initial and final samples were very similar, indicating that the final sample was a good representation of all students enrolled in the participating classrooms. Nonetheless, although all efforts were made to create comparable

treatment and comparison groups, the treatment group appeared to be slightly more advantaged in terms of student background characteristics, fall achievement levels, and teacher characteristics, such as years of service, obtaining a Master's degree, and reading certification level. Preliminary analyses indicated, however, that years of teaching, having an advanced degree, and being certified in reading were not significant predictors of classroom mean outcome values after considering the other level-two background variables that ultimately were used in the analyses. The student-level characteristics in which the groups differed did predict the outcomes, and thus, were included as covariates.

The HLM analyses indicated that Superkids classrooms outperformed comparison classrooms after adjusting for the key covariates on the Reading Comprehension, Word Analysis, Reading Words, and Total Reading ITBS tests. The groups did not differ significantly on the Vocabulary and Language tests, perhaps indicating the Superkids is no more effective than other programs at promoting kindergarten students' vocabulary skills and language proficiencies regarding using prepositions, comparatives and superlatives, and singular-plural distinctions.

Superkids apparently is more effective than other programs implemented in the comparison group at facilitating students' proficiencies related to analyzing and reading words and at comprehending what they are reading. Treatment students might have demonstrated higher achievement levels in both word analysis and comprehension because they were provided a stronger foundation in word decoding, which is necessary to understand written text. The comprehensive approach of Superkids, which involves a combination of alphabetical skill development with comprehension building, seemed to be advantageous to students across a full spectrum of language arts outcomes.

Two measures of program effect were provided: standardized adjusted mean differences between the treatment and control, and the amount of between-group variance explained by the treatment. Standardized mean differences between .20 and .50 commonly are defined as small to medium effect sizes (Cohen, 1988), and medium r^2 values typically range from .09 to .25. For all four significant outcomes, the effect size indicators were within these ranges, with Superkids demonstrating the largest effect on Word Analysis.

Perhaps the most suitable means of interpreting effects, however, is to compare effect sizes to normative expectations or the effects of other programs designed to reach similar goals (Hill, Black, Bloom, & Lipsey, 2007; Konstantopoulos & Hedges, 2008). Among the two core programs for beginning readers that the What Works Clearinghouse (2009) has reviewed, the average effect of .22 (range .01 to .41) found in this study for Superkids was slightly larger compared to one reviewed program ($M = .20$), and considerably higher than the other intervention ($M = .09$). Furthermore, the two reviewed programs yielded lower bound effect sizes for some outcomes that were negative in value. Thus, in comparison with at least two programs with adequate empirical evidence, Superkids produced rather favorable achievement effects.

Yet even larger effects were yielded for the teacher reported variables. As was found in Borman and Dowling (2007), Superkids teachers actually reported less time spent per day on language arts instruction than did control teachers, but the Superkids teachers claimed to use their core program for more minutes per day, and thus, had a significantly larger average proportion of total language arts instructional time that was devoted to their core program materials relative to their peers. Ostensibly, teachers rely on fewer supplemental materials to teach language arts with Superkids compared to many other available programs, perhaps

reflecting the more inclusive nature of the Superkids materials. Teachers seemingly feel that Superkids is a more complete package that requires far less supplementation.

Based on the effect sizes that ranged from .83 to 1.32, treatment teachers demonstratively reported greater average levels of satisfaction, ease of use, and effectiveness regarding Superkids than control teachers reported about their core programs. Given that students who learned language arts with Superkids outperformed students taught with other programs on the majority of outcome measures, the teachers' appraisals seemingly were accurate. The evidence collected in this study indicates that Superkids is a complete program that is easy and efficient to implement, and that produces positive learning outcomes that are most vital for student success.

