Using Self Selected Reading Materials to Improve the Achievement of Children Poorly Served by Traditional Instruction: African American and Early Independent Readers

Robert M. Offenberg
Independent Evaluation Consultant and
St. Joseph's University Adjunct Professor
Philadelphia, Pennsylvania

Mailing address: Robert M. Offenberg, Ed.D
137-B North 22 Street
Philadelphia PA 19103.
E-mail: roffenb@aol.com
Phone and fax: (215)-972-6792

Word Count: 8,491

Abstract:

This ex post facto study is an analysis of the effect of *100 Book Challenge,* a reading program component emphasizing extensive independent reading of self-selected materials, on the reading achievement scores of Philadelphia pupils in grades 1 through 3 during the 2002 to 2004 school years. The study involved 89 different elementary schools—14 *100 Book Challenge* or Treatment schools and 75 Comparison schools. A preliminary study replicated previous findings by showing that adding the Treatment to the ongoing reading programs of schools significantly improved 2004 TerraNova reading score averages. The main study showed that the preliminary findings were due to the test score gains of two groups of pupils with needs often not satisfied by traditional early elementary grade reading programs: African American children; and children who, according to their 1st grade, first marking period report card marks, were early, independent readers. Together these groups constituted 79% of the nearly 16,000 children who were subjects of the study. The observed effect sizes were large enough to show that *100 Book Challenge* made meaningful improvements to the test scores of these groups of students, and that it should be considered when designing reading programs for schools where they are prevalent.

Key Words: Reading
Self-selected materials
African American
Early Independent Readers
Hierarchical Design

Using Self Selected Reading Materials to Improve the Achievement of Children Poorly Served by Traditional Instruction: African American and Early Independent Readers

The goal of this ex post facto inquiry was to determine whether urban schools' making a multi-year, multi-grade commitment to incorporate the *100 Book Challenge* system, which emphasizes children's independent reading of self-selected books, into their reading programs led to higher standardized reading test scores of children in grades 1 through 3. If incorporating the system appeared to lead to higher test scores, the next goal was to find out whether the effect was a general one, or whether subgroups of pupils were its principal beneficiaries, and then to measure the magnitude of the effects. The findings in this study will hopefully help schools and school districts make suitable decisions about whether *100 Book Challenge* or other similar approaches should be part of their reading instruction, and make good decisions about which schools and students would benefit the most from using it. As the system is an exemplar of a multiple-path, curriculum-embedded, independent reading approach, the findings may also contribute to a general understanding of the effects of other programs using similar methods.

The following description of the system is based on material provided by its company's Chief Academic Officer. More information about the program can be found at the company's website: http://www.americanreading.com.

100 Book Challenge, developed and owned by the American Reading Company, is designed to enhance pupils' reading competencies through use of self-selected reading materials that are not part of traditional, formal instruction. Like other independent reading efforts, it is typically added to programs that have a reading-skills focus and complements them. It includes on-going, embedded teacher training, parent involvement, and student-centered components. The program gives pupils access to a large number and a wide variety of books that students are expected to find attractive, chosen from the titles of over 200

publishers. A standards-based, color coded leveling system is used to classify books based on difficulty according to national standards for acquisition of phonemic awareness, phonics, fluency, comprehension and vocabulary skills. Participating teachers help students to learn to self-select books for daily in-school and at-home independent reading from those levels at which the students are most fluent. The books are stored in libraries (each containing between 200 and 300 titles) that are rotated among the classrooms and the grades of a school. When teachers have evidence that a pupil has mastered one reading level they move the pupil to the next. These activities are supported by teacher training, parent training and commitments to manage children's reading at home, record-keeping components, and a system of rewards recognizing students whose *quantity* of reading meets program standards. An important goal of the program is to provide differentiated instruction to all students facilitated by use of the proprietary, standards-based book leveling system and regular one-on-one teacher-student reading conferences. The company's website gives more information about how the program is conducted.

100 Book Challenge is an independent-reading, curriculum-embedded program that recognizes the value of using pupils' reading preferences. In this type of program, children are expected to learn to read, in part, by engaging them with materials they have chosen for pleasure or for obtaining information related to their interests (Cullinan, 2000; Guthrie, 2004). As such, it is an important complement to the skills-oriented and phonics-based materials that are currently dominating the field.

The key components of *100 Book Challenge*, encouragement for children to read more in order to read better and the provision of books matched to students' personal levels of reading skill, have a long history. For example, as early as the 1930's, educators were advised to provide students with books they could read easily and were helped to understand that a child learns to read by reading often (Betts, 1936).

Since the 1930's, the practice of encouraging independent reading to improve reading achievement has moved in and out of favor, and been the

subject of intense debate. Two camps have coalesced; the skills-based advocates, who tend to emphasize the mechanics of reading, especially phonics, and the meaning-based advocates who tend to favor holistic approaches emphasizing meaning-construction.

Reading instruction in the 1940's and 50's was dominated by the whole-word, or look-say method, a holistic approach in which students were expected to experience and then remember new words as whole units when they encountered them in text. As students accumulated more and more of these sight words, it was assumed their reading levels would advance. Near the end of this period a popular book, *Why Johnny Can't Read: And What You Can Do About It* (Flesch, 1955) described the poor reading abilities of American students and called for reading instruction that emphasized letter-sound associations or phonics, a skills-based approach. Chall (1967) furthered the argument, by describing two sides of a 'Reading War', and coming down in favor of phonics instruction for most young readers, and phonics became the dominant approach for teaching beginning readers for about a decade.

From the late 1970's to the 1980's meaning-based instruction was in fashion. Several prominent educational leaders pushed for increased emphasis on volume of reading experience as the key element in the creation of young, successful readers (Allington, 1977; Goodman, 1971; Smith, 1982; Greaney and Hegarty, 1987). Their ideas fueled the whole language movement with its focus on processes and authentic reading, and de-emphasis of discrete skills. Although some educators eschewed phonics instruction entirely, most in this camp considered themselves advocates of a more eclectic approach that combined contextual elements, especially engaging students with large amounts of reading materials they were expected to find interesting, with phonics elements when students needed them.

