# An Evaluation of the 100 Book Challenge Program in the Schools Funded by the William Penn Foundation 

January, 200I

Dr. Joseph DuCette

Department of Educational Psychology
Temple University

## Executive Summary

This report presents a series of analyses on data collected to evaluate the "100 Book Challenge Program." The "100 Book Challenge Program" is designed to help students in the School District of Philadelphia improve their level of proficiency in reading. An integral part of the Program is the stipulation that students read a minimum of roo books during a designated period of time. Initially, the Program was implemented for approximately I 4 months in the first, second, and third grades in twelve elementary schools in the District. The current evaluation focuses on ten schools that were added to the Program as part of a grant from the William Penn Foundation. At the time of data collection, these schools had been participating in the Program for less than one full school year.

The evaluation utilized a quasi-experimental design in which students in Program schools were compared to students in matched schools. The matching was based on four criteria: the School District's Achievement Accountability Index; the grade range serviced in the school; the poverty level of the students; and, the racial distribution of the students within the school. The Stanford Achievement Test 9th Edition (SAT-9) was used to assess the reading level of Program and comparison students. The SAT-9 was given at the end of the 1999-2000 academic year. Data for the second and third grades were obtained as part of the District's regular testing program. Data for first grade students were obtained through a special administration of the SAT-9.

The results of the evaluation indicated that:

- First grade students are reading at levels higher than would be expected if the Program had not been implemented.
- Program students in the second grade attained significantly higher levels of reading achievement as compared to students in non-Program schools.
- Although the difference between Program and comparison students at the third grade was not significant, the average score of Program third graders was higher than the score for students in the comparison schools.
- Program students in classes with a higher degree of Program implementation attain significantly higher reading scores than students in classes with a lower degree of implementation.
- Data from a questionnaire administered to principals and teachers in the Program schools indicated strong support for the Program.
Overall, the evaluation indicates clear support for the "100 Book Challenge Program." Suggestions for future research on the Program are provided at the end of the Report.


# An Evaluation of the " 100 Book Challenge Program": Year Two 

Dr. Joseph DuCette<br>January, 2001

## A • Introduction

Presented in this report are data concerning the effectiveness of the "100 Book Challenge Program" (hereafter referred to as "the Program"). A previous evaluation of the Program was submitted in October, i999. As described in that report, the "100 Book Challenge Program" is a motivational and instructional support system where students are asked to read a minimum of ioo books at levels appropriate to their age and reading proficiency throughout a specific period of time. Students may read more than this, and data are collected through the Program on the number of children who meet the goal of ioo books read, as well as 200 and 300 books. Both students and parents are asked to verify that the books have actually been read. Teachers who participate in the Program are provided with an array of books at various difficulty levels for their classroom. The central purpose of the Program is to improve the reading achievement of participating students.

In the previous evaluation report, data were presented which demonstrated that students in Program classes in the first, second, and third grades were reading at significantly higher levels than students in matched classrooms. Reading achievement was measured through the Stanford Achievement Test, 9th Edition (or, SAT-9 as the test is often known), which is used throughout the School District of Philadelphia. The data indicated that the Program was somewhat less effective in the second grade, although the average SAT-9 scores of the Program second graders were higher than the average score for the matched control students. The previous evaluation was conducted after the Program had been implemented in the School District for fourteen months. The current evaluation uses data collected one year later, at the end of the 1999-2000 academic year. The specific focus of this evaluation is on ten schools which are part of a new cohort of Program participants funded by the William Penn Foundation. These cohort schools have been participating in the Program for less than a year.

## B - Methodology and Instrumentation

Data were collected from students in the ten elementary schools in the District which were participating in the Program as part of a grant from the William Penn Foundation. A matched sample of schools was chosen as the comparison group by the Office of Research and Evaluation of the School District. The schools were matched with the Program schools on four variables, in the following priority:

- Achievement Accountability Index Score
- Grade range serviced in the school (e.g., K-8; K-6)
- Poverty level of the students
- Racial distribution of the students

To make this comparison as valid as possible, each Program school was matched with two comparison schools, one immediately above the Program school in rank order on the Achievement Accountability Index Score and one immediately below, with the further stipulation that the grade range of the school was identical. Twenty comparison schools were chosen by this method. Consistent with the previous evaluation study, this study employs what Campbell and Stanley (1966) call a quasi-experimental design. This design is both common and acceptable in the educational evaluation literature.