References

- Borman, G.D., & Dowling, N.M. (2007). *Student and teacher outcomes of The Superkids Quasi-Experimental Study*. Unpublished manuscript.
- Center for Research in Educational Policy. (2006). The effects of the Superkids program on beginning reading achievement in kindergarten. Unpublished manuscript.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, 2nd edition. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hill, C.J., Bloom, H.S., Black, A.R., & Lipsey, M. W. (2007). *Empirical benchmarks for interpreting effect sizes in research*. New York: MDRC.
- Institute of Educational Sciences. (2009). *What Works Clearinghouse beginning reading programs*. Retrieved on January 30, 2009 from <http://ies.ed.gov/ncee/wwc/reports/topic.aspx?tid=01>.
- Institute of Educational Sciences. (2008). *What Works Clearinghouse: Procedures and Standards Handbook* (Version 2.0). Washington, DC: United States Department of Education.
- Institute for the Advancement of Research in Education. (2004). *Superkids research report*. Unpublished manuscript.
- Konstantopoulos, S., & Hedges, L.V. (2008). How Large an effect can we expect from school reforms? *Teachers College Record*, 110, 1613-1640.
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: National Institute of Child Health and Human Development.
- Pikulski, S. (1994). Preventing reading failure: A review of five effective programs. *The Reading Teacher*, 48, 30-41.

Pressley, M., Wharton-McDonald, R., Allington, R., Block, C., Morrow, L., Tracey, D., Baker, K, Brooks, G., Cronin, J., Nelson, E., & Woo, D. (2001). A study of effective first-grade literacy instruction. *Scientific Studies of Reading, 5*, 35-58.

Rowland Reading Foundation. (2007). *Theoretical foundations of the Superkids program*. Madison, WI: Rowland Reading Foundation.

Rowland Reading Foundation. (2006). *A quasi-experimental study of the Superkids program*. Madison, WI: Rowland Reading Foundation.

Snow, C. E., Burns, S. M., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academies Press.

Spiegel, A. (1995). A comparison of traditional remedial programs and reading recovery: Guidelines for success for all programs. *The Reading Teacher, 49*, 86-98.

U.S. Department of Education. (2008). *Conditions of Education 2008*. Washington, DC: United States Department of Education, Institute of Educational Sciences.

Table 1

Initial and Final Sample Characteristics

Variable	Sample					
	Total (829)	Initial		Final		Total (750)
		Control (403)	Treatment (426)	Control (368)	Treatment (382)	
<i>Ethnicity</i>						
African-American	13%	12%	14%	11%	13%	12%
Asian-American	<01%	<01%	<01%	01%	<01%	<01%
Hispanic	24%	24%	23%	24%	24%	24%
Native-American	01%	02%	<01%	02%	<01%	01%
Pacific-Islander	<01%	<01%	<01%	<01%	<01%	<01%
White	63%	63%	64%	63%	64%	64%
Minority	38%	38%	38%	37%	38%	37%
<i>Free/Reduced Lunch</i>	45%	47%	43%	47%	43%	45%
<i>English-Language Learner</i>	04%	05%	04%	05%	03%	04%
<i>Public School</i>	73%	72%	73%	72%	73%	73%
<i>Mean (SD)NCE</i>	44.9	43.8	45.9	43.9	46.0	45.0
<i>Pretest Word Analysis</i>	(17.7)	(17.3)	(18.1)	(17.1)	(18.1)	(17.6)

Note. Sample sizes are in parentheses. Minority status included African-American, Hispanic/Latino, Native-American, and Pacific Islander students.

Table 2

Teacher and Classroom Characteristics

Variable	Total Sample (<i>n</i> =43)	Control Classrooms/Teachers (<i>n</i> =22)	Treatment Classrooms/Teachers (<i>n</i> =21)
<i>Percent Teachers with Master's Degrees</i>	47	46	48
<i>Percent Teachers Reading Certified</i>	14	05	24
<i>Mean Years Teaching</i>	14.6 (10.4)	14.0 (10.7)	15.2 (10.3)
<i>Mean Percent Minority Students</i>	35.9 (31.4)	35.6 (30.8)	36.1 (32.7)
<i>Mean Percent Free/Reduced Lunch Students</i>	47.2 (25.0)	49.2 (26.8)	45.1 (23.5)
<i>Mean Percent English-Language Learner Students</i>	4.0 (7.7)	4.2 (8.3)	3.9 (7.2)
<i>Mean NCE Pretest Word Analysis</i>	44.4 (7.5)	43.4 (7.7)	45.3 (7.4)
<i>Mean Class Size</i>	17.4 (6.5)	16.7 (3.5)	18.2 (8.6)

Note. Standard deviations in parentheses.