This group was followed by skills-based proponents (Foorman, Francis, Novey and Liberman, 1991; Griffith, Klesius and Kromey, 1992; Kame'enui, Stein, Carnine and Maggs 1981; Moats, 1994; Torgesen, Morgan, and Davis, 1992; Torgesen, and Rashotte, 1997; Wagner and Torgesen, 1987) who

currently dominate thinking in the field. They revived the discrete-skills focus, with research on skills they thought to be prerequisites to reading, especially phonemic awareness and phonics. The composition of the National Reading Panel, a federally sponsored advisory group, was weighted toward these researchers, resulting in its report's declaration of five reading pillars, phonemic awareness, phonics, fluency, vocabulary, comprehension, as primary foci of instruction, and its discrediting of independent reading as an instructional strategy (National Reading Panel, 2000). The NRP advised: *it would be unreasonable to conclude that research shows that encouraging reading has a beneficial effect on reading achievement* (p. 3-28). The NRP report became the basis for the No Child Left Behind legislation and the reading instruction of elementary school children moved strongly toward discrete skills teaching, especially the systematic teaching of phonics to young children.

One of the key points, perhaps one of the mistakes of the National Reading Panel in making this decision, was to regard correlation based studies as non-scientific, and therefore unworthy of consideration in determining effective approaches to teaching children to read. As a result, they undervalued the studies that demonstrated the efficacy of encouraging children to read more, most of which were correlation-based and not traditional studies with clearly defined experimental and control groups, preferably created through randomization. So the NRP concluded that although *there is an extensive amount of correlational data linking amount of reading with reading achievement (Cunningham & Stanovich, 1998; Krashen, 1993), such studies do not provide a clear delineation of what is antecedent and what is consequent* (p. 3-10). They felt that it was unclear whether readers got better by reading more, or whether better readers chose to read more because they were good at it. From their initial literature search of over 30,000 published articles as possibly germane to deciding this question, the NRP found only 14 studies that met reasonable review criteria. Further, they decided that the studies found did not satisfy their criteria for meta-analysis. They performed an informal analysis instead which resulted in their de-emphasizing of independent reading.

However, the ignoring of the large body of correlational data supporting independent reading by the NRP was questioned by a number of researchers (Yatvin, 2000; Allington, 2002; Garan, 2002; Krashen, 2001) some of whom noted that some of the most accepted scientific principles have been established almost solely on the basis of correlational research (e.g., that smoking causes cancer).

In making their decision, NRP clearly gave priority to what Campbell and Stanley (1966) called 'internal validity', assurance that the cause-effect relationships of the experimental design are unambiguous; at the expense of what Campbell and Stanley called 'external validity' or assurance that the findings can be generalized to the population of interest, which is often supported by consistent findings over a larger number of studies using different strategies. All of the *100 Book Challenge* evaluations have been conducted in urban neighborhood elementary schools of a district, not laboratories; in other words, at places and under conditions that will be most helpful to people who will be considering acquiring the program or using its approaches. Thus they tend to have high external validity for this specific purpose, even when they do not meet all the internal validity standards of NRP. As we will show, these evaluations have moved in the direction of increased internal validity, with the latest evaluation clearly meeting the NRP's acceptable quasi-experimental design standard (see National Reading Panel (2000) page 1-7). The present study provides evidence that that *100 Book Challenge* and probably other well developed, independent reading programs are effective additions to the skills-oriented programs that have been dominant.

Recent evaluations of Reading First, the best funded part of NCLB that addressed reading in grades K-3, have generally shown that the emphasis on discrete skills has been a failure. The Reading First Impact Study Final Report stated: *Reading First did not produce a statistically significant impact on student reading comprehension test scores in grades one, two or three* (Ganse, et al, 2008). The consideration of a reading initiative that combines skills with meaning emphasis would appear timely.

Prior to the current research effort, the William Penn Foundation supported two evaluation studies of *100 Book Challenge*, DuCette (1999) and DuCette (2001). Both were quasi-experimental designs conducted in Philadelphia public schools that concluded the program contributed meaningfully to the reading achievement of students. DuCette (1999) studied 12 elementary schools in each of which at least 10 teachers of first through third grade classes had agreed to participate in the program. Stanford Achievement Tests (SAT-9) NCE scores and performance level scores were the outcome measures. Classes at the same schools taught by teachers who chose not to participate in the program comprised the comparison group. DuCette found that students who were in program classes attained higher reading levels than students who were in non-program classes. Within the program, students in classes with fuller implementation attained higher scores than students enrolled in classes where it was less fully implemented. Since the classes were self-selected and no school, student or teacher information was used in the analyses, there was no way to disentangle the effects of *100 Book Challenge* from other likely differences between program and comparison classes.

The second study addressed some of the shortcomings of the first. In DuCette (2001) the units of analysis were grades and classrooms within schools. Ten program schools were matched with twenty comparison schools on the previous year's Philadelphia School Performance Index (a local school effectiveness measure heavily weighted with Stanford Achievement Test-9 performance levels), racial composition, poverty level, and grade range. SAT-9 reading-subtest NCE and performance level scores again were the outcomes studied. DuCette concluded that first grade program students were reading at higher levels than expected; but unfortunately, no first grade comparison group students were tested. In second grade, reading test performance was significantly higher in program schools than in comparison group schools. In third grade, the average program group reading test score was higher than the comparison group average, but the difference was not significant. DuCette's

previous finding that, within the program fuller implementation in classes led to higher test scores was replicated. In addition, students who had been in the program for two consecutive years out scored students who had been in the program for only one, an indication of a possible cumulative program effect. A survey of principals showed they thought the program was valuable.