The primary dependent variable for the evaluation are scores from the SAT-9. The data from the SAT-9 are expressed as Normal Curve Equivalent Scores (NCE's). NCE's are a form of derived scores which are standardized so that the average is set at 50 for each grade. NCE's are similar to, but are not identical with, percentile rank scores. NCE's are most commonly used to report aggregated statistics on school districts or schools within districts. For students in the third grade and above, the Stanford-9 produces scores in three areas for reading: a multiple choice reading comprehension score; an open-ended reading comprehension score; and a composite reading score. For students in the first and second grade, only the open-ended score is available. Moreover, since the previous report found that the composite score and the multiple choice score were redundant, it was decided not to use the multiple choice score for students in the third grade. The data presented in this study were collected at the end of the 1999-2000 school year. As in the previous study, the data for the first grade students were collected through a special administration of the SAT-9, since the SAT-9 is not normally given by the School District to students lower than second grade.

## C • Differences Between the Current and Previous Evaluations

There are several differences between the two evaluations which should be noted. The most notable is that the previous evaluation design used classes within the same school as the comparison group. Since the Program has been more widely implemented in participating schools during the time between the two evaluations, it was felt that a reasonable comparison group could not be created by using the method of within-school matching. Moreover, there was some concern expressed about the prior design that there might have been uncontrolled differences between the two sets of classrooms which may have affected the results. The current design, therefore, was not only necessitated by the nature of the schools used in the Program as part of the William Penn cohort, but may also be somewhat stronger than the previous study.

A second difference between the two studies involves the collection of direct evaluation data from the principals and teachers in the Program schools. Brief questionnaires were
constructed for the principals and teachers asking them their perception of the Program. Descriptive data from these questionnaires will be presented in this report.

An additional difference between the two studies involves the unit of analysis. In the previous study, data were available only at the classroom level. Therefore, the class mean served as the unit of analysis. For the current study, data were available for individual students. This differing data set provides considerably more power in the statistical analyses. In order to make direct comparisons between the two studies possible, however, the data will be analyzed using both the student and the class mean as the unit of analysis.

A final difference between the two evaluations concerns the existence of comparison schools for the first grade. In the previous evaluation, comparison first grade classrooms were chosen and were included in the special administration of the SAT-9. This proved to be both costly and problematic. As a consequence, there were no comparison first grades chosen for this study. Rather, the comparison that will be made is between the scores of the Program schools and the population mean for the SAT-9. In addition, data from this year's evaluation will be compared to the data from the previous study.

## D • Sample Description

Table i presents a breakdown of classes and students for the Program schools. Comparable data on the comparison schools are presented in Table 2.

Table 1: Sample Description of Program Schools

|  | First Grade |  | Second Grade |  | Third Grade |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
| School | Classes | Students | Classes | Students | Classes | Students |
| Adaire | 3 | 59 | 2 | 52 | 3 | 59 |
| Blankenburg | 4 | 80 | 3 | 76 | NA | NA |
| Edmonds | 2 | 40 | 1 | 26 | 2 | 49 |
| Elkin | 3 | 67 | 1 | 29 | 3 | 94 |
| Greenberg | 2 | 47 | 3 | 64 | 2 | 61 |
| Houston | 3 | 78 | 2 | 57 | 3 | 70 |
| Kelly | 4 | 116 | 1 | 29 | NA | NA |
| Mayfair | 2 | 51 | 2 | 55 | 2 | 55 |
| Steel | 5 | 141 | 4 | 118 | NA | NA |
| Wright | 3 | 66 | 3 | 70 | 1 | 29 |
| Total | $\mathbf{3 1}$ | $\mathbf{7 4 5}$ | $\mathbf{2 2}$ | $\mathbf{5 7 6}$ | $\mathbf{1 6}$ | $\mathbf{4 1 7}$ |

Note: NA indicates that there were no classrooms in the sample for that grade in the specified school.

