FIVE YEARS OF READ 180 IN DES MOINES: IMPROVING LITERACY AMONG MIDDLE SCHOOL AND HIGH SCHOOL SPECIAL EDUCATION STUDENTS

Gina M. Hewes
Nancy Palmer
M. Bruce Haslam
Monica B. Mielke

April 2006

Prepared for
Scholastic, Inc. and
Des Moines Independent School District
1801 16th Street
Des Moines, IA 50314

This report was prepared under a contract from Scholastic, Inc.
The findings and conclusions are those of the authors. They do
not necessarily represent the opinions of Scholastic or the
Des Moines Independent School District and no endorsement should be inferred.

Contents

List of Exhibits

Page

List of Exhibits (continued)

Summary of Findings

Over the first five years of implementation, more than 1,200 special education students have participated in Scholastic READ 180 in the Des Moines Independent Community School District. During this period, district personnel have produced annual reports on student outcomes and they have collected evidence that indicates that the program has been implemented reasonably well across the district.

This report builds on the school district's reports and provides additional analyses that take advantage of the full complement of five-year longitudinal data. Analyzing the data as a longitudinal data set, rather than sequentially as a series of yearly data sets, offers an opportunity to investigate and identify growth in student learning outcomes over time.

The longitudinal database available for this study covers the school years 2000-01 through 2004-05, and provides student-level achievement results on the following assessments:

- Stanford Diagnostic Reading Test (SDRT) Comprehension (Version 4)
- Stanford Diagnostic Reading Test (SDRT) Vocabulary (Version 4)
- Stanford Diagnostic Reading Test (SDRT) Total Reading (Version 4)
- Scholastic Reading Inventory (SRI)
- Iowa Tests of Basic Skills / Educational Development (ITBS/ITED) Comprehension.

For the analyses reported here, student scores on each of these achievement measures were treated as outcome variables in five separate series of analyses. Each series included a variety of statistical procedures, ranging from the application of simple descriptive statistics, such as computing group means and standard deviations, to the multivariate, multilevel inferential technique of growth curve modeling. Special education students who were no longer in READ 180 served as the comparison group.

Overall, the study findings indicate that for special education students in Des Moines, the effect of READ 180 is positive and statistically significant. While former READ 180 students continued to make achievement gains after leaving the program, their gains were not as large as gains for students who remained in READ 180 for an additional year. Specifically, there was a statistically significant effect on three of the achievement measures, and no measurable effect on the other two. Thus, each additional year of participation in READ 180 was associated with annual increases of approximately 30 Lexiles on the SRI, 6 scale-score points on SDRT4 Comprehension, and 5 scale-score points on SDRT4 Total – above and beyond the observed yearly growth in achievement. Respectively, these differences in scale-score points can be translated into differences of 0.26, 0.40, and 0.33 years of growth, above and beyond expected annual growth. These findings are displayed in Exhibits A, B, and C, which follow below.

Exhibit A
Annual Achievement Growth on SDRT4 Comprehension: READ 180 Students versus Comparison Students

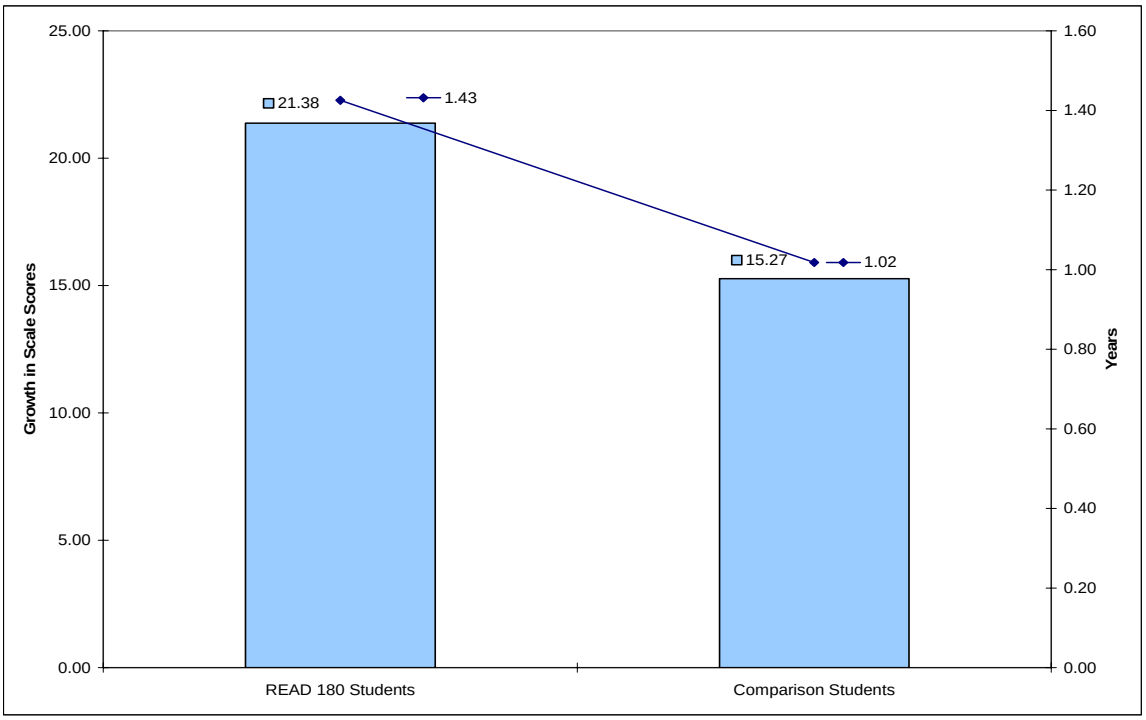

Exhibit reads: READ 180 students showed an annual growth of 21.38 scale-score points on SDRT4 Comprehension, which corresponds to 1.43 years of growth. In contrast, comparison students showed an annual growth of 15.27 scale-score points on the same measure, corresponding to 1.02 years of growth. The difference between the annual increases in scale scores achieved by the two groups of students is statistically significant at $p<0.05$.

Exhibit B
Annual Achievement Growth on SDRT4 Total: READ 180 Students versus Comparison Students

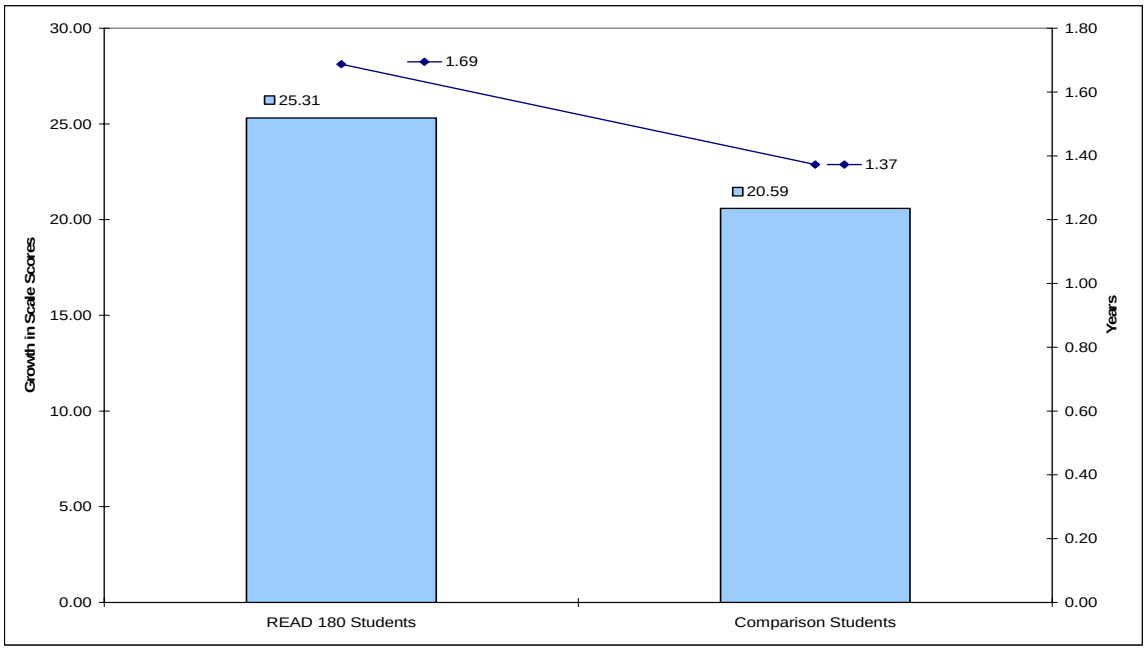

Exhibit reads: READ 180 students showed an annual growth of 25.31 scale-score points on SDRT4 Total, which corresponds to 1.69 years of growth. In contrast, comparison students showed an annual growth of 20.59 scale-score points on the same measure, corresponding to 1.37 years of growth. The difference between the annual increases in scale scores achieved by the two groups of students is statistically significant at $p<0.05$.

Exhibit C
Annual Achievement Growth on the SRI: READ 180 Students versus Comparison Students

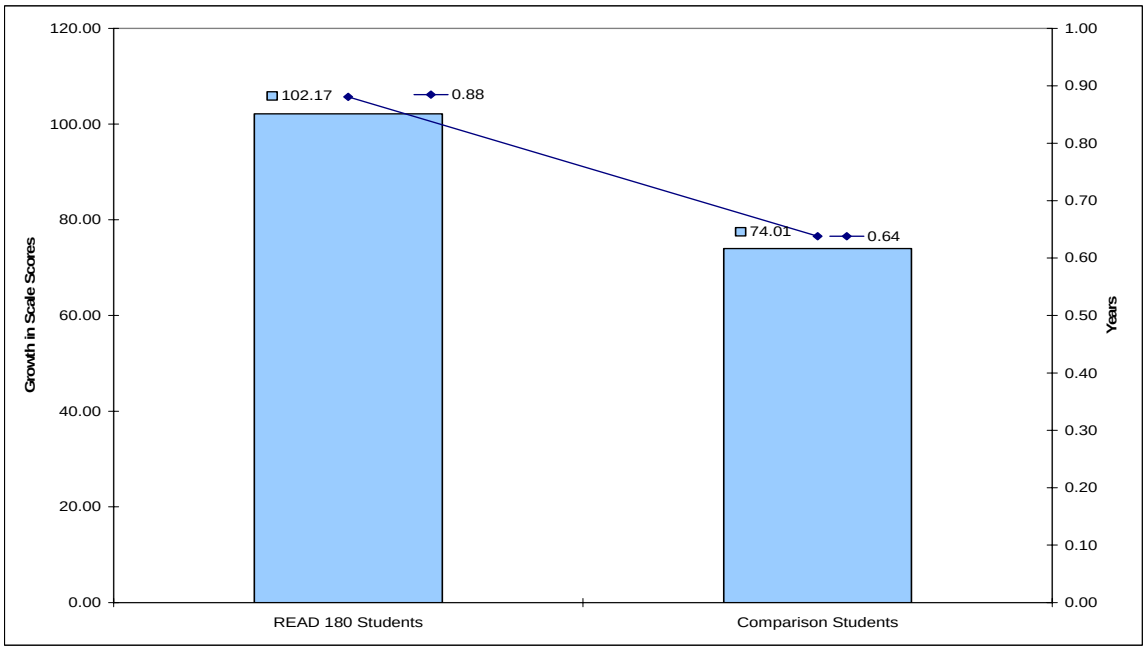

Exhibit reads: READ 180 students showed an annual growth of 102.17 scale-score points on the SRI, which corresponds to 0.88 years of growth. In contrast, comparison students showed an annual growth of 74.01 scale-score points on the same measure, corresponding to 0.64 years of growth. The difference between the annual increases in scale scores achieved by the two groups of students is statistically significant at $p<0.05$.

Introduction

Scholastic Inc.'s READ 180 program was first implemented in the Des Moines Independent Community School District in the 2000-01 school year. Participants included sixth- to eighth-grade special education students in 10 middle schools. Participation expanded to one high school in 2001-02, and to the remaining high schools in 2005-06. Also in 2005-06, READ 180 participants for the first time included students from the regular education program.[1] Overall, more than 1,200 Des Moines special education students participated in Scholastic READ 180 over the first five years of implementation. This report focuses exclusively on these special education students because data are not yet available from the 2005-06 school year.

During the first five years of implementation, district staff prepared annual reports on student outcomes and they also reported on implementation. This report builds on the district reports and provides additional analyses that take advantage of the full complement of five-year longitudinal data.

The first section of this report reviews previous research on READ 180 in Des Moines, including district staff's look at the fidelity of the implementation and research on the program's effectiveness. Next, there is a discussion of the data and analytic procedures used in this study. The third section presents findings about student achievement gains and growth trends. The fourth section reviews the context and implications of the findings. The report also contains two appendices. Appendix A discusses the study's descriptive results, with information about means and correlations. Appendix B is a technical appendix that provides detail about the study's growth curve modeling methods.