The current study was conducted when No Child Left Behind, which began in 2002, and its test-based, Adequate Yearly Progress evaluations of student achievement were strongly influencing the policies of the Philadelphia public schools, which also had developed their own rating system. Moreover, the state of Pennsylvania was threatening to intervene in the management of schools by replacing the locally appointed board of education with the School Reform Commission, a governing body with both state and locally appointed members. During this period, the sub-regions comprising the district had chosen reading instructional materials, and local schools had received funding to buy supplementary programs that they considered appropriate for the students. The focus of this study is, therefore, on differences in pupil reading achievement resulting when Philadelphia schools contracted to make *100 Book Challenge* a component of the reading programs of all their 1st through 3rd grade classes. The study compared all schools in the district that had made this contract during the years the subjects were in these grades, the Treatment group, to all of the district's schools serving the grade range that had made no contract with this vendor, the Comparison group. Only special admission schools, schools that had made contracts to use *100 Book Challenge* in some but not all classes, and a few schools using non-traditional first grade report cards that could not be analyzed were excluded. As the presence of a contract was the factor that distinguished Treatment schools from Comparison schools, the level of *100 Book Challenge*, implementation levels could vary among the Treatment schools. Moreover, both the Treatment and Comparison schools were making efforts to meet federal standards. For these reasons, and especially because of the multivariate design used in the main part of the study, the findings are a conservative answer, with high external validity, to the management question,

'How will purchasing *100 Book Challenge* for schools in a large urban system whose main reading programs tend to reflect the current skills-oriented approach affect the reading achievement of students?'

Vernez, Karam, Mariano, and DeMartin (2006), conducted a study of validated comprehensive school reform models that had become parts of on-going school operations when they studied them (i.e. were implemented to achieve educational rather than research goals). They concluded that schools and districts adopting the 'validated' reforms often realized little or no improvement in student achievement because of relatively low levels of program implementation. These findings suggest that the ex post facto design of this study has more external validity than most traditional program evaluation designs. Unlike most traditional program validation studies that produce findings under ideal conditions, the results of this study *could* be affected by the implementation variations likely to be found among schools in large urban districts purchasing *100 Book Challenge* or developing their own reading instruction plans. Moreover, the Comparison group was composed of schools being pressed to provide what they considered to be the best program possible, not the artificial, 'control-group' condition of formal studies.

As the Treatment group of this study was defined by the presence of a *100 Book Challenge* contract, and the company's Chief Academic Officer reported that they showed variations in levels of implementation and provision of professional development and support for teachers like those observed by Vernez et al*,* the findings were obtained despite this variation*.* All schools received the initial teacher and parent training components, materials and a rubric designed for self-assessment of the level of program implementation. However, the levels of additional professional development support depended on school-based decisions regarding investments of time and funds that could change from year to year. Thus, the results of this study are more likely to be useful for making urban school district policy than are findings from traditional program validation research.

Study Design

In the preliminary phase of the study, a simple analysis of variance design was used to examine whether schools that made a contractual commitment to *100 Book Challenge*, the Treatment group, outperformed others. In addition, the inconsistent grade level results of the second DuCette study were revisited, by examining whether schools' having made the commitment benefited children in all three grade levels of the study. Then, recognizing that non-programmatic factors could explain the preliminary findings and that managers must make rational, school-by-school choices from among competing reading programs, the main study used a more rigorous, hierarchical, multivariate design to statistically control pre-existing differences among schools and students, measure broad program effects and then to identify the types of pupils whose reading achievement was most affected by having had *100 Book Challenge* as a component of their school's reading program. Thus, the results of the main study can provide guidance to educational leaders deciding whether to choose *100 Book Challenge* or to develop similar approaches when planning reading programs for their schools.

Variables:

Dependent. The dependent variable of the study was the Normal Curve Equivalent (NCE) Reading score of TerraNova tests (CTB McGraw Hill, 2002) administered by the School District to its entire first through third grade pupil population in April 2004. NCE scores were used because they comprise an equal interval scale and, in the population used to standardize the TerraNova tests, they are normally distributed, with a mean of 50.00 and a standard deviation of 21.06 at each grade level. These characteristics make NCEs the easiest form of standardized test scores for describing the impact of instructional programs that span several grade levels, as this one does; and, as Goldschmidt, Choi and Martinez (2004) have shown, evidence about the effects of programs on schools

and students is the same as that yielded by other, more complicated forms of test scores.

Independent. All the independent variables but one were extracted from databases maintained by the School District of Philadelphia and are official student records. The independent variables are:

- *Grade in 2004*: the pupil's grade level, in years above first grade, when he/she took the TerraNova reading test.
- *Fall 1st Grade Reading Level*: The independent reading level mark reported on the pupil's earliest first grade report card. These marks were based on well-specified, district-wide assessment procedures and criteria. They were converted to grade-equivalent (GE) levels using a scale provided by the school district, which assigns a value of '0' to non-readers.
- *Ever Repeated a Grade*: Inferred from the grade assignments of pupils.
- *In Special Education:* Inferred from the presence of a special education indicator in pupil's 2004 record. Speech impaired and gifted pupils were not regarded as special education.
- *African-American*: Inferred from race codes appearing in school district records.
- *Latino*: Inferred from race codes appearing in school district records.
- *Prevalence of Title 1 Eligible Students at School:* The proportion of pupils in grades 1-3 of a school who were low income according to state and federal standards.
- *Treatment and Comparison Schools:* All neighborhood elementary schools in the Philadelphia district that, according to American Reading Company records, had purchased *100 Book Challenge* for all first grade classes in the 2001-2002 school year, first and second grade classes in 2002-2003, and first, second and third

grade classes in 2003-04 were considered part of the Treatment group. Records show that all of these schools made major investments in the purchase of materials and staff training, but day-to-day classroom management of the program was not monitored. The Comparison group was comprised of all the neighborhood elementary schools in the district that had made no *100 Book Challenge* purchases for any of these grades and years. It is extremely unlikely that any of the Comparison group schools followed the procedures or had access to the materials outlined above or described on the company website, However they were subject to the same funding rules as the Treatment schools and may have purchased other programs or services.

The first six independent variables are attributes of the individual students in the study. The last two independent variables are attributes of the schools they attended.