Table 2: Sample Description of Comparison Schools

|  | Second Grade |  | Third Grade |  |
| :--- | :---: | :---: | :---: | :---: |
| School | Classes | Students | Classes | Students |
| H.A. Brown | 3 | 81 | 3 | 77 |
| J.H. Brown | 3 | 73 | 4 | 70 |
| Belmont | 3 | 71 | NA | NA |
| Bethune | 6 | 148 | 8 | 154 |
| Disston | 2 | 58 | 4 | 80 |
| Dick | 2 | 51 | 2 | 44 |
| Duckry | 3 | 77 | NA | NA |
| Fitler | 2 | 56 | NA | NA |
| Frank | 6 | 119 | 6 | 97 |
| Hancock | 3 | 78 | 3 | 85 |
| Jenks | 2 | 51 | 2 | 50 |
| Logan | 3 | 58 | NA | NA |
| Marin | 2 | 62 | 2 | 55 |
| McKinley | 2 | 43 | 2 | 39 |
| Pastorious | 5 | 134 | 3 | 93 |
| Potter-Thomas | 3 | 79 | 3 | 63 |
| Powel | 2 | 56 | NA | NA |
| Pratt | 2 | 54 | NA | NA |
| Prince Hall | 5 | 118 | 4 | 112 |
| Wister | 4 | 98 | 3 | 87 |
| Total | $\mathbf{6 3}$ | $\mathbf{1 5 6 5}$ | 47 | $\mathbf{1 1 0 6}$ |
|  |  |  |  |  |

Note: NA indicates that there were no classrooms in the sample for that grade in the specified school.

As demonstrated in Tables 1 and 2, there were 69 classes represented in the Program sample (3I first grades, 22 second grades, and 16 third grades), with a total sample size of 1738 students. The
comparison sample consisted of no classes ( 63 second grades and 47 third grades) with a total of 2767 students.

## E • Sat-9 Data for Students in the First Grade

The means for the open-ended scores on the SAT-9 for Program first grade students are presented in Table 3, by school.

## Table 3: Mean SAT-9 Scores for First Grade Program Students

| School | Mean SAT-9 Score |
| :--- | ---: |
| Adaire | 37.69 |
| Blankenburg | 55.93 |
| Edmonds | 54.46 |
| Elkin | 42.20 |
| Greenberg | 62.49 |
| Houston | 46.20 |
| Kelly | 45.76 |
| Mayfaire | 50.12 |
| Steel | 48.95 |
| Wright | 36.20 |
| Total | 47.33 |

As demonstrated in Table 3, the average scores per school range from a high of 62.49 (Greenberg) to a low of 36.20 (Wright). The average across all of the first grade students tested was 47.23 . Two statistical analyses were conducted to provide additional perspective on these data. First, the mean of each school, as well as the total mean, were compared to the population mean for the SAT-9 through a one-sample $z$ test. A one-sample $z$ test answers the question: what is the probability that a specific sample is taken from a population, when the mean and the standard deviation of the population are known? This translates in this case into the question: are the students in each school scoring at a level significantly different (either higher or lower) than the national average on the SAT-9. Assuming that the population mean of the SAT-9, expressed as NCE's, is 50 , and that the population standard deviation is 21.06 , this test can be used to ascertain if the average score in each school is significantly different from the population mean. The results of this analysis are presented in Table 4.

Table 4: Results of One-Sample Z tests Comparing Program Schools to the National Average on the SAT-9

| School | Deviation From <br> Population Mean | Z score | Level of <br> Significance |
| :--- | :---: | :---: | :---: |
| Adaire | -12.31 | 4.71 | .000 |
| Blankenburg | 5.93 | 2.64 | .008 |
| Edmonds | 4.46 | 1.41 | NS |
| Elkin | -7.80 | 3.18 | .001 |
| Greenberg | 12.49 | 4.26 | .000 |
| Houston | -3.80 | 1.67 | NS |
| Kelly | -4.24 | 2.99 | .002 |
| Mayfair | .12 | .42 | NS |
| Steel | -1.15 | .68 | NS |
| Wright | -13.8 | 5.58 | .000 |
| Total | -2.67 | 3.83 | .000 |

As demonstrated in Table 4, two of the Program schools are significantly higher than the national average (Blankenburg and Greenberg), four are not significantly different from the national average (Edmonds, Houston, Mayfair and Steel), and four are significantly lower than the national average (Adaire, Elkin, Kelly and Wright). Since all of these schools are urban, it should be viewed as positive that the students in the first grade in six of the ten schools are scoring at levels at or above the national average.