I. Prior Research on READ 180 in Des Moines

District reports on program implementation are based on two rounds of classroom observations and teacher interviews conducted by district staff (Palmer, 2004). The first round of data collection took place in January-April, 2002, and included 23 READ 180 teachers. The second round of data collection took place in April-May, 2004, and included 19 teachers, six of whom had been included in the earlier round of data collection. Overall, data were collected on 36 teachers in a total of 42 observations and interviews. These samples represent nearly all of the teachers who were implementing READ 180 during the two school years.

Findings from these two reports suggest that the program was reasonably well-implemented, at least in terms of the time allocations called for in the program model (i.e., a 90-minute instructional block broken into 20 minutes of whole-group instruction, followed by three 20-minute rotations for small-group instruction, individual work with READ 180 software, and

[1] As of the 2005-06 school year, there are two READ 180 programs in nearly all 10 middle schools and all five high schools in Des Moines: one program in each building is targeted for special education students and one is for regular education students. All of the buildings serve primarily regular education students.

independent and modeled reading, and a 10-minute wrap-up).[2] Program materials were available in the classrooms and the required hardware was also available, although observers and teachers both reported some problems with the functionality of the hardware. Observers and teachers generally agreed that classroom environments were orderly and conducive to students' working on required activities. Teachers also expressed confidence in their ability to use the program effectively.

An important limitation of the implementation reports was the lack of data on instruction. Therefore, while it was possible to conclude that the program was reasonably well-implemented in terms of the presence of key artifacts and resources, it was not possible to draw any conclusions about overall instructional quality.

District staff's annual reports on student learning outcomes document gains on the Stanford Diagnostic Reading Test (SDRT4) Comprehension[3] and Total, Scholastic Reading Inventory (SRI), Iowa Tests of Basic Skills/Educational Development (ITBS/ITED) Comprehension, and a district fluency probe[4] after one, two, and three years of participation in the program (Palmer, 2003a, 2003b, 2004, forthcoming). These reports present their results in terms relative to expected gains and in terms of grade-equivalents for test scores and gains.

One distinct pattern persists across all years and across most achievement measures: *Students in their first year of READ 180 make gains larger than expected for a single year of achievement growth.* That is, students gained more than the equivalent of a single grade level in a single year. This was true for SDRT4 Total, SDRT4 Comprehension, SRI, and the district-developed reading fluency probe. These results are particularly impressive given that these students had previously been gaining much less than expected for a year of growth. Furthermore, in the first two years of implementation, approximately 18 percent of participating students achieved gains that were sufficiently large for the district to reduce the level of special education support for those students.

Achievement gains for 2002-03 were not as high as for 2000-01 and 2001-02, but the results for 2003-04 were once again similar to the first two years, so the lower 2002-03 gains may simply be an anomaly in the otherwise strong results for students in their first year of READ 180.[5] Generally, achievement growth for students in a second year of READ 180 was not as large as growth from their first year of participation, and similarly, achievement growth for students in the third year were not as large as growth in the first or second years of participation.

---

[2] Not surprisingly, teacher reports about implementation and use of the program were somewhat more positive than those of observers. The data do not permit a valid assessment of which set of perspectives is more accurate, although it is not unreasonable to suggest that the reality lies somewhere in between. It is also important to note that there was some convergence of observational data and teacher reports in the second round of data collection.

[3] The district reports do not present results from the SDRT4 Vocabulary and Scanning subtests.

[4] The district fluency probes were described as "short reading passages taken from curriculum-based materials that the students are asked to read orally for one minute. The number of words read correctly is computed…Passages selected were one grade level below students' grade-level placement" (Palmer, 2003a). The fluency probes were not analyzed in the current research and are not elaborated further.

[5] The 2002-03 school year was also the first year a computer support person was no longer available to teachers on a daily basis (Palmer, 2003b).

In examining the effectiveness of READ 180, the district's research indicates that the program accelerated the learning of special education students in the middle grades, relative to the rate at which the students had been learning prior to enrolling in READ 180. While this improvement was clear and consistent, it should be noted that the program did not produce effects strong enough to bring many of the participants up to grade level during the period of their participation in the program.

Another study, conducted by Policy Studies Associates, Inc. (PSA), used a matched comparison group of non-READ 180 students in Des Moines to determine the program's effectiveness (Hewes & Haslam, 2004). This study, which focused on results of the ITBS comprehension subtest, found no statistically significant effects of READ 180 on student outcomes. Several differences in research design between the PSA study and the district's analyses may account for the different results. For example, the district's analyses focused on results from the SDRT4 and SRI while the PSA analysis relied on ITBS results.

The present study addresses some of the drawbacks of the earlier studies by presenting findings for all five achievement measures. It differs from the district reports in that, rather than reporting a sequential series of annual data on student achievement, it examines the full complement of five-year longitudinal data to identify trend patterns over time. Analyzing the data as a longitudinal data set permitted investigation of trend patterns such as the rate of achievement growth associated with READ 180, effects of number of years of a student's participation in READ 180, effects of patterns of participation/non-participation in READ 180, and interactions among these factors. Additionally, the present study investigates the achievement consequences for students who left the READ 180 program.

II. Data and Data Analysis

This section of the report describes the data and analytic procedures used in this study. Before turning to the particulars, we provide a brief review of our main analytical method, *growth curve modeling,* to help guide the reader through the complex mosaic of data, measures, and analyses that are described in this section.

Growth curve modeling, in contrast to a series of annual analyses, provides the opportunity to sort out different causes associated with change over time. Specifically, achievement change that is observed in a series of annual analyses could be attributable to three possible sources: (1) an intervention, such as READ 180; (2) normal individual growth that would occur anyway; or (3) changes that occurred in the population at large. These three sources of change are not mutually exclusive, and are probably all operating simultaneously. What growth curve modeling does is help sort out the three sources of change, and thereby provide more confidence in results associated with an intervention. In particular, growth curve modeling allows us to measure the sources of change and include them in analyses

Growth curve modeling requires multiple, sequential measures of the same concept (here, reading achievement) using the same test over time, with the same set of individuals (students), while controlling for other measures. Conceptually, there is a distinction between two types of

control measures: time-varying measures and non-time-varying measures. Some variables are inherent in the student – a good example is student gender. Such a variable will not change its value, no matter how many times it is measured. Other variables, in contrast, do vary from time point to time point. One example of a time-varying measure is a student's special education status. Across a series of testing time points, a student may move in and out of a special education program, and thus the value for this measure changes from time point to time point.

The distinction between time-varying and non-time-varying measures parallels the two levels of analysis in growth curve modeling. Growth curve modeling, as implemented for this study, has two levels of analysis – students *and* test points – which are modeled simultaneously. Modeling both levels simultaneously means we can include measures that are related to either level of analysis. In fact, measures related to the student level are the measures that are non-time-varying as discussed above. Measures related to the test point level are measures that are time-varying: their value may change from test point to test point.

A practical implication of this complex form of analysis is that having two levels of analysis means there are two sets of everything else as well: two sets of measures, as already discussed; two sets of sample sizes; two sets of coefficients. The remainder of this section of the report describes the data that were available, the measures used in analyses, and the statistical models employed.

Data

The Des Moines Independent Community School District's Department of Student and Family Services provided the data for this report, and these are the same data used in the annual status reports (Palmer, 2003a, 2003b, 2004). The database includes data for all students who participated in READ 180 in Des Moines during one or more of the first five school years during which READ180 was implemented: 2000-01 through 2004-05. In order to protect student confidentiality, all student names and other identifiers were deleted before the data were transferred to PSA for analyses.

Information available for each student included basic demographic information as well as test scores on several achievement measures. For each student for each year[6] – starting with the year he or she first participated in READ 180 (or occasionally the prior year) and continuing beyond participation in READ 180 – the database includes information about grade level, level of special education services provided, status of READ 180 participation, school attended, and student's teacher assignment. The database also includes scores on one more of the five achievement measures discussed earlier: SDRT4 Comprehension, Vocabulary, and Total;[7] SRI, and ITBS/ITED Comprehension. Some of these pieces of information were used directly as

[6] Most students have incomplete data across the five-year span – that is, most students do not have information for every year, but only for some years.

[7] On the SDRT4, reading comprehension and vocabulary are components of the Total Reading score, along with a third element, scanning. For SDRT4 Scanning, analyses showed that the relationship with READ 180 could not be determined: Standard errors of model coefficients were unstable, possibly due to wide variation in scanning scores, suggesting that the coefficients were not reliable estimates of effects. SDRT4 Scanning results are thus not available and not discussed further.

variables in statistical analyses; others were used to construct new variables for use in analyses. We will return to this distinction later in describing measures.[8]

Across the five years covered in this study, the annual number of students who enrolled in READ 180 for the first time ranged from just under 200 to almost 350. (See Exhibit 1.) Cohort 1, which included students whose first enrollment was in 2000-01, was the largest READ 180 cohort, enrolling 110 sixth-graders, 130 seventh-graders, and 102 eighth-graders – a roughly even distribution across the middle school grades that were the initial focus of READ 180 in Des Moines. Subsequent cohorts enrolled similar numbers of sixth-graders, but most seventh- and eighth-graders in subsequent years were students continuing from previous cohorts, with many fewer seventh- and eighth-graders identified for first-time enrollment. Cohorts 2 through 5 thus are somewhat smaller than the first cohort, with Cohort 4 the smallest of all of the cohorts.

Exhibit 1
First-Time Enrollment in READ 180 by *Cohort Year* and *Starting Grade Level*

Grade	Cohort 1 2000-01	Cohort 2 2001-02	Cohort 3 2002-03	Cohort 4 2003-04	Cohort 5 2004-05
6	110	112	123	95	124
7	130	52	57	40	70
8	102	49	45	31	10
9		5	1	4	6
10		16	6	4	5
11		10	2	2	2
Totals	342	244	234	176	217

Exhibit reads: Cohort 1 included 110 sixth-graders who participated in READ 180 for the first time.

Source: Authors' tabulations from Des Moines school district database

For analyses, we pooled all five of the cohorts into a single, large sample of students. The rationale for pooling was to increase sample size and thereby improve our statistical power for detecting effects. Additionally, pooling the cohorts while also controlling for a measure of cohort is a statistical technique that allows the growth curve modeling to tease out the three possible sources of change. After pooling the cohorts, the total number of students in the database who participated in a year or more of READ 180 is 1,213 – the sum across the bottom row of Exhibit 1.

The database includes pretest and posttest scores on at least one of the five achievement measures for all but one student. However, the number of students included in any one achievement analysis is smaller than the total number of students in the database. Specifically, scores for over 90 percent of the 1,213 students were available for the analyses of results on the

[8] It is important to note the distinction among three terms: data, measures, and variables. Data are the raw information in the database; data cannot be created, only collected. Measures are the conceptual factors we wish to study; measures may be directly available as raw data or they may be created from raw data. Variables are the technical version of measures, also called the "operationalization" of measures, and are used in analyses; variables are usually created. In some cases, the distinction among these three terms is purely conceptual, with little or no practical consequence. Usually, however, there are important practical, procedural differences among the three.

SDRT4 (Comprehension, Vocabulary, Total) and the SRI, and scores for about 60 percent of these students were available for the ITBS/ITED analyses. The reason for this difference is that the SDRT4 and the SRI in Des Moines are specifically targeted for READ 180 students. Every effort is made to administer the tests to as many students as possible as they enter READ 180 and again at the end of each year in READ 180. The Department of Student and Family Services has been extremely successful in this endeavor. In contrast, the ITBS/ITED is a test administered district-wide on a yearly basis. The extent to which any one student will have test scores in the two relevant consecutive years (to serve as pretest and posttest) is much less common on a yearly testing cycle.[9]

As discussed earlier, the analyses conducted for this study use two levels of analysis: students and their test points. For growth curve modeling, there thus are multiple data points for each student, with each data point representing an individual test score. Consequently, the number of data points in the database is greater than the number of students in the database. Exhibit 2 illustrates the number of data points that result from structuring the database in this way.

In Exhibit 2, we see that the number of students who have scores on SDRT4 Comprehension at pretest was 1,163. These students' test scores represent 1,163 data points in the database. At the time of the first posttest (the second overall test point), 1,175 students have scores on the SDRT4 Comprehension test; these students' test scores represent an *additional* 1,175 data points in the database, for a total of 2,338 data points after two testing time points. If we continue in this manner, adding the number of test scores available at each time point, we obtain a total of 2,806 students' test scores after the fifth posttest. Similar interpretations hold for the other achievement measures presented in Exhibit 2. For SDRT4 Vocabulary, there are 2,788 test scores; for SDRT4 Total, 2,738 test scores; for SRI, 2,740 test scores; and for ITBS/ITED Comprehension, 2,352.

[9] Note that, for any single year of ITBS data, the participation rates are well over 90 percent.