In the main study a Hierarchical Linear Modeling or HLM (see below) is used to assess the value of adding *100 Book Challenge* to the reading programs of schools. HLM analyses results are often easier to interpret when independent variables have been scaled to have meaningful 'zero' values. With 'No' coded as zero for binary variables, almost all of the independent variables of this study, including whether a school was or was not a *100 Book Challenge* school, had this property. However, three variables, *Prevalence of Title 1 Eligible Students at School, Fall 1st Grade Reading Level* and *Grade in 2004*, were centered to create scales with meaningful zeros. After centering, the grand means of these variables were zero, and the values of individual cases represented deviations from the grand mean. With the centering, the intercept of the analysis was an estimate of the NCE score of a hypothetical, but meaningful 'baseline' student—one attending a Comparison school where the prevalence of Title 1 students was at the mean of the study, who was enrolled in the mean grade level of students in the study, and had had the mean early first grade reading level of students in the study. This Baseline student also had never repeated a grade, was not in special

education, and was neither African American nor Latino. All the Main Study findings showed how the expected NCE score of a student was influenced by type of school—Treatment or Comparison and by each personal attribute (Grade in 2004, Fall 1st Grade Reading Level, African-American, etc.). Among them were personal-attribute by school-type interactions that described how students were affected by attending a *100 Book Challenge* school.

Sample:

Schools. The study sample comprised most students who attended grades one through grade three of 89 different Philadelphia neighborhood public schools in 2004—fourteen *100 Book Challenge* and seventy-five Comparison schools. Contract records of the American Reading Company and a data-base provided by the School District of Philadelphia were used to identify the schools of the sample.

Although these were two self-selected groups of schools, the communities they served were similar. The percentage of Title 1 eligible students, which reflects community social class factors typically predictive of children's standardized reading test scores, ranged from 3% to 100% in each group of schools, and the mean Title 1 percentages of 61% in the Comparison schools and 72% in the *100 Book Challenge* schools did not differ significantly ($t(87)=.914$, $p=.36$). Still, to guard against contamination of the findings by community social class of schools, the proportion of Title 1 students in schools was controlled in the Main Study.

Students. There were 17,130 potential students in the study, 15,972 with complete data. All attended one of the study-sample schools continuously from when they began first grade until they were tested in spring 2004. The students at the *100 Book Challenge* Treatment group schools comprised 16.0% of the sample, those at the Comparison schools comprised 84.0% of the sample. Since continuous attendance at the same school since beginning first grade was a study requirement, students whose records showed they had moved from one school to another within the school district or entered the district after they began

first grade were not part of the study. Because of these requirements there were more first than third grade students in the sample—41% were in first grade, 31% in second, 27% in third when they were tested.

Data Analyses:

Preliminary Study. An ordinary, two-way analysis of variance was computed to answer the preliminary, basic research question, 'Did students in the*Treatment* schools outperform students who attended the same grades of Comparison schools on the TerraNova reading tests?' The researcher of this study recognized that there was an audience for the answer; even though there were risks that differences between the schools choosing to use the Treatment and the Comparison schools affected the results. That is why the main study was also conducted.

Main Study. After the preliminary analysis, a Hierarchical Linear Analysis or HLM (Raudenbush, Bryk, Cheong, and Congdon, 2000) was used to control the effects of school and student variables, and assess the degree that the preliminary findings were results of *100 Book Challenge*. A model was used that could measure both the degree that it had a general effect on the reading achievement of students and the degree it had specific effects answering the main research question, 'Which students benefited most from *100 Book Challenge*?'[1]

HLM is particularly useful in evaluations like this, where there are naturally occurring groups that form a hierarchy—in this case students were enrolled in schools, and their schools belonged to either the Treatment or the Comparison conditions. It has three major benefits: First, it recognizes that students attending the same school may have more in common with each other than they have with students in other schools, so outcomes occurring only at one or two schools are less likely to be attributed to all schools comprising a treatment condition. Second, it uses a family of student and school level 'control' variables to equate the backgrounds of students, thus partially fulfilling the role played by random assignment of students to treatment groups in traditional, laboratory-based

research studies. Third, it validly tests interactions between student characteristics and school variables so the types of students whose reading test scores were improved the most by the Treatment could be identified.

The design was a two level hierarchy. Level 1, students within schools, explored the relationships of the first six variables listed above—the student variables—to their TerraNova NCE scores, separately for each of the 89 schools. The Level 1 products, which would then be analyzed in Level 2, were regression analyses *for each school* comprised of weights relating its students' backgrounds to TerraNova reading scores, and a constant that was an estimate of the test score the baseline student described above would have earned if he or she attended the school.

In Level 2, the constants and the regression weights yielded by Level 1 were analyzed to find patterns typical of the Comparison schools and the Treatment schools, and then determined whether the patterns differed. Finding statistically significant differences between the Treatment school and the Comparison school patterns would be evidence that 100 Book Challenge changed some or all students' reading scores.

The arrangement of the Main Study results in Table 1 in the following section departs from traditional HLM-study practice, in which the findings derived from Level 2 are organized by their associated Level 1 variables. The author of this study felt that this approach would make the findings unnecessarily difficult for the educational policy makers, who are considered to be an important part of this study's audience. For this reason, the findings have been arranged so the effect of poverty on achievement which is associated with all schools appears first; those that describe students attending the Comparison schools follow, and the differences between students attending Treatment and Comparison schools, the results of greatest interest to policy makers, appear last. This allowed discussing all the effects of *100 Book Challenge* in a cogent way.

Results

The preliminary study showed that students attending Treatment schools had significantly higher reading test scores than Comparison school students' scores. The main study showed that the preliminary study finding was due to the elimination by the Treatment of most of the reading test score gap between African American and non-minority students, and to its improving the reading test scores of children whose first first-grade report cards showed they were early independent readers.

Preliminary Study: The goal of the preliminary study was to answer the basic question, 'Did students in the Treatment schools outperform students who attended the same grades of Comparison schools on the TerraNova reading tests?' The simple two-way analysis of variance that was computed to contrast the *100 Book Challenge* students with the Comparison students showed that, at every grade studied, the Treatment students outperformed the Comparison students. Figure 1 shows the mean NCE scores of the two groups of students in each grade.