Table 3

Test Score Summary Statistics

Outcome	All Students	Treatment	Comparison
<i>Fall Word Analysis</i>	45.0 (17.64) 41 st (750)	46.0 (18.1) 42 nd (382)	43.9 (17.1) 39 th (368)
<i>Spring Vocabulary</i>	60.0 (18.11) 68 th (747)	60.3 (19.2) 69 th (381)	59.7 (17.0) 68 th (366)
<i>Spring Reading Comprehension</i>	65.7 (21.6) 77 th (740)	69.0 (20.0) 82 nd (378)	62.3 (22.6) 72 nd (362)
<i>Spring Total Reading</i>	64.9 (21.5) 76 th (735)	68.3 (20.5) 81 st (377)	61.4 (22.0) 71 st (358)
<i>Spring Word Analysis</i>	60.5 (19.1) 69 th (745)	64.7 (21.1) 76 th (381)	56.1 (15.8) 61 st (364)
<i>Spring Reading Words</i>	68.0 (20.3) 80 th (738)	70.8 (20.3) 84 th (377)	65.1 (20.1) 76 th (361)
<i>Spring Language</i>	71.5 (18.7) 85 th (745)	72.3 (19.2) 86 th (381)	70.7 (18.1) 84 th (364)

Note. The first cell entry is the mean normal curve equivalent (NCE) score. The standard deviation of the NCE mean is in parentheses on the first row of each cell. The percentile rank of the NCE mean score is the first value on the second row of each cell, and the sample size is the second value on the second row in parentheses.

Table 4

Student Level Variable Correlations

	1	2	3	4	5	6	7	8	9	10	11
1. Superkids	--	-.04	.00	-.03	.06	.03	.15**	.23**	.14**	.16**	.09*
2. Free/Reduced		--	.25**	.12**	-.17**	-.21**	-.09*	-.20**	-.22**	-.18**	-.27**
3. Minority			--	.22**	-.04	-.20**	.02	-.15**	-.02	.00	-.23**
4. ELL				--	-.07	-.15**	-.07	-.10**	-.02	.00	-.23**
5. Fall Pretest Word Analysis					--	.38**	.49**	.49**	.56**	.60**	.49**
6. Vocabulary						--	.28**	.49**	.36**	.36**	.57**
7. Reading Comprehension							--	.48**	.62**	.84**	.41**
8. Word Analysis								--	.63**	.63**	.60**
9. Reading Words									--	.92**	.48**
10. Total Reading										--	.51**
11. Language											--

Note. Superkids was coded 0 for comparison classroom and 1 for Superkids treatment classroom. Free-Reduced was coded 0 for “not eligible” and 1 for “eligible.” Minority was coded 0 for non-minority status (Asian or White) and 1 for minority status (African-American, Hispanic, Native American, or Pacific Islander). ELL was coded 0 for “not an English language learner” and 1 for “English language learner.”

** $p < .01$

* $p < .05$

Table 5

Classroom Level Variable Correlations

	1	2	3	4	5
1. Superkids	--	-.08	.01	-.02	.13
2. Percent Free/Reduced		--	.31*	.16	-.39**
3. Percent Minority			--	.16	.20
4. Percent ELL				--	-.12
5. Average Word Analysis Pretest					--

Note. ** $p < .01$, * $p < .05$

Table 6

Final Conditional HLM 2-Level Model, Literacy Outcomes

<i>Fixed Effect</i>	<i>Vocabulary</i>	<i>Reading Comprehension</i>	<i>Word Analysis</i>	<i>Reading Words</i>	<i>Total Reading</i>	<i>Language</i>
Model for mean Literacy Outcome, β_0						
Intercept, γ_{00}	-0.01(0.08)	-0.11(0.09)	-0.19(0.07)**	-0.10(0.07)	-0.12(0.07)	-0.07(0.06)
Superkids, γ_{01}	0.02(0.13)	0.24(0.10)*	0.41(0.14)**	0.23(0.11)*	0.27(0.11)*	0.16(0.11)
Percent Free/Reduced, γ_{02}	0.44(0.28)	0.34(0.32)	0.13(0.30)	-0.36(0.23)	-0.04(0.27)	0.02(0.24)
Percent Minority, γ_{03}	-0.56(0.20)**	0.44(0.25)	-0.61(0.23)*	0.15(0.18)	0.39(0.21)	-0.60(0.17)
Percent ELL, γ_{04}	0.11(0.85)	0.15(0.66)	0.88(0.95)*	1.54(0.74)*	0.93(0.64)	0.55(0.63)
Average Word Analysis Pretest, γ_{05}	0.04(0.02)*	0.01(0.02)	0.02(0.02)	0.00(0.01)	0.01(0.01)	0.01(0.01)
Model for Free/Reduced Slope, β_1						
Intercept, γ_{10}	-0.23(0.07)**	-0.10(0.07)	-0.17(0.06)	-0.22(0.06)**	-0.19(0.06)**	-0.28(0.06)**
Model for Minority Slope, β_2						
Intercept, γ_{20}	-0.17(0.08)*	-0.11(0.08)	0.00(0.06)**	0.03(0.08)	-0.08(0.09)	-0.12(0.04)**
Model for ELL Slope, β_3						
Intercept, γ_{30}	-0.38(0.11)**	-0.17(0.14)	-0.26(0.12)*	-0.22(0.10)	-0.11(0.08)	-0.22(0.08)**
Model for Word Analysis Pretest Slope, β_4						
Intercept, γ_{40}	0.30(0.04)**	0.45(0.04)**	0.45(0.04)**	0.52(0.03)**	0.56(0.03)**	0.44(0.04)**