Exhibit 2
Numbers of Students Providing Test Data at Each Time Point

Testing Time Point	SDRT4 Comprehension	SDRT4 Vocabulary	SDRT4 Total	SRI	ITBS/ITED Comprehension
0 – Pretest (First test point)	1,163	1,161	1,140	1,168	990
1 – First posttest (Second test point)	1,175	1,161	1,139	1,122	888
2 – Second posttest (Third test point)	379	375	371	361	421
3 – Third posttest (Fourth test point)	75	75	74	74	45
4 – Fourth posttest (Fifth test point)	12	13	12	13	5
5 – Fifth posttest (Sixth test point)	2	3	2	2	3
Totals	2,806	2,788	2,738	2,740	2,352

Exhibit reads: One thousand one hundred sixty-three students provide test data from SDRT4 Comprehension at the pretest time point. Slightly more students (1,175) provide test data at the first posttest (that is, the second test point), and 379 provide test data at the second posttest (third test point) time point. The total number of student-test points available for growth analysis for SDRT4 Comprehension sums to 2,806.

Note: The counts given in this exhibit reflect observations that have complete data on all analysis variables.

Source: Authors' tabulations from Des Moines school district database

We now turn to a description of how the available data were used to create measures used in analyses.

Measures Used in Growth Curve Modeling

As noted earlier, it is important to maintain the distinction among data, measures, and variables. Data are the raw information in the database, while measures are the conceptual factors we wish to study, and variables are the technical, procedural version of measures used in analyses. We discussed previously the information that was available in the data file from Des Moines. Some of the information in the data file was useful as-is, while other information was useful for creating new measures. Recall also that we are here interested in two levels of analysis: students and test points. Measures can be associated with either students or test points, but not both.

Analyses in this report rely on the outcome and predictor measures listed in Exhibit 3. All five of the outcome measures are taken directly from the Des Moines database and are time-varying test-point-level measures. Of the six predictor measures, three are time-varying, test-point-level measures: *READ 180 status* and *special education status* are taken directly from the Des Moines database, and *time count* was created to designate the time between pretest and each posttest using the school year as the metric. The remaining three predictor measures are student-level non-time-varying measures: *cohort year, starting grade level*, and *participation pattern*.

Exhibit 3
Summary of Measures used in Analyses, by Type

***Type*/Name of Measure**	**Description**
Outcome Measures	
SDRT4 Comprehension Subtest Score	A measure of how well students understand and analyze reading material
SDRT4 Vocabulary Subtest Score	A measure of the range of words a student knows relative to grade-level norms
SDRT4 Total Reading Score	A composite measure of three reading subtests, scaled independently based on grade-level norms
SRI Lexile Score	A computer-adaptive measure of students' reading comprehension
ITBS/ITED Comprehension Subtest Score	A measure of students' ability to make inferences or generalizations about passages they have read
Predictor Measures – Time-Varying, Test Point Level	
READ 180 status	Student's participation status in READ 180 at each test point
Special education status	Student's participation status in special education services at each test point
Time count	A measure of the time between pretest and each testing point, in school-year increments
Predictor Measures – Non-Time-Varying, Student Level	
Cohort year	The school year during which a student first enrolled in READ 180
Starting grade level	Student's grade level assignment when first participated in READ 180
Participation pattern	The five-year pattern of entering, withdrawing, and re-entering READ 180

Cohort year and *starting grade level* are taken directly from the district database; *participation pattern* was created from information on student enrollment in READ 180 across the five years. The remainder of this section provides additional descriptions of each measure, starting with the achievement measures, moving to the time-varying predictor measures, and finishing with the non-time-varying predictor measures. Where relevant, a brief description is also provided of how measures were "operationalized" as variables in models. (For additional detail, please see the Appendix B.)

SDRT4. The SDRT4 is a multiple-choice, group-administered, norm-referenced assessment of students' strengths and needs in reading, for grades 2-12. The SDRT4 offers eight levels and two forms. In Des Moines, the purple level (geared towards grade levels 4.5-6.5), Form J, was administered to all test-takers, regardless of current grade-level assignment, in its paper-and-pencil form rather than online. The paper-and-pencil SDRT4, purple level, which covers literature, informational text, and functional material, provides subtest results as well as a composite score. The following three measures are treated separately in all analyses and discussions:

- ***SDRT4 Comprehension subtest*** measures how well students understand and analyze reading material, and includes some cloze questions and comprehension questions at informational, inferential, and prediction levels of analysis.

- ***SDRT4 Vocabulary subtest*** measures the range of words a student knows, relative to grade-level norms, and uses picture-word matching at the elementary level and synonym matching at higher levels.

- ***SDRT4 Total Reading*** is a composite of these two subtests plus a third, scanning, but is scaled independently based on grade-level norms rather than being a simple average of the three scores.

SRI. The SRI, which is an internal component of READ 180, is a computer-adaptive assessment for grades 1-12 designed to measure students' reading ability in terms of reading comprehension. Scores are reported as Lexile levels, or simply Lexiles, which range from 200 to 1,800. Because the assessment is computer-adaptive, the difficulty of the questions that are asked is automatically adjusted up or down, depending on student responses, to assess each student's reading ability more accurately. This feature reduces the potential for floor or ceiling effects that are associated with assessments that are not adaptive. The SRI uses "authentic" text passages drawn from children's and young adults' literature, classics, and periodicals. It measures students' reading comprehension skills such as identifying details, cause and effect, and event sequence; drawing conclusions; and making comparisons and generalizations.

ITBS/ITED Comprehension subtest. The ITBS Comprehension subtest is a two-part, group-administered assessment of students' ability to make inferences or generalizations about passages they have read. The passages range in length and can be of various types, including fiction, poetry, interviews, scientific information, and other non-fiction. In Des Moines, the ITBS was administered to sixth- through eighth-graders. In 2000, 2001, and 2002, some READ 180 students were assessed using a test level lower than their grade placement. Starting in 2003, all students took the test level corresponding to their assigned grade level at the time of testing. For all analyses in this report, we used scale scores, which allowed cross-level and cross-grade comparison of scores and investigation of growth over time.

The ITED Comprehension subtest is the high school correlate of the ITBS. The ITED Comprehension subtest is a group-administered assessment of high school students' ability in reference to factual understanding, inferential understanding, and analytic understanding of passages they have read. The passages range in length and can be of various types, including fiction, poetry, expository non-fiction, memoirs, essays, biographical sketches, and editorials. The ITED was administered to ninth- through eleventh-graders in Des Moines. In 2000, 2001, and 2002, some READ 180 students were assessed using a test level lower than their grade placement. Starting in 2003, students took the test level corresponding to their assigned grade level at the time of testing. For all analyses in this report, we used scale scores, which allowed cross-level and cross-grade comparison of scores and investigation of growth over time.

In addition to student achievement results on the assessment measures just discussed, the statistical analyses included six predictor measures (as listed in Exhibit 3). The measures are all taken directly from the district database or are based on other information available in the database. Three of the six predictor measures are time-varying test-point-level measures: *READ 180 status, special education status,* and *time count.* The other three are non-time-varying student-level measures (*cohort year, starting grade level,* and *participation pattern*).

READ 180 status. Student participation in READ 180 at the time of each posttest was the measure of most interest in this study because the effect associated with this measure provides an estimate of the effectiveness of the READ 180 program for this sample of students. For the pretest and the first posttest, all students were enrolled in READ 180. However, due to different participation patterns over the years, students may or may not have been participating in READ 180 at the time of subsequent posttests. For analytic purposes, we assign this variable a value of 1 for testing time points at which students had been enrolled in READ 180 for a full year and a value of 0 for testing time points at which they were not enrolled or were enrolled for less than a full year. For testing time points at which students had a value of 0 for *READ 180 status,* they served as comparison students. Thus, although all students included in the analyses had been in READ 180 for at least one year, individual students served as comparisons only for the time points at which they were no longer enrolled in READ 180. The effects associated with this variable in the analyses are the effects attributed to the READ 180 program for this sample of students, after controlling for other variables in the analysis.

Special education status. *Special education status* refers to the categories of special education services provided by the district, and it is a measure whose value could vary from one time point to the next. For the present study, at each time point, a student's special education status was measured as one of four levels: Up to 2 hours per day (referred to as Resource Level); Special class 1-3 hours per day; Self-contained most of the day; and Self-contained all day.[10] The measure of special education was operationalized as a series of four "dummy" variables, one for each category of special education, each coded 1 for students in that category and 0 for students not in the category. For analyses, one of the dummy variables must be omitted to serve as the reference category. In the final models for this study, the lowest level, Up to 2 hours per day, served as the omitted reference category.

Time count. The longitudinal growth curve analyses necessary for this study require a measure of the passage of time. For this, we created a simple counting variable, *time count*, with assigned values of 0 at a student's pretest time point, 1 at the first posttest time point, 2 at the second, and so on, up to a maximum possible of 5, representing the fifth posttest time point, for six (0-5) possible test points overall. The length of time between each posttest is approximately one school year. In the longitudinal analyses, the effect associated with this variable represents the rate of change in achievement over time, or *growth,* which represents the second source of change that growth curve modeling sorts out – the change associated with normal individual growth that occurs naturally.

The remaining three measures listed in Exhibit 3 were student-level non-time-varying measures:

Cohort year. *Cohort year* refers to the school years and cohorts discussed previously, and has five possible values: 1, 2, 3, 4, or 5. *Cohort year* is an important measure to control for in any longitudinal analysis because each cohort is subject to slightly different historical conditions, such as changes in district policy, that affect students in different grades or different cohorts in different ways. Since we are unable to control for the countless separate factors that

[10] These are the levels of special education services as defined by the State of Iowa. Iowa identifies students for special education services as "entitled individuals" rather than assigning disability labels.

contribute to the unevenness of historical conditions, controlling for cohort is the best way to even out the uneven reality of local circumstances. This is one facet of how growth curve modeling teases out the three sources of change: cohort year represents the third source of change discussed previously, that related to changes that occur in the population at large.

Starting grade level. *Starting grade level* ranged from sixth to eleventh grades. *Starting grade level* is an important factor to control for because intervention programs often have different impacts at different grade levels. In fact, previous research in Des Moines indicated that there may be differential impacts for READ 180, in particular for seventh- versus eighth-graders who are in their second year of READ 180 (Palmer, 2003a and 2003b).

Participation pattern. *Participation pattern* refers to students' histories of entering, leaving, and re-entering the READ 180 program over the years of the study. There were seven distinct patterns of participation, defined in terms of the number of years of participation and the extent to which participation was consecutive, but several of these had very few students that followed them – too few, in fact, to allow meaningful analysis. Exhibit 4 displays the seven patterns and the numbers of students for each pattern.

Exhibit 4
Seven READ 180 Participation Patterns

Participation Pattern	Participating Students	
	Number	Percent
One year only	831	68.51
Two consecutive years only	308	25.39
Three consecutive years only	53	4.37
Two non-consecutive years	14	1.15
Three non-consecutive years	2	0.16
Four consecutive years	4	0.33
Five consecutive years	1	0.08
Totals	1,213	100.00

Source: Authors' tabulations from Des Moines school district database

From Exhibit 4, we can see that the two most common participation patterns accounted for 94 percent of all students: just over two thirds of students participated in READ 180 for a single year without ever returning to the program during the five years of the study; and another quarter participated for two consecutive years only, leaving after those two years and not returning during the five-year span of the study. In addition, just over 4 percent participated for three consecutive years only. For analyses, we operationalized *participation pattern* as a series of four dummy variables: one year only, two consecutive years only, three consecutive years only, and Other (for the latter, we collapsed the four smaller categories into one catch-all category). Recall that a series of dummy variables must omit one category in analyses, to serve as a reference category. In this study, three consecutive years only served as the omitted reference category in analyses because it was the category that was most different from the others (i.e., had the most statistically significant contrasts with the other categories).

Statistical/Analytical Models

To investigate growth trends over time, we employed growth curve modeling as described above, using the statistical software HLM 6.0. We conducted a separate series of statistical analyses for each of the achievement measures available in the database, resulting in five sets of analyses.[11]

Before turning to a discussion of the growth trend results, it may be useful to review the concept of growth curve modeling and to explain how analysis models can be created. Recall that we have two levels of analysis – students and their test time points – that correspond to two types of measures – non-time-varying measures and time-varying measures, respectively. The two types of measures play different roles in analyses. Time-varying measures, that is, test-point-level measures, can only be used to predict achievement directly. For example, *special education status* at each time point directly predicts achievement at that time point, such that we can obtain an average effect associated with special education status' relationship with achievement. In contrast, non-time-varying measures, or student-level measures, can be used to predict achievement directly as in the previous example, and can also be used to predict *relationships* between the time-varying measures and achievement. For example, as we discuss below, *participation pattern* has an effect on the relationship between achievement and *time count* that is, some patterns of participation have higher average annual amounts of growth than other patterns of participation. This dual analytic role for the non-time-varying student-level measures will become more familiar as we discuss results in the next section of the report.

III. READ 180 Achievement Growth Trends

This section begins with a discussion of findings about overall trends in student achievement, and how achievement is related to two explanatory measures, *READ 180 status* and *time count*. It then examines, in order, how student achievement was associated with the other four predictor measures: *special education status*, *cohort year*, *starting grade level*, and *participation pattern.*

Overall Achievement Results

Overall student achievement results suggest that participation in READ 180 had a positive, statistically significant effect on three of the achievement measures but that it had no measurable effect on the other two achievement measures. Specifically, for each additional year that students participated in READ 180, their scores increased by approximately 30 Lexiles on the SRI, 6 scale-score points on SDRT4 Comprehension, and 5 scale-score points on SDRT4 Total – above and beyond the observed yearly growth in achievement. For ITBS/ITED Comprehension and SDRT4 Vocabulary, the relationship with years of participation in READ 180 was positive but not statistically significant and deemed "no effect."