As an additional comparison, the results from the first grade testing for the current evaluation were compared to the data collected last year on the Program and comparison students. These data are presented in Table 5.

## Table 5: First Grade Results from Previous and Current Evaluations

| Group | Mean SAT-9 Scores |
| :--- | ---: |
| 1998-1999 Program Students | 43.72 |
| 1998-1999 Comparison Students | 33.16 |
| 1999-2000 Program Students | 47.33 |

Separate sample t-tests were computed on the data in Table 5 which indicated that the current first grade Program students are scoring at a level significantly higher than both last year's Program and comparison students.

As another assessment of how well the first grade students in the William Penn cohort are reading, an analysis was performed using level of performance as the metric. The School District of Philadelphia uses a four-level system to classify performance on the SAT-9: Below Basic; Basic; Proficient; and, Advanced. The percentages of first grade Program students in the William Penn cohort, as well as data from the previous Program and comparison first graders, are presented in Table 6.

Table 6: Percentages of First Grade Students at Four Levels of Performance for the Current and Previous Evaluations

|  | Current First Grade <br> Cohort Students | Previous Program <br> Students | Previous Comparison <br> Students |
| :--- | :---: | :---: | :---: |
| Below Basic | $22.4 \%$ | $34.94 \%$ | $51.53 \%$ |
| Basic | $39.9 \%$ | $26.69 \%$ | $31.31 \%$ |
| Proficient | $30.8 \%$ | $29.28 \%$ | $14.91 \%$ |
| Advanced | $6.9 \%$ | $6.08 \%$ | $2.21 \%$ |

These percentages were compared to ascertain if they differed from each other. These results are presented in Table 7.

Table 7: Comparison of Percentages for Current Program, Previous Program and Previous Comparison Students

|  | Current Program vs. <br> Previous Program | Current Program vs <br> Previous Comparison | Previous Program vs. <br> Previous Comparison |
| :--- | :---: | :---: | :---: |
| Below Basic | .013 | .000 | .024 |
| Basic | .016 | .041 | NS |
| Proficient | NS | .000 | .007 |
| Advanced | NS | NS | NS |

It is evident from the data in Table 6 that the current Program students are performing less at the Below Basic level and more at the Basic and Proficient levels as compared to students in the previous Program classes.

## F • Analysis of the SAT-9 Data for Second and Third Grade Students Using the Student as the Unit of Analysis

The means for the open-ended and composite scores of the SAT-9 are presented in Table 8 by grade level for the Program and comparison students. As mentioned above, only the openended sub-scale is available for the second graders. The results of separate sample t-tests are also reported in this table.

Table 8: SAT-9 Data for Program and Comparison Students Using the Student as the Unit of Analysis
ProgramMean ComparisonMean $\quad \mathrm{t}$-test $\quad$ Significance

| Grade2: |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- |
| Open-Ended | 47.63 | 44.70 | 2.55 | .009 |
| Grade3: |  |  |  |  |
| Open-Ended | 43.39 | 42.46 | .83 | NS |
| Composite | 43.27 | 41.92 | 1.36 | NS |

It is evident from Table 8 that there is a significant difference between the Program and comparison students in the second grade in favor of the Program students. The difference between Program and comparison students is not significant in the third grade, although the means for both the open-ended and the composite scales are higher for Program students than for comparison students.

## G • Analysis of the SAT-9 Data For Second and Third Grade Students Using the Class Mean as the Unit of Analysis

As mentioned previously, the prior study used the class mean as the unit of analysis. To allow direct comparisons between the two evaluations, this analysis was replicated for the current study. Consistent with the previous study, means for each class were computed which contained a minimum of io students. The data from this analysis are contained in Table 9.

Table 9: SAT-9 Data for Program and Comparison Students Using the Class Mean as the Unit of Analysis

|  | ProgramMean | ComparisonMean | t-test | Significance |
| :--- | :--- | :--- | :--- | :--- |
| Grade2: |  |  |  |  |
| $\quad$ Open-Ended | 46.63 | 44.76 | 2.03 | .03 |
| Grade3: |  |  |  |  |
| Open-Ended | 44.91 | 43.45 | .48 | NS |
| Composite | 43.97 | 42.71 | .44 | NS |

It is evident from Table 9 that the results of this analysis using the class mean as the unit of analysis are identical to the analysis using individual student data. As before, there is a significant difference in favor of the Program classes in the second grade, but not a significant difference in the third grade.