<INSERT FIGURE 1 ABOUT HERE>

The analysis of variance upon which the figure is based confirmed that the 1.7 NCE overall difference between the two groups of students was statistically significant, $F (1/15,966) = 16.43$, $p < .001$. As shown by the figure, the average NCE scores of both the Treatment and the Comparison students were higher among first graders than among third graders, a change that was also highly significant, $F (2/15,966) = 19.89$, $p < .001$, but not relevant to this study.[2] Although the figure appears to show that the difference between Treatment and Comparison students changed with grade, these changes were not statistically significant, $F (2/15,966) = 1.97$, NS), suggesting that they could be due to chance.

While this analysis shows that Treatment school students out performed Comparison school students, it does not adequately control for preexisting school or student difference, it does not address the non-independence statistical issue

raised by the shared experiences of students who attended the same school and it does not allow distinguishing students who benefited most from *100 Book Challenge* from those who benefited least. The main study in the next section addresses these issues.

Main Study: The goal of the HLM-based main study was to answer the question, 'Which students or schools benefited most from the Treatment?' The findings are summarized by Table 1, which shows that *100 Book Challenge* nearly eliminated the test score gap between non-Minority and African American students found in Comparison schools, and raised the scores of students who had acquired some beginning reading skills by the first marking period of first grade, suggesting that it was meeting the unique needs of early independent readers.

<INSERT TABLE 1 ABOUT HERE>

Poverty Adjustment. The first section of the table, *Proportion of Schools' Students Who Were Title 1 Eligible* adjusts all of the subsequent findings for the prevalence of poverty in the schools of the study. The negative, significant weight of -6.955 NCEs shows that the average reading scores of students attending a school where poverty was more prevalent tended to be lower than those of otherwise similar students who attended a school where poverty was less prevalent. It neither changed the impact of the student history variables (Grade in 2004, Ever Repeated a Grade, In Special Education, African American, Latino, Fall 1st Grade Reading Level) nor the impact of students' attending a *100 Book Challenge* school on reading test scores.

Comparison Schools. The first data line of the next section of the table shows the TerraNova NCE score of the baseline student described in the discussion about centering of variables (an average, non-minority, regular-education student who attended an average Title 1-prevalence school) if the student was attending a Comparison school. Its value, 48.350, shows that the reading competence of this type of student was slightly below the national

average score of the test. The next lines of this section of the table show how six student background and history variables affected the test scores of Comparison school students, regardless of the Title 1-prevalence at their school: As students' grade levels increased from first to third, their scores declined by -2.108 NCEs per grade level, a significant result replicating the finding of the preliminary study. In the Comparison schools, students identified as Special Education, African American and Latino children had significantly lower test scores than the base-line student. The differences were -8.167 NCEs for students in Special Education, -4.637 NCEs for African American students, and -3.057 NCEs for Latino students. However, students who had repeated grades did not differ significantly from their peers once the effects of the other variables were taken into account.

Comparison school students who had begun to develop early independent reading skills, as reflected by their Fall 1st Grade Reading Level report card marks, had significantly higher reading test scores even though the marks and the tests were separated by two-thirds to three and two-thirds school years depending upon a student's 2004 grade and his or her grade retention history. The analysis shows that test scores increased by 13.498 NCEs for each reading grade equivalent (GE) mark on the report card, a finding that demonstrates the importance of the reading skills children bring with them to school on later performance, and the validity of using a Fall 1st Grade Reading Level to measure them.

Treatment Schools. The first two sections of Table 1 both describe the relationships between the school community and student history in Comparison schools and control them statistically. As a result, the third section of the table, "Treatment Schools," shows the ways attending a *100 Book Challenge* school altered the relationships between student history variables and student scores that had been found at the Comparison schools. There were two statistically significant, large, and educationally meaningful results: the Treatment raised the scores of African American students and significantly enhanced the subsequent reading achievement of students who had been early readers in first grade, but

this paper will begin by discussing an equally important, but non-significant result.

As shown by the first data row of the third part of the table, the effect of *100 Book Challenge* on the baseline student was small and had a non-significant associated t-value. This showed that adding the Treatment neither enhanced nor reduced the reading test scores of this type of student. This meant that the significant result of the preliminary study was not likely to be due to the Treatment's yielding a general effect, but was probably due to its affecting the scores of identifiable student subgroups.

100 Book Challenge significantly reduced the gap between the scores of African American students and the baseline students in this study. At Comparison schools, the reading scores of African American students were 4.637 NCEs below the scores of their non-Minority peers; the statistically significant effect on African American students who were at Treatment schools showed that this gap was reduced by 3.303 NCEs. Thus, having *100 Book Challenge* as part of the reading program of a school reduced the gap between African American and non-Minority students to 1.334 NCEs, a 71% reduction. This is evidence that adding the Treatment to the reading program of a school is one way of overcoming factors that lower the reading achievement of urban, African American elementary grade students.

A second effect that *100 Book Challenge* produced is that it appears to have significantly enhanced the subsequent reading achievement of students who had been early independent readers in first grade. The Comparison school results were evidence that students whose Fall 1st Grade Reading Levels showed they were early readers had advantages that were maintained months and years later, when they were tested with TerraNova. Attending a *100 Book Challenge* school added to these advantages— 2.269 NCEs per Fall 1st Grade Reading Level GE were added to the 13.498 NCEs per Fall 1st Grade Reading Level found at the Comparison schools, a 20% increase.

As the difference between the two groups of schools was per *Fall 1st Grade Reading Level*, and every student had his or her own fall first grade

reading level, the result of attending a Treatment school on a pupil's NCE score varied from child to child. The more the child was an early independent reader the more he or she benefited from it. From the 'Effect in NCEs' values in the table, it is easy to compute the size of the effects. For example, the average Fall 1st Grade Reading Level of students was 0.9 GE; the Treatment increased the test scores of students who had been at that level by 2.4 NCEs. Children whose Fall 1st Grade Reading Levels were 1.6 GEs or higher composed the top 10% of students in the study; the Treatment increased the test scores of students at this level by 4.3 or more NCEs.