[11] We follow standard model-building procedures to arrive at a final model for each achievement measure. This section provides a general overview of model-building and the final models. Additional information is provided in the Technical Appendix to this report (Appendix B).

Exhibit 5 displays the achievement effects of each additional year of participation in READ 180, after controlling for *cohort year*, *starting grade level*, *participation pattern* and *special education status*. On three of the five achievement measures, the effect of READ 180 was positive and statistically significant. Students gained 6.11 additional scale-score points on the SDRT4 Comprehension subtest for each year they participated in READ 180. With multiple years of additional participation, scale-score points increase in multiples of just over six scale-score points per year. Thus, the gap in achievement between students who remained in READ 180 for more than one year and those who participated for only one year increased over time, other things (*participation pattern*, *cohort year*, *starting grade level*, *special education status*) being equal. Using achievement on SDRT4 Comprehension for students who began READ 180 as seventh-graders, Exhibit 6 illustrates this pattern. In this exhibit, Time 0 represents the pretest for seventh-graders, administered in the fall just prior to their participation in READ 180. Time 1 represents their first posttest, administered the spring after their first year of participation. At Time 2, the scores begin to diverge into two groups: students who remained in READ 180 for another year have slightly higher scores than students who no longer participated in the program. Scores for the two groups continued to diverge in subsequent years such that each additional year of participation in READ 180 was associated with additional increases in the gap between the two groups' scores. After three years, the gap widened to 12 scale-score points between the two groups, 658 versus 646. Since a student's annual expected gain in SDRT4 scores was 15 scale-score points (Scholastic, 2002), this 12-point gap represented a difference equivalent to four-fifths of a year of growth. The same pattern held for the other grade levels on SDRT4 Comprehension, with the same 12 scale-score point advantage for students who continued in READ 180.

Exhibit 5
Effect of READ 180 by Achievement Measure

	SDRT4 Comprehension	SDRT4 Vocabulary	SDRT4 Total	SRI Score	ITBS/ITED Comprehension
Scale-score points associated with each add'l year of participation in READ 180	6.11 **	0.72	4.72 *	28.16 *	0.31
Years of growth	0.41	0.05	0.31	0.24	0.02
Effect size	0.16	0.01	0.12	0.11	0.01
NCEs	3.33	0.30	2.57	2.30	0.26
No. of students	1,204	1,204	1,203	1,204	1,159
No of test points	2,806	2,788	2,738	2,740	2,352

Exhibit reads: READ 180 students achieved an additional 6.11 scale-score points in SDRT Comprehension for each year of participation in READ 180, other things being equal. This figure corresponded to 0.41 years of growth and an effect size of 0.16, which in turn corresponded to a difference equal to 3.33 normal curve equivalent points.

Note: *Results are reliable at statistical significance levels of $p<0.05$; ** $p<0.01$. NCE = Normal Curve Equivalent.
Source: Authors' tabulations from Des Moines school district database.

Exhibit 6
Effects of READ 180 on Achievement, SDRT4 Comprehension, for Students Starting READ 180 as Seventh-Graders

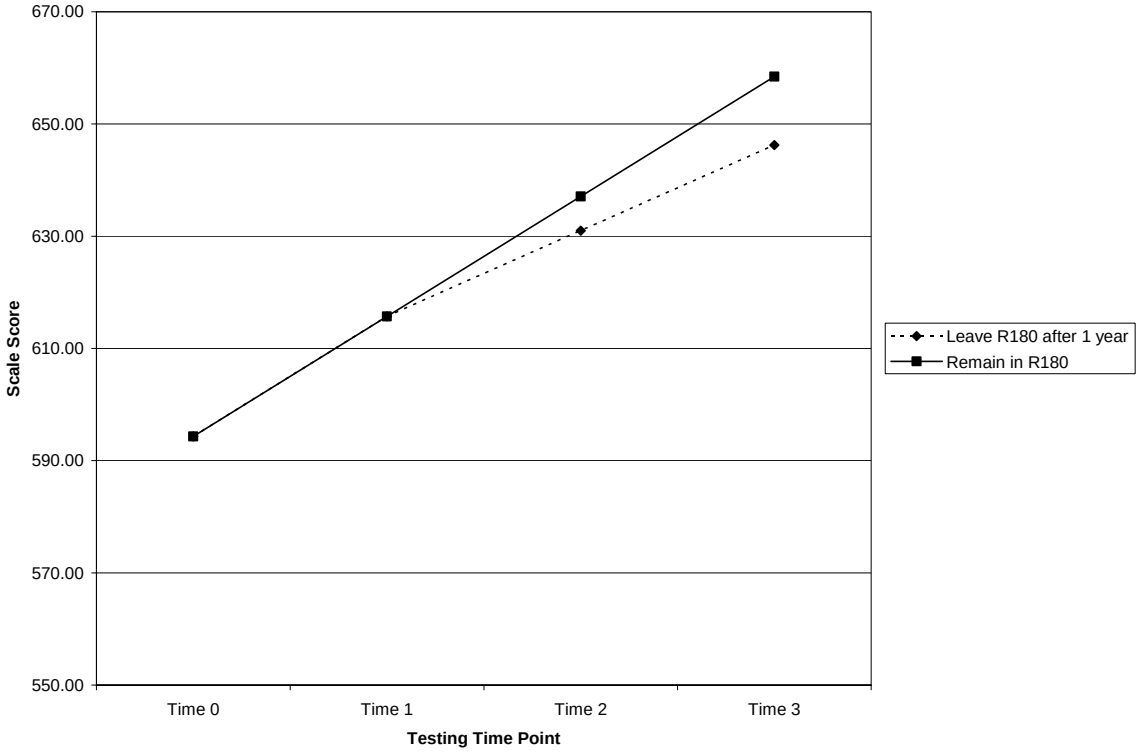

Exhibit reads: READ 180 students who remained in the program beyond the first year have higher mean scale scores on SDRT4 Comprehension than students who left the program after only one year of participation, controlling for other factors in the model, based on growth curve analysis whereby sample sizes equal 1,204 students and 2,806 test points.

Source: Authors' tabulations from Des Moines school district database.

The results for the SDRT4 Total and the SRI followed the same pattern as the SDRT4 Comprehension results, with somewhat different start and end points and rates of divergence. Student scores on SDRT4 Total diverged at a slightly lower rate than they did on the comprehension subtest, earning students an additional 4.72 scale-score points for each additional year they participated in READ 180. After three years of posttests, the gap between participants and students who were no longer participating was 9.4 points, representing a READ 180 advantage of 0.63 years of growth. On the SRI, each additional year of READ 180 participation was associated with 28.16 additional Lexiles. Although this appeared much higher than the SDRT4 results, the expected annual gain was also much higher (116 versus 15), so it translated into the same range of advantage: after three years of posttests, READ 180 students' advantage on SRI was 56.32 Lexiles, or approximately half a year of growth.

It is important to bear in mind that these increases in READ 180 students' achievement occurred *on top* of other growth. So, while the equivalent of a half year of advantage over two years may seem small, this is growth *in addition to* other growth students experienced. Furthermore, in one sense, these results are conservative estimates of the effect of READ 180

because READ 180 students are being compared to students who have had at least one year of READ 180. Comparing READ 180 students to students who never enrolled in READ 180 but who otherwise were similar might yield even larger effects of participation in READ 180.

In a second sense, though, some of the results are not conservative. The students who served as comparison students in the SDRT4 and SRI analyses were also more mobile by definition, and therefore more disadvantaged than the READ 180 students in those analyses.[12] This difference exists because only students who had spent a full year or a portion of a year in READ 180 were administered the SDRT4 and SRI tests each year, with full-year students designated by the district as READ 180 participants, and less-than-full-year as comparison students. However, the fact of receiving only a portion of a year of READ 180 indicates that the student had changed schools, possibly to a school that could not accommodate the mobile student into their READ 180 program. That is, for SDRT4 and SRI analyses, we know with certainty that the less-than-full-year-READ 180 students (the comparison students) were students who experienced mobility during the school year. For the full-year READ 180 students, we do not have enough information to know the extent of mobility – it is possible that some full-year READ 180 students were also mobile, but that the school they moved into was able to accommodate them into their READ 180 program and thereby provide the full year of participation.

It is thus possible, and even likely, that the comparison students (less-than-full-year participation in READ 180) for the SDRT4 and SRI analyses were more disadvantaged than the READ 180 students. Since we do not have the same information for full-year READ 180 students, we cannot control for the effects of mobility. This means that any effects observed for the SDRT4 and SRI tests may be larger than we would observe if we could control for mobility. Despite these limitations, having a flawed comparison group is better than having none at all, especially since we *are* able to control for other important measures, including pretest score. Nevertheless, interpretation of these results must reflect these caveats.

Another way to look at the results is to examine their effect size. Effect size is a measure of the size of the association between student outcomes and student participation in an intervention program (versus not participating). In this study, effect size can be calculated by dividing the effect of READ 180 by the standard deviation of the relevant outcome measure (Lipsey & Wilson, 2001). Thus, for the SDRT4 Comprehension analyses, the effect of READ 180, 6.11, was divided by the standard deviation on SDRT4 Comprehension, 38.66, to obtain an effect size of 0.16. This figure indicates that the difference between READ 180 and non-READ 180 for each year of participation in READ 180 is 0.16 of a standard deviation. The corresponding effect size for SDRT4 Total is 0.12; and for SRI, 0.11. The effects for SDRT4 Vocabulary and ITBS/ITED Comprehension are not statistically significant.

In addition to the effects associated with READ 180, an aspect of the overall results that deserves attention is the relationship between achievement growth and time. The results confirmed a linear growth trend such that a student's achievement score increased an average amount each year. This held true for all five achievement measures and was statistically

[12] This caution does not apply to the ITBS/ITED tests, which are administered district-wide to all students, not just READ 180 students.

significant in all five analyses. For SDRT4 Comprehension, the average number of scale-score points associated with each year's increment of achievement growth was 15.27; for SDRT4 Vocabulary the increment was 17.21 points per year increase; for SDRT4 Total, 20.59; for SRI, 74.01; and for ITBS/ITED Comprehension, 4.58.

When we combine the findings associated with participation in READ 180 with the findings associated with achievement growth over time, we gain an understanding of the annual achievement difference between READ 180 students and non-READ 180 students. For SDRT4 Comprehension, for example, all students gain about 15 scale-score points per year; and READ 180 students gain an additional 6 scale-score points per year, for a total of 21. Thus, non-READ 180 students on average gain 15 scale-score points per year on SDRT4 Comprehension while READ 180 students on average gain 21 scale-score points per year. Dividing by the expected annual growth (15, as discussed above) translates these scale-score point gains into years of growth: 1.02 and 1.43, respectively. Exhibit 7 displays these results as described for SDRT4 Comprehension. Exhibit 8 and Exhibit 9 present the results for SDRT4 Total and for the SRI, respectively.[13]

[13] Results for SDRT4 Vocabulary and for ITBS/ITED are not presented because the differences between READ 180 and non-READ 180 students were not statistically significant on those two measures. Details are available in the Technical Appendix.

Exhibit 7
Annual Achievement Growth on SDRT4 Comprehension: READ 180 Students versus Comparison Students

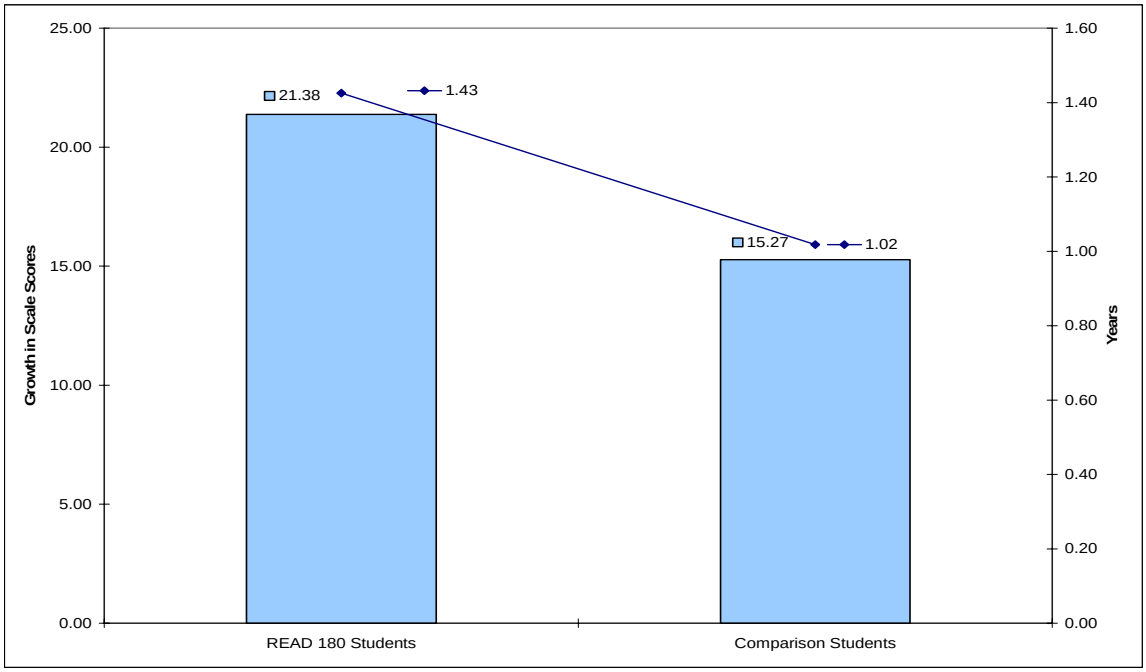

Exhibit reads: READ 180 students showed an annual growth of 21.38 scale-score points on SDRT4 Comprehension, which corresponds to 1.43 years of growth. In contrast, comparison students showed an annual growth of 15.27 scale-score points on the same measure, corresponding to 1.02 years of growth. The difference between the annual increases in scale scores achieved by the two groups of students is statistically significant at $p<0.05$.