## H • Analysis of the Performance Level of Program and Comparison Students

As mentioned previously, the School District categorizes performance on the SAT-9 into four levels. To ascertain if the Program and comparison schools at the second and third grade levels differed on this metric, an analysis was performed on these data. The data for second grade students are presented in Table io.

## Table 10: Percentages of Second Grade Students at Four Levels of Performance for the Program and Comparison Students

|  | Second Grade Program <br> Students | Second Grade Comparison <br> Students |
| :--- | :--- | :--- |
| Below Basic | $14.6 \%$ | $19.8 \%$ |
| Basic | $51.8 \%$ | $48.2 \%$ |
| Proficient | $28.4 \%$ | $26.2 \%$ |
| Advanced | $5.7 \%$ | $5.7 \%$ |

The chi square computed on the above data equaled $7.908, \mathrm{p}=.048$. As can be seen from the table, Program students are more typically at the Basic and Proficient Levels, and less typically at the Below Basic level, as compared to non-program students. A comparable analysis for the third grade data was not significant.

## I - Analysis Using the Implementation Scores

The Program routinely collects data which indicates how extensively the Program is actually been implemented in each school and each classroom. These data include the percentage of students reading roo books, the percentage reading 200 books and 300 books, and an implementation score derived from classroom observation. In the previous evaluation, these data were used as a form of within-Program evaluation since it should be the case that students in classrooms in which the Program has been implemented to a greater extent should be reading at higher levels as compared to students in classrooms in which the Program had been implemented less extensively. An identical analysis was conducted for the present study.

Two analyses were conducted: one using the school as the unit of analysis and one using the individual classroom. Pearson correlations were computed between the SAT-9 open-ended
reading scores and the percentage of students reading 100,200 and 300 books, as well as the Implementation Score. The correlations are presented in Table II.

Table 11: Pearson Correlations Between SAT-9 Scores and Implementation Indices

|  | Pearson Correlations <br> Using the School as the <br> Unit of Analysis | Pearson Correlations <br> Using the Classroom as <br> the Unit of Analysis |
| :--- | :---: | :---: |
| Percentage Reading 100 Books | .40 | .05 |
| Percentage Reading 200 Books | .44 | $.39^{*}$ |
| Percentage Reading 300 Books | $.56^{*}$ | $.58^{*}$ |
| Implementation Score | .29 | NA |

NA: Implementation score data were not available for the individual classrooms.
It is evident from Table ir that there is a relationship between implementation and Program success. This is most evident in the correlations for the percentage of students reading 200 or 300 books. It should also be remembered that the sample size for the analysis using the school as the unit is only 10 , making it difficult to obtain significance. In general, the more extensively the Program has been implemented, the better the students are reading.

## J•Analysis of Questionnaire Data

The following section presents the results of the brief questionnaire which was administered to the principals and teachers in the Program schools. These results are presented below, with the data from the principals contained in Table 12 and the teacher's data in Table ${ }_{3}$ Nine of the ten principals completed the questionnaire. A total 89 teachers completed the questionnaire.

[^0]$\dagger \mathrm{p}<.01$

## Table 12: Responses of Principals to the Attitude Questionnaire

| Question from Questionnaire | 1 | 2 | 3 | 4 | 5 | Mean |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 1 | Usage of the 100 Book Challenge has stimulated teacher <br> professionalism and creativity. | 0 | 0 | 1 | 4 | 4 | 4.33 |
| 2 | Usage of the 100 Book Challenge has stimulated positive <br> teacher attitude | 0 | 0 | 0 | 4 | 5 | 4.56 |
| 3 | Involvement in the 100 Book Challenge has strengthened <br> the overall reading instruction program in the school. | 0 | 0 | 0 | 2 | 7 | 4.78 |
| 4 | The support provided by the 100 Book Challenge |  |  |  |  |  |  |
| staff has assisted in strengthening the overall reading <br> instruction program in the school. | 0 | 0 | 0 | 1 | 8 | 4.89 |  |
| 5 | Following the initial training and follow-up support by <br> 100 Book Challenge staff, school staff are comfortable <br> with continuing the Challenge in subsequent years | 0 | 0 | 0 | 4 | 5 | 4.56 |
| 6 | I intend to continue the 100 Book Challenge program in <br> subsequent years. | 0 | 0 | 1 | 0 | 8 | 4.78 |
| 7 | 0 | 0 | 1 | 0 | 8 | 4.78 |  |
| Overall, I would rate the 100 Book Challenge as a <br> valuable program. | 0 |  |  |  |  |  |  |