In the schools of this study, 79% of the students were either African American or had Fall 1st Grade Reading Level marks showing they were early independent readers. Thus the two statistically significant findings would encompass most of the students in the study. Moreover, 29% of the students were both African American and had Fall 1st Grade Reading Levels indicating that they were early independent readers. The results of this study suggest that the benefits of *100 Book Challenge* were additive, and that the scores of these students were likely to have been increased by both statistically significant trends.

Discussion

The developers of *100 Book Challenge* believe that the changes in dominant reading programs occurring over the last seventy years were because neither the skills-and-phonics approach nor the reading-for-meaning approach meets the needs of all young children. When one approach became the norm, groups of children whose needs were not being met fully became evident, leading to the eventual adoption of the other approach. They believed that this need not be the case, so in an era when skills and phonics approaches to teaching young children were dominant, they developed a way to add a carefully controlled program emphasizing children's reading for pleasure at their comfort level to the skills-based approaches that were prevalent in the main reading program of many schools. This study shows that in one large, urban school

district, elementary-grade students with particular needs benefited, and that the beneficiaries were concentrated in the African American community and among early independent readers.

As discussed earlier, the time of the study, the 2002-2004 school years, was one in which all schools in the nation, including the ones in Philadelphia that were subjects of this study, were under pressure to succeed. The passage of the No Child Left Behind legislation in 2002 demanded that all schools succeed with those subgroups of students traditionally underserved and low-achieving. This study indicates that *100 Book Challenge* is probably a useful tool for meeting the needs of one of these sub-groups. It also indicates potential usefulness for meeting the needs of a second group that was completely ignored by No Child Left Behind legislation despite their importance to our society.

The findings of the DuCette studies, that adding the *100 Book Challenge,* to reading programs improved the standardized reading test scores of elementary grade students, were supported by the preliminary study findings. However, the main study showed that these outcomes are largely the result of the Treatment improving the scores of two components of the student population, African American pupils and early independent readers, without significantly affecting the scores of other pupils. As these findings link the effectiveness of the system to student characteristics, they should be useful for helping educational leaders decide whether *100 Book Challenge* or another approach emphasizing independent reading of self-selected materials is an appropriate strategy for a particular school, especially now during the federally mandated quest to meet the Adequate Yearly Progress (AYP) standards of No Child Left Behind.

The author believes that similar aspects of this Treatment helped the African American and the early independent reader groups, but that the ways that they did so were different. The landmark Coleman (1966) study and a more recent research summarized by Rothstein (2004) show how social factors prevalent among urban, low-income African American families result in lower levels of school achievement. Rothstein summarizes extensive literature about the ways these differences, among them fewer children’s books in the home and

differences in the ways parents tend to use them with their children, contribute to the race-related achievement gap that is the focus of his research. As a result, recent papers focus on the importance of helping African American parents to become active participants in their children's education. For example, Frazier-Trotman (2001), Edwards (1992), Drummond and Stipek (2004), and Epps (1995) advocate parental participation in their children's education in ways that can enhance the parental authority, enable parents to view themselves as partners of educators, and help parents to view their children as being successful. The *100 Book Challenge* system directly deals with several of these factors. It provides reading materials for children that their homes may lack, and shows parents how to engage in child-centered reading activities. It provides leveled materials matched to children's personal interests and reading fluency so that students at lower reading levels can still feel successful before their parents, and helps parents know what is expected of their children. Thus, the effect of *100 Book Challenge* on the achievement of African American children may be due to its addressing some of the education-related family factors that contributed to the discrepancy between African American and non-Minority student reading scores found in the Comparison schools.

The second finding was that, attending a *100 Book Challenge* school increased the standardized reading test scores of students whose beginning of first grade report cards showed they were early independent readers; and the higher the student's early reading level, the more the student benefited from attending a Treatment school. No Child Left Behind, and the educational policies and literature that have grown out of it, tend to overlook this higher achieving group because of their emphases on remedying the deficiencies of students who are below average (see National Conference of State Legislatures*,* 2005 for criticisms of this aspect of federal policy). However, this study found that the number of early independent reading students in elementary schools was substantial and the Treatment was one way that their needs, which are more likely to be for enrichment than for remediation, can be met.

The*100 Book Challenge* system results in there being both class and homework time to address the instructional needs of these early independent readers better than traditional classroom instruction and homework can. While activities of the early readers typically parallel those of other students, the reading level of the materials the students use often does not. Thus, the more advanced students are helped to choose books that are more appropriately demanding and the instructional frameworks used in 100 Book Challenge, e.g., individualized teacher-student conferences, allow them to interact with interesting texts using higher order thinking, helping them develop higher levels of proficiency.

The results suggest that many of the students who came to school knowing how to read were able to reach difficulty levels that were well beyond those their classmates encountered. It may be that clarity of expectations in *100 Book Challenge* and the empowerment of students through choice helped create conditions that led to early independent readers taking fuller advantage of their inherent motivation to learn and the opportunity to use higher-level materials both at home and during school. The higher standardized test scores of the early independent readers who attended the Treatment schools and the finding that the gap with their peers at Comparison schools grew as students' early independent reading levels increased were the result.

Together, the African American and the early independent reader findings suggest that *100 Book Challenge* or other programs emphasizing reading self selected materials can help school managers to deal with a common dilemma, the need to simultaneously serve pupils whose cultural histories tend to de-emphasize reading at home and to serve students who come to school ready to read higher-level text than their peers.

The evidence of improved achievement among African American students—gains that have reduced by more than two thirds their achievement gap with non-Minority students found in Comparison schools—while also benefiting early independent readers suggests that the approaches used by this

program should be of particular interest to school managers working with urban schools serving diverse populations.