Exhibit 8
Annual Achievement Growth on SDRT4 Total:
READ 180 Students versus Comparison Students

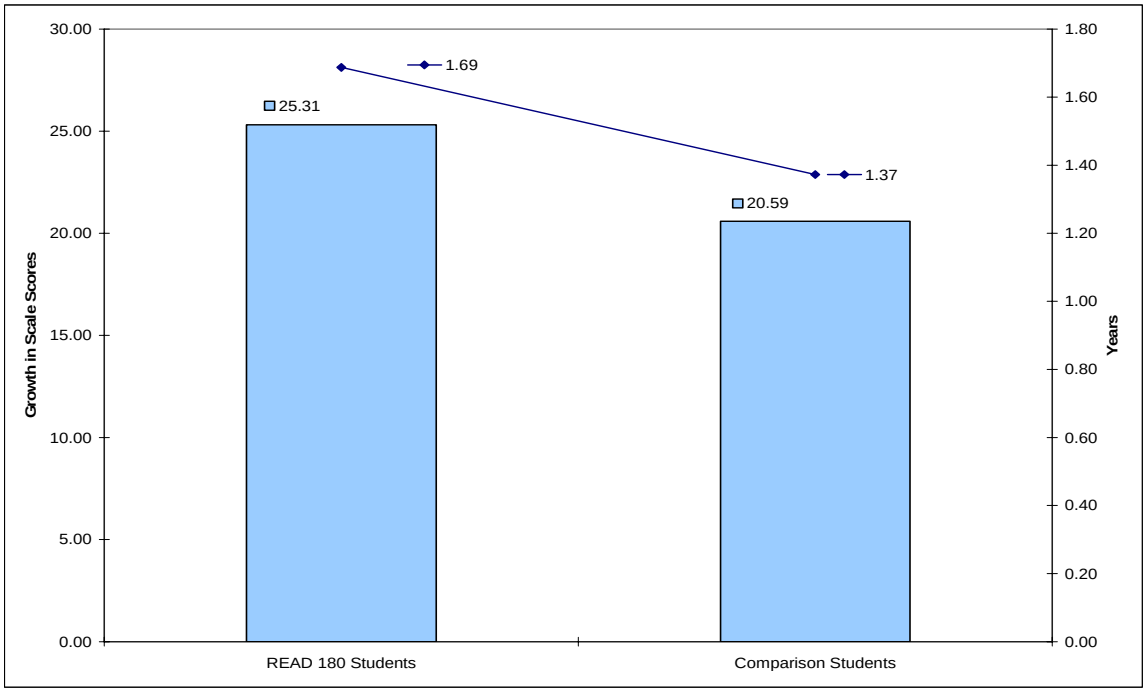

Exhibit reads: READ 180 students showed an annual growth of 25.31 scale-score points on SDRT4 Total, which corresponds to 1.69 years of growth. In contrast, comparison students showed an annual growth of 20.59 scale-score points on the same measure, corresponding to 1.37 years of growth. The difference between the annual increases in scale scores achieved by the two groups of students is statistically significant at $p<0.05$.

Exhibit 9
Annual Achievement Growth on the SRI:
READ 180 Students versus Comparison Students

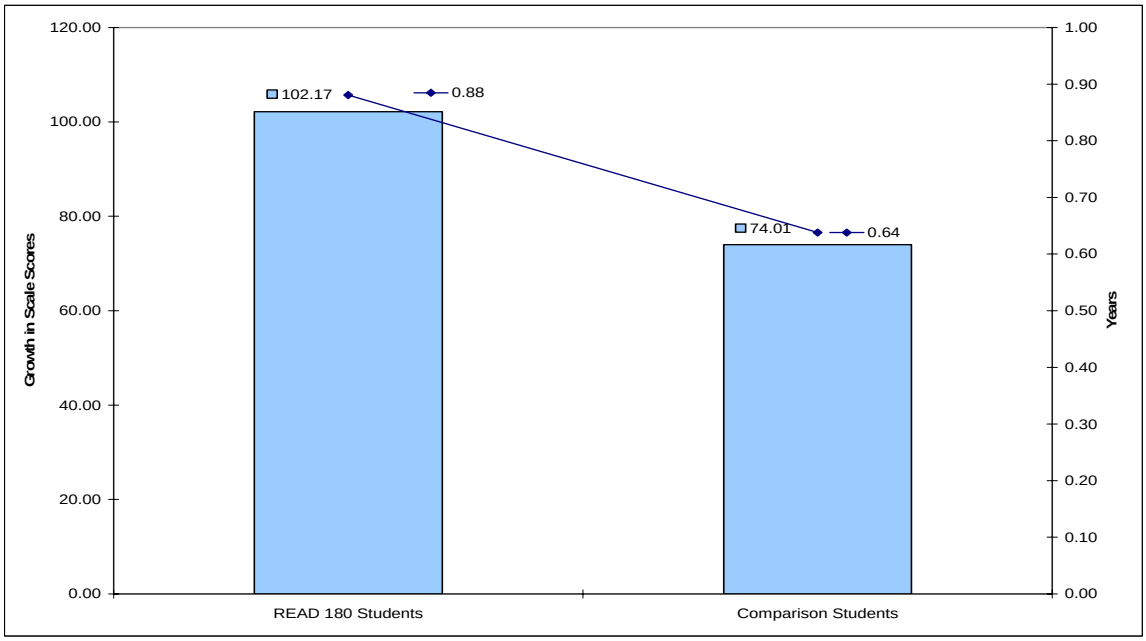

Exhibit reads: READ 180 students showed an annual growth of 102.17 scale-score points on the SRI, which corresponds to 0.88 years of growth. In contrast, comparison students showed an annual growth of 74.01 scale-score points on the same measure, corresponding to 0.64 years of growth. The difference between the annual increases in scale scores achieved by the two groups of students is statistically significant at $p<0.05$.

Each series of achievement analyses also included a subset of the four other predictor measures described earlier – not all predictor measures were used in all models. Preliminary models were run testing for statistically significant effects of the predictor measures on achievement. Where effects on achievement or on growth were statistically significant, the measures were included in final models; where not, they were omitted. The following sections discuss the results of the applications of the final models to the data.

The Relationship between Achievement and *Special Education Status*

Special education status, which refers to one of four levels of special education services, is a time-varying measure that predicts achievement directly. *Special education status* at each time point can be used to predict achievement at that time point. We thereby obtain an average effect associated with special education status' relationship with achievement.

The results indicate that, at any given time point, students receiving higher levels of special education services had lower achievement scores than did students receiving the lowest level of special education services. Exhibit 10 presents these results. Each number in Exhibit 10

Exhibit 10
Differences in Mean Achievement Scores among Students, by *Special Education Status*

Special Education Status	SDRT4 Comprehension	SDRT4 Vocabulary	SDRT4 Total	SRI Score	ITBS/ITED Comprehension
Resource level, up to 2 hours per day, is the reference category.					
1-3 hours per day	-10.37 ***	-10.41 **	-10.44 ***	-39.83 *	-6.60 ***
(Years of achievement)	(0.69)	(0.69)	(0.70)	(0.34)	(0.44)
Most but not all of the day	-15.13 ***	-11.00 ***	-12.95 ***	-64.56 ***	-9.74 ***
(Years of achievement)	(1.01)	(0.73)	(0.86)	(0.56)	(0.65)
All day	-12.01 ***	-5.64	-9.48 **	-17.83	-4.53 *
(Years of achievement)	(0.80)	(0.38)	(0.63)	(0.15)	(0.30)
No. of students	1,204	1,204	1,203	1,204	1,159
No. of test points	2,806	2,788	2,738	2,740	2,352

Exhibit reads: Students receiving 1-3 hours per day of special education services have a mean achievement score 10.37 scale-score points lower than students receiving the resource level of special education services, at any given test point, controlling for other factors in the model. This difference is statistically significant at the $p < 0.001$ level.

Note: Results are presented relative to the resource level of special education services, up to 2 hours per day. Results are reliable at statistical significance levels of p<0.05; ** p<0.01; *** p<0.001. Numbers in parentheses are the years of achievement represented by the scale-score points, based on expected annual gain of 15 scale-score points on the SDRT4 and ITBS/ITED and 116 on SRI.

Source: Authors' tabulations from Des Moines school district database

depicts the difference, in scale-score points, between the mean achievement score for students receiving resource level special education services[14] and students receiving the special education services listed in the far left column. Negative numbers (all are negative) indicate a mean achievement for these students lower than that for the students receiving resource level services; positive numbers, if there were any, would reflect a mean achievement higher than that for students receiving resource level services.

The scale-score effects of *special education status* listed in Exhibit 10 can be translated into estimates of years of achievement. Recall that the expected annual growth on SDRT4 and on ITBS/ITED is 15 points. Dividing the effect of *special education status* by 15 thereby produces a proportion of a year of achievement represented by the scale-score points related to level of special education. For example, students receiving 1-3 hours per day of special education services have achievement lower than students receiving up to 2 hours per day, by 10.37 scale-score points on SDRT4 Comprehension. This difference represents seven-tenths of a year of achievement (10.37/15=0.69).

The Relationship between Achievement and *Cohort Year*

Cohort year refers to the first year that a student participated in READ 180. Since *cohort year* is a non-time-varying student-level measure, it can have two analytic functions: predicting achievement directly and predicting relationships of achievement with the time-varying measures. Of special interest are effects on the relationship between achievement and the *time count* measure. An effect on this relationship is equivalent to an effect on the rate of growth that occurs in year-to-year achievement scores. Preliminary analyses were conducted to investigate these two possible roles for *cohort year*. Where effects were found to be statistically significant in preliminary analysis, that effect was included in final models; where not statistically significant, it was omitted.

Findings from the final models indicate that the effects of *cohort year* were rather limited in both potential roles. Only for SDRT4 Total Reading and for SRI Lexile scores was *cohort year* associated directly with achievement, and only for SDRT4 Total Reading was *cohort year* associated with rate of achievement growth. Specifically: For SDRT4 Total, Cohort 2 had a baseline achievement 6 points higher than Cohort 1, representing a 0.40 years of achievement (6/15=0.40). Also, Cohort 4 had a higher rate of achievement growth than did Cohort 1, by 6 points per year (0.40 years); and Cohort 5 had a lower rate of growth than did Cohort 1, by 9 points per year (0.60 years). For SRI scores, Cohort 2 and Cohort 3 each had starting pretest scores higher than did Cohort 1, by 57 and 59 points respectively (0.49 and 0.51 years, respectively). These differences, as stated, are statistically significant.

The Relationship between Achievement and *Starting Grade Level*

As a non-time-varying measure, *starting grade level*, like *cohort year,* can function in two ways: predicting achievement directly and predicting relationships of achievement with the

[14] And a handful of students whose level of service was coded as No-Service.

time-varying measures. As with *cohort year*, preliminary analyses were run to test for effects of *starting grade level*. Where such effects were found to be statistically significant, the measure was included in final models. Here we report on the results of the final models.