Note: 1 is rated as "Do not Agree"
5 is rated as "Strongly Agree"
It is evident from the data in Table i2 that the principals are overwhelmingly positive about the Program. With the exception of Question \# I (which concerns the increase in teacher professionalism), all of the means are above 4.5 (and this mean is only slightly below this level). In general, then, the principals of Program schools believe that the Program is having a positive effect on the students in their schools and strongly support its value.

Table 13: Responses of Teachers to the Attitude Questionnaire

| Question from the Questionnaire | 1 | 2 | 3 | 4 | 5 | Mean |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 Using the 100 Book Challenge has stimulated my personal professionalism and creativity. | 4 | 3 | 23 | 28 | 21 | 3.78 |
| 2 Involvement in the 100 Book Challenge has strengthened the overall reading instruction program in my class. | 0 | 2 | 12 | 27 | 48 | 4.36 |
| 3 The support provided by the 100 Book Challenge staff has strengthened my personal skills in providing reading instruction. | 3 | 7 | 22 | 36 | 21 | 3.73 |
| 4 Following the initial training and follow-up support by 100 Book Challenge staff, I am comfortable with continuing the Challenge in subsequent years | 1 | 3 | 7 | 24 | 54 | 4.43 |
| 5 As a result of participating in the 100 Book Challenge program, my students care more about reading. | 0 | 0 | 6 | 18 | 65 | 4.66 |
| 6 As a result of participating in the 100 Book Challenge program, my students have increased their vocabulary. | 0 | 3 | 12 | 41 | 33 | 4.17 |
| 7 As a result of participating in the 100 Book Challenge program, my students have increased their writing capability. | 0 | 6 | 29 | 40 | 14 | 3.70 |
| 8 As a result of participating in the 100 Book Challenge program, my students more likely will go to the library. | 0 | 3 | 25 | 37 | 24 | 3.92 |
| 9 Overall, I would rate the 100 Book Challenge as a valuable program. | 0 | 1 | 5 | 23 | 60 | 4.60 |

Note: 1 is rated as "Do not Agree"
5 is rated as "Strongly Agree"
Although the means of the questions in Table 13 are somewhat lower than the comparable means in Table I2 it is evident from the data in Table I3 that the teachers are also positive about the Program. This is evidenced by an overall mean of 4.6 (out of a possible 5) for Question \# 9 which assesses the teacher's overall perception of the Program. It is also evident that the teachers believe that the Program causes the students to care more about reading (a mean of 4.66 for Question \# 5), and that they are comfortable continuing the Program in subsequent years ( 4.43 for Question \# 4). The two lowest means are for Question \# 7 which concerns the impact of the Program on students' writing skills, and Question \# 3 which asks about the support from the Program staff. Overall, however, the rating data from both the principals and teachers are clearly positive about and supportive of the Program. As further evidence of this, several
of the teachers and principals included comments at the end of the questionnaire. A complete list of these comments is contained in Appendix A. With only minor exceptions, all of these comments are positive.

## K • Summary and Conclusions

In general, the data presented in this report support the contention that the "100 Book Challenge Program" is having a significant impact on the reading achievement of students participating in the Program who are in the iо schools funded by the William Perm Foundation. There are several aspects of the data which support this contention:

- The SAT-9 scores for second grade students are significantly higher than the scores for the students from the matched schools. Moreover, the percentages of students performing at the Basic and Proficient levels in the second grade Program schools are higher than the percentages at these levels in the comparison schools. This is coupled with the fact that the percentage of Program students performing at the Below Basic level is somewhat lower than the percentage of students in the comparison schools performing at this level.
- Although the evidence for the performance of students in Program first grade classes is indirect (since no comparison first grade classes were included in the sample), the evidence that was presented indicated that the first grade Program students are also performing at higher levels than would be expected if the Program had not been implemented.
- Although significant differences were not found for third grade Program students, there was a slight trend for these students to be performing at higher levels than comparison students.
- The implementation data indicated that students in classrooms in which the Program had been more strongly implemented demonstrate higher reading achievement than students in the classrooms with lower implementation scores.