Acknowledgements:

This study was funded by American Reading Company, King of Prussia Pennsylvania, developers of *100 Book Challenge*. It is a component of the company sponsored program of research on the effectiveness of its products. Two members of the company, Alison Nathan and Gaeton Zorzi, contributed to this paper. A version was presented at the 2006 American Educational Research Association (AERA) Conference.

Biography:

Dr. Offenberg led research groups for the School District of Philadelphia for 35 years, most recently as its Senior Policy Researcher. In addition, he has served as an independent consultant, taught experimental design and statistics, and served as a member of dissertation committees. He is a member of the St. Joseph's University Inter-disciplinary Doctor of Education for Educational Leaders program faculty.

References:

Allington, R. L. (1977). If they don't read much how are they ever gonna get good?, *J. of Reading, 21,* 57-61.

Allington, R. L. (2002). *Big brother and the national reading curriculum: how ideology trumped evidence.* Portsmouth, NH: Heinmemann.

Anderson, R.C., P. Wilson, & L. Fielding. (1988). Growth in reading and how children spend their time outside of school. *Reading Research Quarterly, 23(3),* 285-303.

Betts, E. A. (1936*). The prevention and correction of reading difficulties.* Evanston, IL: Row, Peterson.

Calkins, L.M. (2000). *The art of teaching reading*. Upper Saddle River, NJ: Pearson Education.

Campbell, D. T. and J. C. Stanley (1963). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally.

Chall, J. (1967) *Learning to read: the great debate: an inquirery into the science, art and ideology of old and new methods of teaching children to read, 1910–1965.* NY, McGraw Hill.

Clay, M. M. (1979). *The early detection of reading difficulties*. Portsmouth, NH: Heinemann.

Coleman, J. (1966). *Equality of educational opportunity,* Washington, D.C.: United States Government Printing Office.

CTB McGraw Hill (2002). *TerraNova, the second edition: CAT technical bulletin 1*. Monterey, CA: CTB McGraw Hill.

Cullinan, B. E. (2000) Independent reading and school Achievement. *School Library Media Research, at http://www.ala.org/ala/aasl/aaslpubsandjournals/slmrb/slmrcontents/volume32000/independent.cfm .*

Cunningham, A. E. and K. E. Stanovich (1998). What reading does for the mind. *American Educator, 22*, 8-15.

DuCette, J. (1999). *An evaluation of the 100 Book Challenge program* (monograph), Philadelphia, PA, Temple University.

DuCette, J. (2001). *An evaluation of the 100 Book Challenge program in the schools funded by the William Penn Foundation* (monograph), Philadelphia, PA, Temple University.

Drummond, K. V. and D. Stipek (2004). Low income parents beliefs about their role in children's academic learning. *Elementary School Journal, 104 (3)*, pp 197-213.

Edwards, P. A. (1992). Involving parents in building reading for the African American learner. *Theory into Practice, 31 (4)*, 350-359.

Epps, E. G. (1995). Race, class and educational opportunity: Trends in the sociology of education. *Sociological Forum, Special Issue, African Americans and Sociology: A Critical Analysis, 10 (4)* pp 593-608.

Flesch, R. (1955) *Why Johnny can't read: and what you can do about It*. NY: Harper and Row.

Foorman, B., D. Francis, D. Novey, and D. Liberman (1991). How letter-sound instruction mediates progress in first-grade reading and spelling. *J. Educational Psychology, 83*, 456-469.

Frazier-Trotman, M. (2001). Involving the African American parent: Recommendations to increase the level of parental involvement within African American families. *Journal of Negro Education, 70 (4)*, 275-285.

Gamse, B.C., Jacob, R.T., Horst, M., Boulay, B., and Unlu, F. (2008). *Reading First impact study final report* (NCEE 2009-4038). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Garan, E. M. (2001). *Resisting reading mandates.* Portsmouth, NH: Heinemann.

Glasser, W. (1990). The quality school. *Phi Delta Kappan, 71(6)* pp 424-435.

Goldschmidt, P., K. Choi, and F. Martinez (2004). *Using hierarchical growth models to monitor school performance over time: comparing NCE to scaled score results*. Los Angeles, CA: CRESST/University of California, Los Angeles.

Goodman, K. S. (1971) Decoding—from code to what? *J. Reading, 14(7),*455-62.

Greaney, V and M. Hegarty (1987). Correlates of leisure time reading. *J. Research in Reading, 10 (February):* 3-20*.*

Griffith, P., J. Klesius, and J. Kromey (1992). The effect of phonemic awareness on the literacy development of first grade children in a traditional or whole language classroom. *J. Research in Childhood Education, 6,* 85-92

Guthrie, J.T., & Humenick, N.M. (2004). Motivating students to read: Evidence for classroom practices that increase motivation and achievement. In P. McCardle and V. Chhabra (Eds.), *The voice of evidence in reading research* (pp. 329-354). Baltimore: Paul Brookes.

Guthrie, J.T. (2002). Preparing students for high-stakes test taking in reading. In A. Farstrup & S.J. Samuels (Eds.). *What research has to say about reading instruction.* (pp. 370-391). Newark, DE: International Reading Association.

Kame'enui, E., M. Stein, D. Carnine, and A. Maggs (1981). Primary level word attack skills based on isolated word, discrimination list, and rule application training. *Reading Education, 6(2),* 46-55.

Krashen, S. D. (2001). More smoke and mirrors: a critique of the National Reading Panel report on fluency. *Phi Delta Kappan, 83,* 119-123.

Krashen, S. D. (1993). *The power of reading: Insights from the research*. Englewood, CO: Libraries Unlimited.

Moats, L. (1994), Knowledge of language, the missing foundation for teacher education. *Annals of Dyslexia, 44,* 81-102.

Moses, R. P. (1989). The algebra project: Organizing in the spirit of Ella, *Harvard Educational Review, 59(4)*, pp 423-443.

National Assessment of Educational Progress (2002), Institute of Education Sciences, US Department of Education. At http://nces.ed.gov/nationsreportcard

National Conference of State Legislatures (2005). *Taskforce on no child left behind final report,* Washington, DC: National Conference of State Legislatures.