The findings suggest that *starting grade level* is directly related to achievement; this was expected and *starting grade* level therefore serves as an important control measure in all five achievement analyses. In addition, in the analyses for SDRT4 Vocabulary and SDRT4 Total, *starting grade level* was related to the rate of growth that occurs in year-to-year achievement scores, even after controlling for the different baseline achievement associated with starting grade level. Exhibit 11 and Exhibit 12 illustrate the relationship between *starting grade level* and rate of achievement growth. Exhibit 11 presents the results in terms of scale-score points

Exhibit 11
The Relationship between Achievement Growth and *Starting Grade Level:*
Scale-Score Points

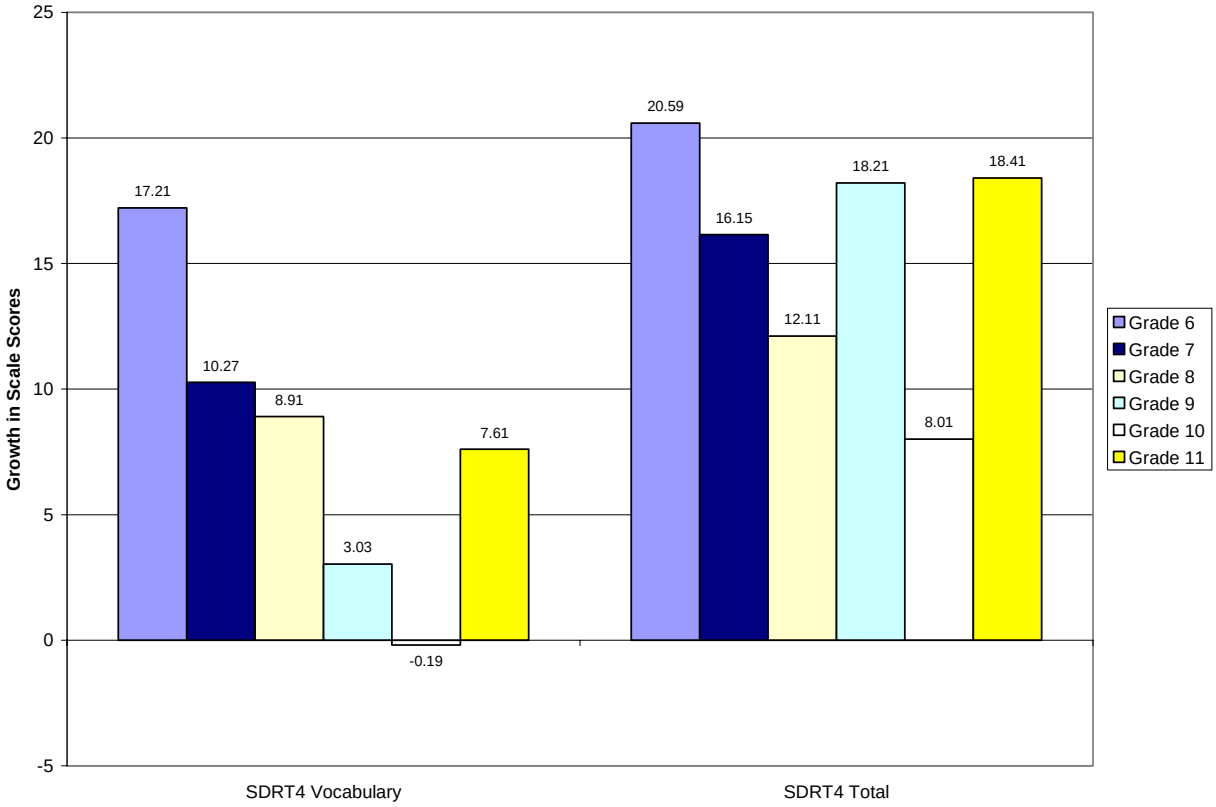

Exhibit reads: Students who start READ 180 in sixth grade gain 17.21 scale-score points per year on SDRT4 Vocabulary, controlling for other factors in the model.

Note: Other achievement measures are omitted from this exhibit because starting grade level was not found to have statistically significant effects on growth for those measures in preliminary models.

Source: Authors' tabulations from Des Moines school district database

while Exhibit 12 presents results translated into years of achievement associated with each scale-score result.

Specifically, for SDRT4 Vocabulary and SDRT4 Total, students who enrolled in READ 180 in the seventh grade or later demonstrated slower achievement growth than did students who began as sixth-graders, as indicated by SDRT4 Vocabulary and SDRT4 Total scale scores. This effect was statistically significant for seventh-, eighth-, and tenth-graders, relative to sixth-graders, for SDRT4 Vocabulary and SDRT4 Total.

The effects of *starting grade level* can be translated into estimates of years of achievement, based on the expected annual growth on SDRT4 of 15 points. Dividing the growth

Exhibit 12
The Relationship between Achievement Growth and *Starting Grade Level:*
Years of Achievement

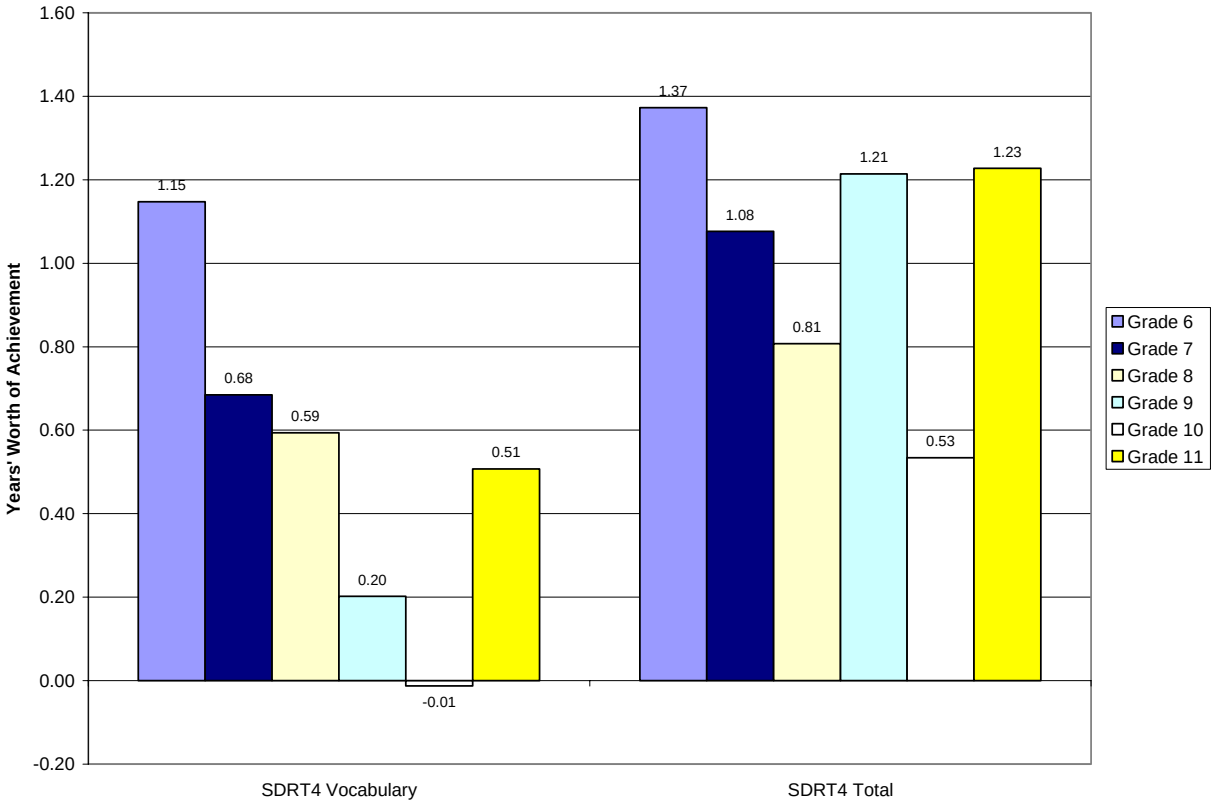

Exhibit reads: Students who start READ 180 in sixth grade gain 1.15 years of expected growth per year on SDRT4 Vocabulary, controlling for other factors in the model.

Note: Other achievement measures are omitted from this exhibit because starting grade level was not found to have statistically significant effects for those measures.

Source: Authors' tabulations from Des Moines school district database

associated with each starting grade level by the expected growth produces the portion of a year of achievement represented by the effects related to *starting grade level*. The corresponding figures are presented in Exhibit 12. (Note that Exhibit 11 and Exhibit 12 present the same results but in different metrics.)

The Relationship between Achievement and READ 180 *Participation Pattern*

As a non-time-varying measure, *participation pattern*, like *starting grade level* and *cohort year,* can function in two ways: predicting achievement directly and predicting relationships of achievement with the time-varying measures. Preliminary analyses were run to test for effects of *participation pattern.* Where such effects were found to be statistically significant, the measure was included in final models. Here we report on the results of the final models.

In its first role, *participation pattern* had a statistically significant relationship with achievement only for SDRT4 Comprehension and SDRT4 Total. In particular, the One Year Only pattern was associated with a higher baseline achievement than was the Three Consecutive Years pattern, by 16 points for SDRT4 Comprehension and 7 points for SDRT4 Total. The Other participation pattern was associated with a higher baseline achievement than was the Three Consecutive Years pattern for SDRT4 Comprehension, by 17 points.

The second role of *participation pattern* in analyses was as a predictor of the rate of growth that occurs in year-to-year achievement scores. (Recall that being a predictor of the rate of growth means the predictor had an impact on the relationship between achievement and *time count*.) For all five achievement measures, *participation pattern* was a statistically significant predictor of rate of year-to-year growth in achievement. Exhibit 13 and Exhibit 14 illustrate the relationship between *participation pattern* and rate of achievement growth. Exhibit 13 presents the results in terms of scale-score points while Exhibit 14 presents results translated into years of achievement associated with each scale score result.

Exhibit 13 displays the extent to which a student's overall *participation pattern* is related to his or her rate of growth in achievement. The main finding was that the Two Consecutive Years participation pattern had the highest rates of achievement growth, and students in the Other category participation pattern had the lowest rates of achievement growth. The One Year Only and the Three Consecutive Years participation patterns fell in between, and the differences between these two in growth rates were rarely statistically significant.

In particular, we found that gains over time are greater for students in the Two Consecutive Years participation pattern than for students in the Three Consecutive Years participation pattern on SDRT4 Vocabulary, SDRT4 Total, SRI score, and ITBS Comprehension. The Two Consecutive Years participation pattern students also had gains over time greater than gains for the Other participation pattern students on SDRT4 Comprehension, SDRT4 Total, and SRI scores. These differences were all statistically significant.

Exhibit 13
Relationship between Achievement Growth and *Participation Pattern*

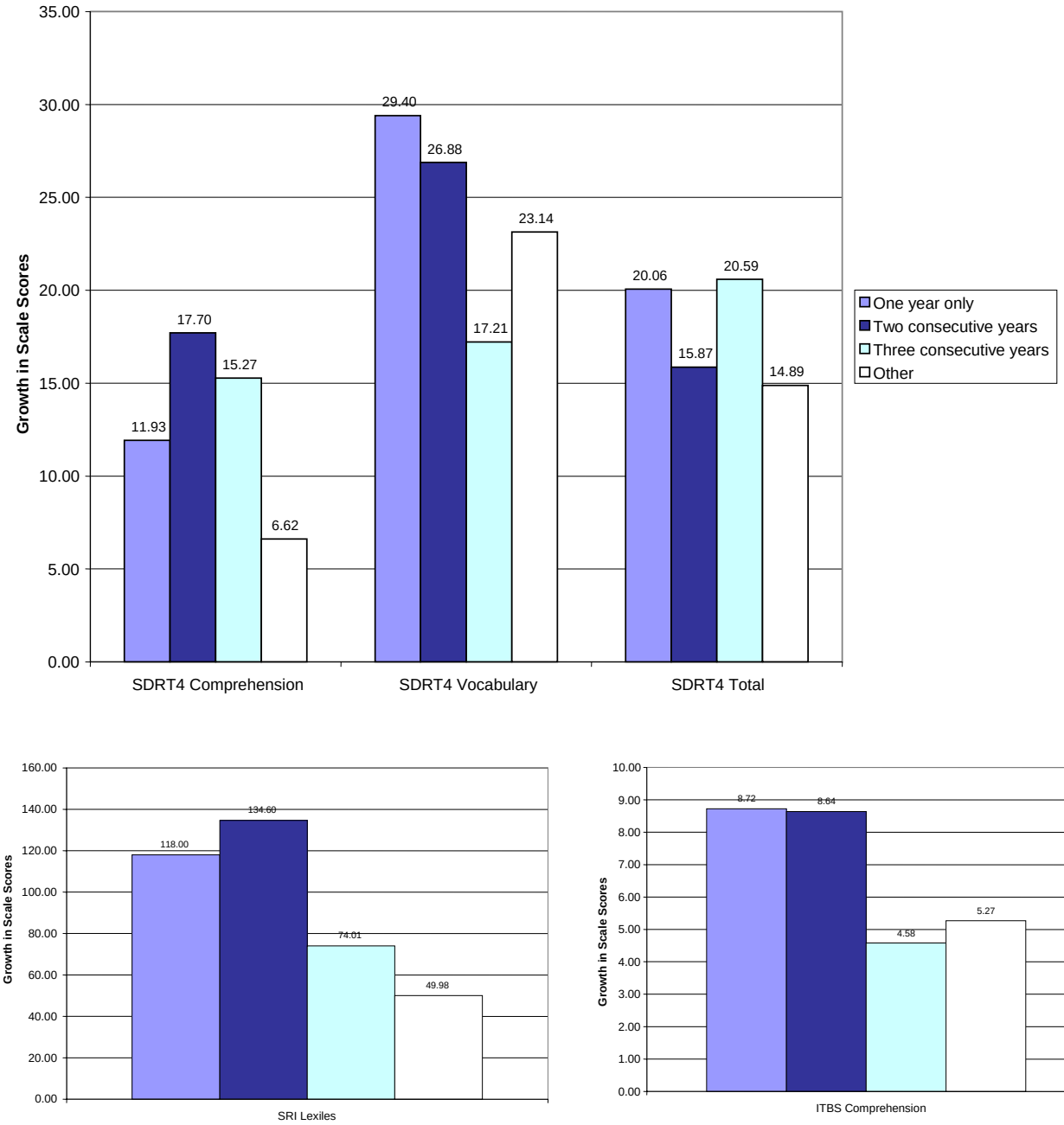

Exhibit reads: Students whose five-year participation pattern was one year only gained 11.93 scale-score points per year on SDRT4 Comprehension. This is less than the 17.70 points per year gained by students in the two consecutive years pattern, controlling for other factors in the model, and the difference is statistically significant based on growth curve analysis where sample sizes equal 1,204 students and 2,806 test points.