In addition to these reading achievement data, the responses from the questionnaire administered to principals and teachers in Program schools were uniformly positive about the Program. What is perhaps most interesting about these positive findings is that the Program has been implemented in the Program schools (that is, the ten schools which constitute the cohort sample funded by the William Perm Foundation) for less than one full school year. It is evident from the data that the extent of Program implementation among the cohort schools, and among classes within single schools, is uneven. Even with this limitation, the Program has been shown to have an impact on the reading achievement of the participating students. When the Program has been more fully implemented, its impact may be even greater.

While in no way diminishing the positive findings of the evaluation, there are some aspects of the results that warrant consideration. Among these are the following:

- The fact that there was not a significant difference at the third grade between Program and comparison students might be worth an investigation into possible discrepancies between the motivational impact of the Program on older students as compared to students in the first and second grades. In particular, if the Program intends to increase its coverage into the upper elementary grades, it is perhaps worth finding out if a somewhat different approach is necessary for these students. One possible source of this information would be the current third grade teachers to ascertain in more depth their perception of the Program (since only one third grade teacher responded to the questionnaire).
- The least positive response to the teacher questionnaire was about support from Program staff. It may be that the participating teachers expected a somewhat higher level of support than they received. If this is a mis-match between perception and reality, or if the Program is not able to provide the support it would like because of resource limitations, it would seem to be worthwhile for the Program to find out why the teachers are somewhat less satisfied with this aspect of the Program.
- Consistent with the research design used in the last study, the impact of the Program is being assessed within a relatively short time after the Program has been implemented in a school. There are two possible effects that this might have. On one hand, it is possible that the Program's impact is being under-estimated since any new program takes time to be fully realized. On the other hand, there is ample evidence in the educational literature that some programs have a short-term impact because they are new and innovative, but lose their impact with time. It is critical that long-term studies of the "100 Book Challenge Program" be completed to ascertain what effect the Program has when implemented in a school for several years.
- As a related comment, it is also important to conduct follow-up studies on the students who participate in the Program to see if the improvement in their reading is maintained as they continue with their schooling. In this regard, follow-up data were collected on students from the previous evaluation who were in the first grade and are now is second grade within the School District. These data are currently being analyzed and will be reported shortly.

All things considered, therefore, these results, like the results from the previous evaluation, support the conclusion that the " 100 Book Challenge Program" is having a positive impact on the reading achievement of its participants.

## Appendix A: Comments from the Open-Ended Question on the Principal and Teacher Questionnaire

Note: The respondent is indicated by a "P" for a principal and "T" for a teacher. For teachers, the grade level taught is also indicated.

- The ratings indicated above speak for themselves. This has been a positive and rewarding program at every level for students, parents and educators. The expense involved, although difficult for schools with small budgets, is not exorbitant. We are working to expand the program in our upper grades. (P)
- Excellent program. I use the program at home with my children. Really believe highly in the validity of the program. (T-K)
- I think the program is very useful overall, for book handling skills and pre-reading skills. (T-K)
- 100 Book Challenge is an excellent program for motivating children to read independently. The skill cards are excellent tools for teaching and reinforcing reading skills. It is also a great dual-purpose reading program. (T-1)
- Excellent program for balanced literacy, especially with regard to the "Independent Reading" by the students ( $\mathrm{T}-2$ )
- The 100 Book Challenge is an excellent program to motivate/stimulate reading. The children have developed a love for books and are avid readers. They look forward to the 100 Book Challenge Program every morning. (T-1)
- I would even use the program in the middle and upper grades as incentives to keep on reading. (P)
- It's going very well. Thank you for helping us with the program. ( P )
- The first year was more difficult. Correcting last year's mistakes has made this year flow much more smoothly. This class is much farther in the number of books than last year's class, at the same point in time. The children love it and so do I. (T-K)
- I love the program. The children are so motivated to read. (T-1)
- My students have become very competitive and motivated by the program. (T-2)
- 100 Book Program provided individual opportunities for growth at a comfortable level. (T)


[^0]:    * $p<.05$