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*; *Reports of the Subgroups.* At: http://www.nationalreadingpanel.org .

Raudenbush, S. A., A. Bryk, Y. F. Cheong, and R. Congdon (2000). *HLM 5 hierarchical linear and nonlinear modeling*, Lincolnwood, IL, Scientific Software International.

Rothstein, R. (2004). *Class and schools: using social, economic and educational reform to close the black-white achievement gap.* Washington, D.C., Economic Policy Institute.

Smith, F. (1982). *Understanding reading, 3rd edition.* NY, Holt Rinehart and Winston.

Torgesen, J., S. Morgan and C. Davis (1992). Effects of two types of phonological awareness on word learning in kindergarten children. *J. Educational Psychology, 84, 374-370.*

Torgesen, J., S. Morgan and C. Davis (1997). Prevention and remediation of severe reading disabilities: keeping the end in mind. *Scientific Studies of Reading, 1(3),* 217-234.

Vernez, G., R. Karam, L. T. Mariano, & C. DeMartini (2006). *Evaluating comprehensive school reform models at scale: Focus on implementation.* Santa Monica CA, RAND Corporation.

Vygotsky, L. S. (1990). *Mind in society: The development of higher psychological processe*s. Cambridge MA, Harvard University Press.

Wagner, R. and J. Torgesen (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101,* 192-212.

Wiggins, G. (1989). The futility of trying to teach everything of importance. *Educational Leadership, 47(3)*, pp 44-48, 57-59.

Yatvin, J. (2000). Minority View, in National Reading Panel, *Teaching Children to Read: an Evidence-Based Assessment of Scientific Research Literature on Reading.* Washington, DC*,* National Institute of Child Health and Development. At: www.nichd.nih.gov/publications/nrp/upload/minorityView.pdf

Footnotes:

[1] In fact, two models were computed. In one, the effects of all student background variables on test scores was allowed to be altered by the prevalence of Title 1 students at the school. In the other, the effect of Title 1 prevalence was limited to estimating the achievement of the average student of a school. As both analyses yielded similar estimates of the effects of the Treatment, the second, simpler one was used for this report.

[2] This study required students to attend one school from the beginning of first grade until April, 2004. This effect may have been the result of higher achieving students changing schools. It may reflect students in the school district losing ground when compared to the national sample used to norm the TerraNova. The study cannot distinguish between these two causes.

Figure 1: Mean NCE Scores of 100 Book Challenge & Comparison School Students

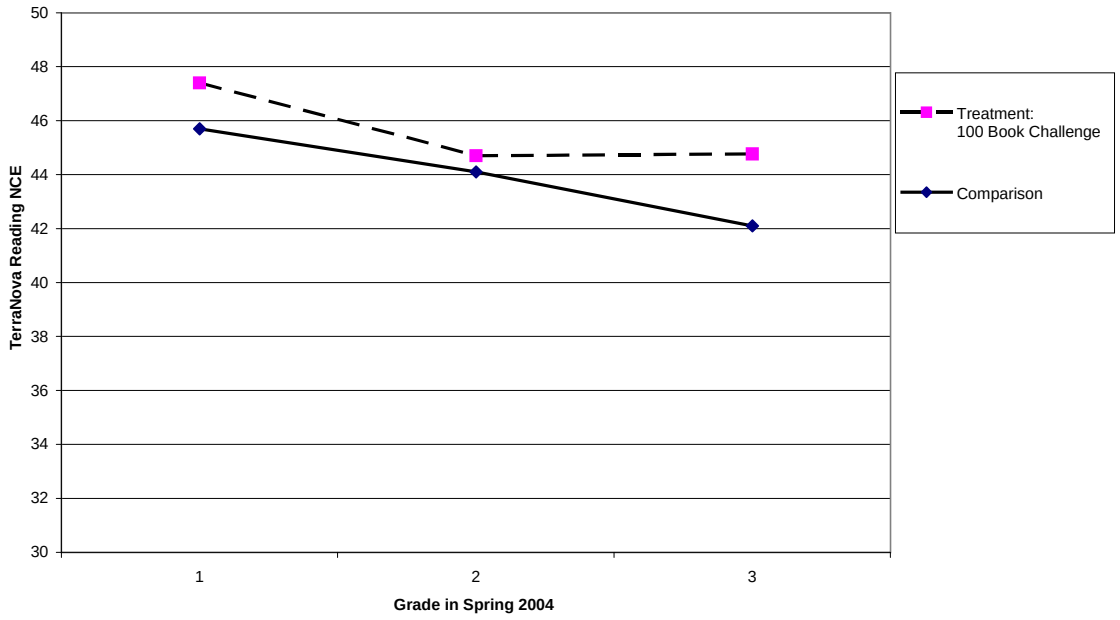

Table 1

Analysis of the Effects of *100 Book Challenge* on

TerraNova Normal Curve Equivalent Reading Scores

of Students in Grades 1-3

Variable		Effect in NCEs	t	Approx. df	p<
Prevalence of Title 1 Eligible Students at School		-6.955	5.216	85	.001
Comparison Schools:					
	The Baseline Student	48.350	71.814	85	.001
	Grade in 2004	-2.108	-12.123	15,957	.001
	Ever Repeated Grade	-0.293	-0.830	15,957	NS
	In Special Education	-8.167	-15.877	15,957	.001
	African American	-4.637	-10.601	15,957	.001
	Latino	-3.057	-5.60	15,957	.001
	Fall 1st Grade Reading Level	13.498	54.347	15,957	.001
Treatment Schools—Effects of Having *100 Book Challenge*:					
	The Baseline Student	-1.131	-0.496	85	NS
	Grade in 2004	-0.272	-0.629	15,957	NS
	Ever Repeated Grade	0.426	0.460	15,957	NS
	In Special Education	1.386	1.038	15,957	NS
	African American	3.303	2.730	15,957	.007
	Latino	1.888	1.262	15,957	NS
	Fall 1st Grade Reading Level	2.696	4.213	15,957	.001