Source: Authors' tabulations from Des Moines school district database

Exhibit 14
Relationship between Achievement Growth and *Participation Pattern*

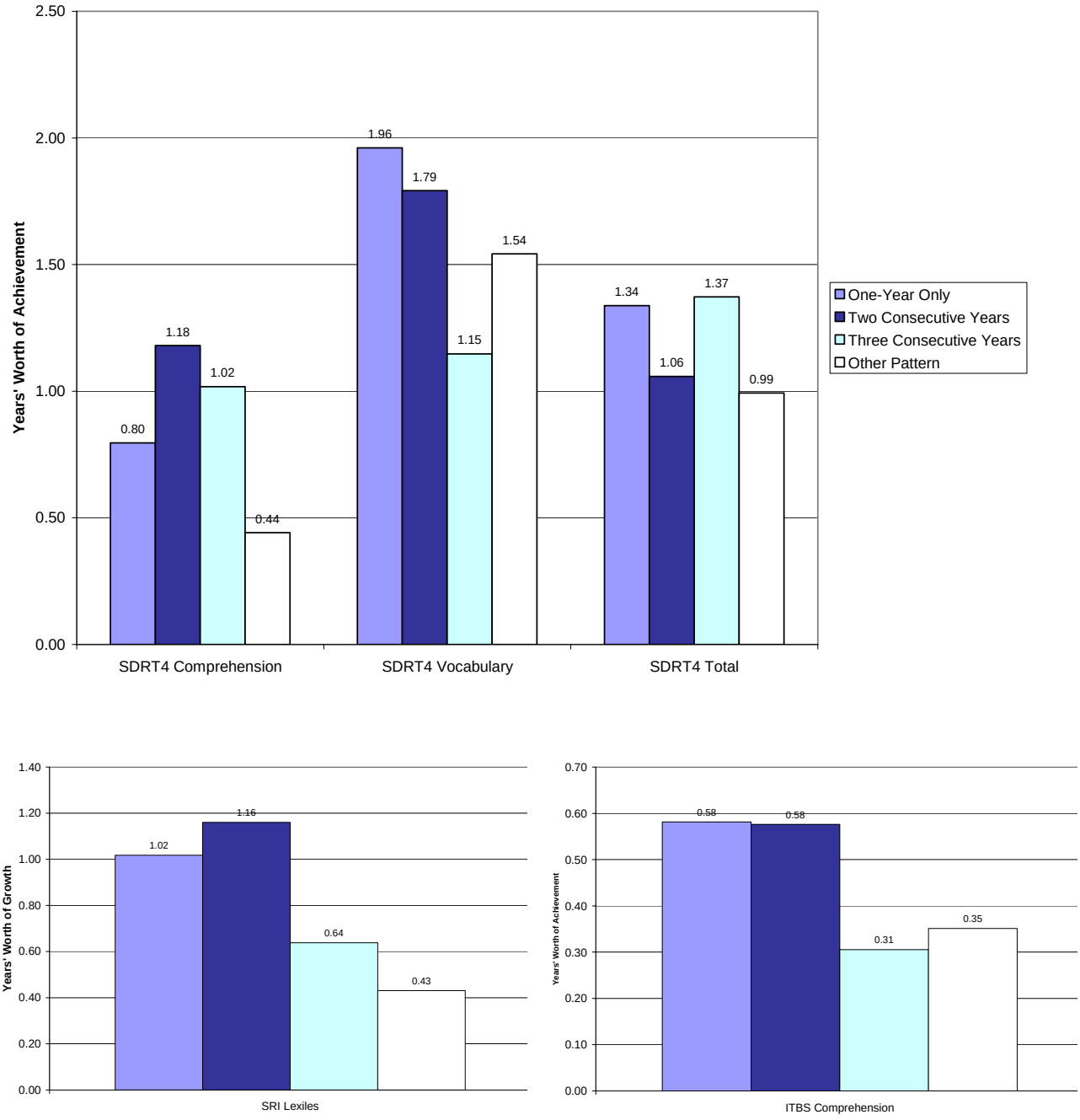

Exhibit reads: Students whose five-year participation pattern was one year only gained 0.80 years of growth per year on SDRT4 Comprehension. This is less than the 1.18 years of growth per year gained by students in the two consecutive years pattern, controlling for other factors in the model. This difference is statistically significant based on growth curve analysis where sample sizes equal 1,204 students and 2,806 test points.

Source: Authors' tabulations from Des Moines school district database

Other results were more mixed. Students in the two consecutive years pattern had gains over time greater than those for students in the one year only pattern on SDRT4 Comprehension and on SRI scores. Students in the one year only pattern had gains over time greater than those for students in the three consecutive years pattern on SDRT4 Vocabulary and on SRI scores, and greater than those for students in the other pattern on SRI scores. On SDRT4 Comprehension, the three consecutive years pattern students had higher rates of growth than did students in the other participation pattern.

The effects of *participation pattern* can be translated into estimates of years of achievement, based on the expected annual growth on SDRT4 and on ITBS/ITED of 15 points, and 116 for SRI. Dividing the growth associated with each participation pattern by the expected growth produces the portion of a year of achievement represented by the effects related to participation pattern. The corresponding figures are presented in Exhibit 14. (Note that Exhibit 13 and Exhibit 14 present the same results but in different metrics.)

IV. Discussion

The central finding of this study of READ 180 in the Des Moines Independent Community School District is that Scholastic's READ 180 had a positive, statistically significant effect on three reading achievement measures, and no measurable effect on two others. Specifically, each additional year of participation in READ 180 contributed approximately 30 additional Lexiles to students' SRI scores, 6 additional scale-score points on SDRT4 Comprehension, and 5 additional scale-score points on SDRT4 Total. These represent annual gains associated with READ180 *above and beyond* other yearly growth in achievement.

For practitioners and policymakers, these gains are perhaps best understood in the context of findings from other research. This section of the report locates the findings from this study in this larger context and suggests some directions for future research.

A Context for Understanding the Observed Effects

In Des Moines in the years covered in this study, the effect sizes associated with special education students' participation in READ 180 are comparable to effect sizes found for other well-known interventions. Exhibit 15 displays the effect sizes for this study alongside mean effect sizes in other intervention research. As the data in this exhibit show, the effect size on SDRT4 Comprehension was 0.16. This represents the difference between READ 180 students' test scores and test scores of non-READ 180 students for each year of the program. The corresponding effect size for SDRT4 Total was 0.12, and for SRI, 0.11.

The modest effect sizes reported here, ranging from .11 to.16, are in good company. Cohen (1988) suggested a commonly cited benchmark for interpreting effect size whereby effect sizes of 0.20 can be considered small, 0.50 moderate, and 0.80 large. However, these guidelines seem inflated when considering that a thorough review of meta-analysis studies concluded that psychological, educational, and behavioral treatment effects of modest values of even 0.10 to

0.20 should not be interpreted as trivial (Lipsey & Wilson, 1993). For example, a synthesis of federal evaluation results of Title I programs found an average effect size of 0.11 (Borman & D'Agostino, 1996). Similarly, a meta-analysis of the research on 29 of the most common comprehensive school reform (CSR) models found an average effect size of 0.15, which was reduced to 0.12 when only two-group comparison studies were considered, and further reduced to 0.09 when only third-party evaluations were included (Borman, Hewes, Overman, & Brown, 2003). Furthermore, considering that READ 180 is a daily 90-minute program, while the others cited here include interventions targeting the entire school day and school year, READ 180's effect sizes were considerable.

Several other considerations can help interpret the magnitude of the effect sizes associated with participation in READ 180 in Des Moines. First, one of the interventions included in Exhibit 15, comprehensive school reform (CSR), is, by definition, an intervention

Exhibit 15
Effect Sizes for READ 180 in Des Moines compared to Effect Sizes for Two Well-Known Interventions

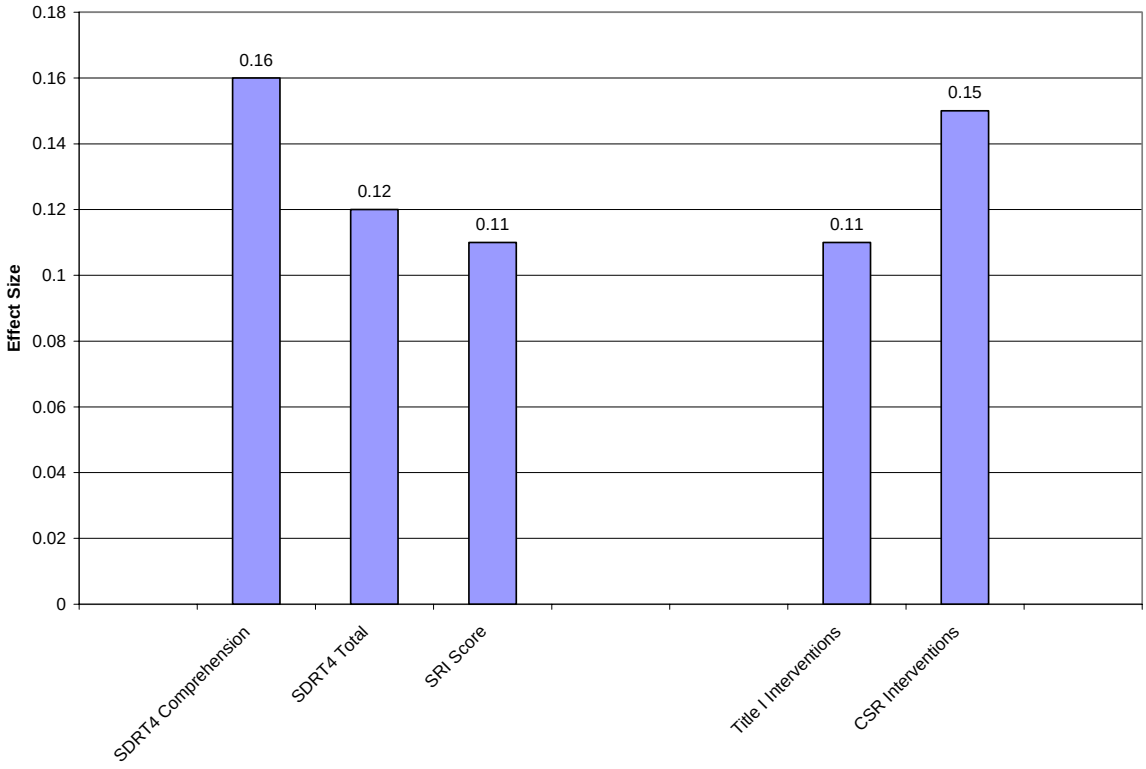

Exhibit reads: The effect size observed for SDRT4 Comprehension in Des Moines was 0.16.

Sources: Authors' tabulations from Des Moines school district database, for SDRT4 and SRI effect sizes; Borman and D'Agostino (1996) for Title I effect size; and Borman, Hewes, Overman, and Brown (2003) for CSR effect size.

that encompasses the entire school day and school year. In contrast, READ 180 is a daily 90-minute program. The fact that READ 180, as a much less intensive intervention, showed effect sizes in the same range as a broadly based extensive intervention is, by itself, very encouraging.

Second, from one perspective, the results of this study represent conservative estimates of the effect of READ 180. Note that the comparison against which READ 180 was judged to have had an effect involved students who had previously had some exposure to the READ 180 program (at least one year). Comparing READ 180 students to students who were never enrolled in READ 180 but who were otherwise similar might yield even larger effects of participation in READ 180. Alternatively, as discussed above, it is possible that the effects of READ 180 are overestimated here because it is likely that the sample of comparison students are more mobile, and therefore more disadvantaged, than READ 180 students.

In addition to reviewing these results in the context of findings about other interventions, it is also useful to view them in the context of other research with special education students. For example, after a comprehensive review of the literature on interventions with learning disabled students, Swanson (1999) and Swanson & Hoskyn (1998) identified 16 studies that met a set of criteria for methodological rigor and that reported on norm-referenced reading comprehension outcomes. Combined, these studies reported on 38 effect sizes that had a mean of 0.45.