Note. Vocabulary and Word Analysis pretest scores were standardized. Superkids was coded 0 for comparison classroom and 1 for Superkids treatment classroom. Free-Reduced was coded 0 for “not eligible” and 1 for “eligible.” Minority was coded 0 for non-minority status (Asian or White) and 1 for minority status (African-American, Hispanic, Native American, or Pacific Islander). ELL was coded 0 for “not an English language learner” and 1 for “English language learner.” Because test scores were standardized with mean = 0 and standard deviation = 1, the Superkids coefficient represents the standardized mean difference effect size.

** $p < .01$, * $p < .05$.

Table 7

Final HLM Model Summary Information by Outcome

	<u>Outcome</u>					
	Vocabulary	Reading Comprehension	Word Analysis	Reading Words	Total Reading	Language
<i>Average Classroom Mean Reliability</i>	.80	.79	.87	.73	.77	.77
<i>Proportion of Variance Between Classrooms</i>	.25	.21	.33	.18	.20	.21
<i>Proportion of Variance Between Classrooms with Student-Level Covariates (Percent Reduction of Between Classroom Variance Due to Student- Level Covariates)</i>	.18 (30)	.17 (19)	.26 (21)	.11 (38)	.13 (35)	.14 (38)
<i>Proportion of Variance Between Classrooms with Superkids Variable Included at Classroom Level (Percent Reduction of Between Classroom Variance Due to Superkids)</i>	.18 (00)	.16 (07)	.22 (16)	.10 (10)	.12 (12)	.13 (02)
<i>Proportion of Variance Between Classrooms with Superkids Variable & Covariates Included at Classroom Level (Percent Reduction of Between Classroom Variance Due to Superkids & Covariates)</i>	.16 (07)	.15 (15)	.21 (18)	.09 (16)	.11 (18)	.12 (11)
<i>Superkids Effect Size (Hedges g)</i>	.01	.24	.41	.23	.27	.16
<i>Average Superkids Classroom as a Percentile Score in the Student-Level Distribution</i>	50th%	59th%	66th%	59th%	61st%	56th%

Table 8

Teacher Outcomes by Control (n=21) and Treatment (n=21) Groups

Outcome	Group	M (SD)	t	d	Percentile
On average, how much time per day do you spend teaching language arts, including reading and writing?	Treatment	99.1(25.1)	-2.62*	.81	79 th %
	Control	141.2(69.4)			
How much of this time is spent using the Superkids or core program?	Treatment	87.8(15.5)	2.22*	.69	75 th %
	Control	67.1(39.6)			
Proportion of amount of language arts instruction spent using core program	Treatment	0.91(0.14)	6.88**	1.46	93 rd %
	Control	0.49(0.24)			
Satisfaction with current core reading program	Treatment	4.86(0.36)	3.41**	1.05	85 th %
	Control	3.95(1.16)			
Effectiveness of current core reading program at teaching students how to read	Treatment	4.76(0.44)	2.94**	.91	82 nd %
	Control	3.90(1.26)			
Effectiveness of current core reading program at motivating students to learn	Treatment	4.90(0.30)	4.28**	1.32	90 th %
	Control	3.95(0.97)			
Ease of use of current core reading program	Treatment	4.86(0.36)	2.70**	.83	80 th %
	Control	4.29(0.90)			

Note. *p<.05, **p<.01, "d" indicates the standardized mean difference effect. The "Percentile" column represents the percentile score of the average Superkids teacher in the control group population distribution.