The drawback to comparing the effect sizes of the present study to effect sizes of other interventions with special education students is that there are important differences between the current study and studies available in the research literature. The differences are generally of two types: those associated with the students receiving the intervention, and those associated with the research design. One critical feature that spans the two types of difference relates to the starting ability levels of students in the studies. Many of the intervention studies in the literature surveyed by Swanson required that students in their samples show a discrepancy between their IQ scores and their reading scores – that is, students had to be of at least average intelligence, but have reading scores lower than would be predicted from their general ability level. There was no such requirement in the Des Moines research. The reason this difference is critical is that the Swanson reviews found that the use of this strict selection criterion (IQ-reading ability discrepancy) is associated with finding higher intervention effects. A study such as the present one that included all students, not just those with a specific IQ-reading ability discrepancy, would find effect sizes lower than studies that did, *and* lower than all studies on average.

For the reason just explained, the effect sizes reported in the literature on interventions with special education students are not strictly comparable to the effects found for READ 180 in Des Moines. A much better benchmark would be the Title I and CSR results discussed earlier. Title I and CSR interventions tend to serve students regardless of any specific IQ-reading ability discrepancy. Furthermore, the research included in the reviews of Title I and CSR interventions represents a broader spectrum of research designs than do the reviews of special education interventions, another difference that makes the comparison of the present research to Title I and CSR effects more applicable than a comparison to special education research as currently conceived.

A final note is in order. Prior research on READ 180 in Des Moines indicated that the program accelerates the learning of special education middle school students, relative to the rate at which the students had been learning prior to enrolling in READ 180. The current findings that participation in READ 180 has positive, statistically significant effects on three different achievement measures are compatible with those earlier results. However, while this improvement was clear and consistent, it should be noted that the program did not produce effects strong enough to bring many of the participants up to grade level during the period of this study. This finding – an unfortunate one – is another feature that the present study shared with other research on interventions with special education students or other low-achieving readers (Calhoon, 2005; Hanushek, Kain, & Rivkin, 1998) as well as the literature on Title I and CSR interventions.

Suggestions for Further Research

We found that the correlations among achievement measures observed for this sample of students are unusually low (see Appendix A for details). One explanation for the low correlations may be that the study sample was composed entirely of special education students, for whom some test experts believe standardized tests such as those used in Des Moines are not appropriate.

The suitability (or not) of standardized tests for special education populations leaves school districts in a conundrum. Under the *No Child Left Behind Act*, all students, including special education students, must be held to the same high standard. This means that if a district has chosen the ITBS, for example, there must be evidence of improvement for all students, including special education students. Unfortunately, many of the assessments that are appropriate for special education students are not appropriate for regular education students, for the opposite reason: regular education students would score too high in the distribution of possible scores for those tests to serve as reliable measurements. The few tests that would be appropriate for both populations are often administered individually, rather than group-administered, and are therefore prohibitively expensive and time-consuming for most school districts. This is a serious issue for which no practical solution seems imminent, but which represents an opportunity for developers and educational researchers interested in special education students to investigate plausible solutions. There is a large untapped market for the development of less-expensive tests that can be used for the full range of students' achievement levels.

Another fruitful avenue of research would be to link teacher-level data on the fidelity of implementation of READ 180 to student achievement. An important caveat to this suggestion is that the definition of full implementation should include instructional quality defined in the context of the program model as well as the presence of appropriate hardware and software and prescribed time allocations (i.e., a daily 90-minute instructional period that includes 20 minutes of whole group instruction, three 20 minute rotations for small group instruction, modeled and independent reading, and work with the READ 180 reading software, and a 10 minute wrap-up session). In addition, as is the case in Des Moines, the local data system should have the capacity to link data on teachers to data on their students. Finally, it is necessary that the number

of READ 180 teachers to be sufficiently large to support an empirical study of the association between implementation and student outcomes. The scope of Des Moines' use of READ 180 meets this requirement.

References

Borman, G., & D'Agostino, J. (1996). Title I and student achievement: A meta-analysis of federal evaluation results. *Educational Evaluation and Policy Analysis*, 18:4, pp. 309-326.

Borman, G., Hewes, G., Overman, L., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, 73:2, pp. 125-230.

Calhoon, M.B. (2005). Effects of a peer-mediated phonological skill and reading comprehension program on reading skill acquisition for middle school students with reading disabilities. *Journal of Learning Disabilities*, 38:5, pp. 424-433.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Hanushek, E.A., Kain, J.E., & Rivkin, S.G. (1998). *Does special education raise academic achievement for students with disabilities?* (Working Paper No. 6690). Cambridge, MA: National Bureau of Economic Research.

Hewes, G.M., & Haslam, M.B. (2004). *Analysis of Scholastic READ 180 in Des Moines.* Unpublished technical report prepared for Scholastic, Inc. Washington, DC: Policy Studies Associates, Inc.

Lipsey, M.W., & Wilson, D.B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, 48, pp. 1181-1209.

Lipsey, M.W., & Wilson, D.B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

Palmer, N. (2003a). *An Evaluation of READ 180 with Special Education Students.* (Alternate title: *READ 180 Middle-School Study, Des Moines, Iowa 2000-2002*) (Effectiveness Report, Item #432595). New York: Scholastic Inc.

Palmer, N. (2003b). *READ 180 Report: 2002-2003.* Des Moines: Des Moines Independent Community School District.

Palmer, N. (2004). *READ 180 Report: 2003-2004.* Des Moines: Des Moines Independent Community School District.

Palmer, N. (forthcoming). *READ 180 Report: 2004-2005.* Des Moines: Des Moines Independent Community School District.

Scholastic/Council of Great City Schools. (2002). *READ 180 Validation Study* (Item #418637). New York: Scholastic Inc.

Swanson, H.L. (1999). Reading Research for Students with LD: A Meta-Analysis of Intervention Outcomes. *Journal of Learning Disabilities,* 32:6, pp.504-532.

Swanson, H.L., & Hoskyn, M. (1998). Experimental Intervention Research on Students with Learning Disabilities: A Meta-Analysis of Treatment Outcomes. *Review of Educational Research,* 68:3, pp. 277-321.

Appendix A
Descriptive Statistics

Means

Group means for each of the five achievement measures at each time point are presented in Exhibit A-1. Standard deviations for each achievement measure are presented as single figures, each representing the variation across all students and all test points.

Exhibit A-1
Group Means and Standard Deviations by Achievement Measure

***Achievement Measure* / Starting Grade Level**	**Pretest**	**First Posttest**	**Second Posttest**	**Third Posttest**	**Fourth Posttest**	**Fifth Posttest**
SDRT4 Comprehension (SD=38.66)						
Sixth-Graders	588.01	609.52	626.00	630.49	637.00	647.00
	548	551	230	53	4	1
Seventh-Graders	598.76	616.59	628.05	633.61	635.00	556.00
	330	339	126	18	6	1
Eighth-Graders	608.43	624.19	630.00	612.67	586.50	
	230	231	19	9	2	
Ninth-Graders	613.63	629.31	611.00			
	16	16	1			
Tenth-Graders	630.90	648.06	643.00			
	31	31	5			
Eleventh-Graders	624.13	650.19				
	16	16				
SDRT4 Vocabulary (SD=51.04)						
Sixth-Graders	578.71	606.51	624.67	630.23	642.00	634.50
	546	546	228	53	5	2
Seventh-Graders	594.79	614.61	630.99	648.39	663.17	540.00
	331	335	125	18	6	1
Eighth-Graders	609.39	634.08	630.47	627.44	592.50	
	230	226	19	9	2	
Ninth-Graders	635.50	649.56				
	16	16				
Tenth-Graders	643.42	657.00	660.00			
	31	31	5			
Eleventh-Graders	645.20	665.19				
	15	16				
SDRT4 Total (SD=38.67)						
Sixth-Graders	584.92	608.39	625.57	631.17	641.50	625.00
	533	542	226	53	4	1
Seventh-Graders	596.76	616.35	625.85	636.65	639.00	557.00
	324	324	123	17	6	1
Eighth-Graders	607.70	624.53	632.89	616.56	591.50	
	228	222	19	9	2	
Ninth-Graders	617.75	635.94				
	16	16				
Tenth-Graders	636.52	648.48	650.00			
	31	29	5			
Eleventh-Graders	634.94	657.00				
	16	15				

***Achievement Measure* / Starting Grade Level**	**Pretest**	**First Posttest**	**Second Posttest**	**Third Posttest**	**Fourth Posttest**	**Fifth Posttest**
SRI Score (SD=257.40)						
Sixth-Graders	334.20	471.45	602.78	535.72	576.80	787.00
	550	542	215	53	5	1
Seventh-Graders	404.18	552.78	661.50	588.18	626.67	113.00
	334	316	123	17	6	1
Eighth-Graders	457.35	612.19	614.53	562.00	352.00	
	232	218	19	10	2	
Ninth-Graders	537.93	615.00	654.00			
	15	16	1			
Tenth-Graders	597.93	744.31	754.40			
	30	26	5			
Eleventh-Graders	618.81	713.23				
	16	13				
ITBS/ITED Comprehension (SD=24.73)						
Sixth-Graders	179.53	187.25	195.41	201.68	194.33	198.00
	408	421	230	95	3	2
Seventh-Graders	181.90	191.34	201.92	207.84	206.00	194.00
	321	262	149	102	2	1
Eighth-Graders	190.87	197.53	200.80	226.57		
	225	172	111	70		
Ninth-Graders	207.00	215.83	220.00			
	11	6	2			
Tenth-Graders	208.00	212.84	237.00			
	28	25	1			
Eleventh-Graders	208.50	171.00	207.00			
	6	2	1			

Exhibit reads: Students who begin READ 180 in sixth grade have a mean pretest score of 588.01 and a mean first posttest of 609.52, on SDRT4 Comprehension. The standard deviation for this achievement measure is 38.66.

Note: SD = Standard deviation, pooled across all students at all time points. Sample sizes are provided below each mean.

Source: Authors' tabulations from Des Moines City School District Database

As expected, there is roughly a linear trend in achievement scores over time – that is, an increase from pretest to first posttest and an increase from first posttest to second posttest, and so on. This pattern holds true for all five achievement measures and all starting grade levels.

We also observe in Exhibit A-1 that the pretest of each grade level is higher than the pretest at the previous grade level, such that seventh is higher than sixth, eighth higher than seventh, and so forth. This confirms our suspicion that starting grade level must be included in statistical analyses as a control measure. That is, because starting grade level is associated with achievement, we want to control for starting grade level in order to observe the direct relationship between other measures and achievement.

Correlations

It is useful to consider the extent to which the five achievement measures are correlated with each other. Exhibit A-2 presents the correlations among the five achievement measures. Aside from SDRT4 Total, which is a composite of the other scores, the five achievement measures are only moderately correlated with each other for this sample of students, with Pearson correlation coefficients usually in the 0.4 to 0.6 range.

Exhibit A-2
Correlations among Achievement Measures and Sample Sizes

	SDRT4 Comprehension	SDRT4 Vocabulary	SDRT4 Total	SRI Score
SDRT4 Comprehension	1.0			
SDRT4 Vocabulary	0.67 2,798	1.0		
SDRT4 Total	0.86 2,762	0.78 2,761	1.0	
SRI Score	0.65 2,694	0.64 2,675	0.64 2,630	1.0
ITBS/ITED Comprehension	0.52 2,143	0.46 2,130	0.49 2,092	0.49 2,108

Exhibit reads: The correlation between SDRT4 Comprehension and SDRT4 Vocabulary is 0.67.

Note: All correlations in this exhibit are statistically significant at $p < 0.05$. Sample sizes are provided below each correlation coefficient.

Source: Authors' tabulations from Des Moines City School District Database

The correlations displayed in Exhibit A-2 are lower than we might have expected for correlations among reading achievement measures. For example, in previous research, Lexile scores – the units of measurement on the SRI – have been reported to be correlated very highly with standardized tests of reading achievement, including 0.91 with SDRT4, and 0.90 or greater with several widely-used nationally-normed tests such as the SAT9, CTBS5/Terra Nova, Gates-MacGinitie, and Metropolitan 8 (MetaMetrics Inc., n.d.). The low correlations observed for this sample of students may be related to the fact that this sample is composed exclusively of special education students. Some test experts believe that standardized tests such as these are not appropriate for special education populations because the scores are too low in the distribution of possible scores to be reliable measures.[15] The existence of the low correlations justifies our interest in all five achievement measures.

[15] Scores at either extreme of a distribution are known to be unreliable in the sense that repeated measures do not provide consistent, "reliable" results: we don't get the same answer every time.

Appendix B: Technical Appendix on Growth Curve Analysis

This appendix is included for readers who are interested in additional technical details of the analytic methods and results reported here. We first provide technical information on methods (data, measures, and models) in sufficient detail to meet standards for replicability of research. Following the technical discussion of methods, we present the results of the HLM growth curve models in their full context.

Methods

Growth curve modeling, in contrast to a series of annual analyses, provides the opportunity to sort out different causes associated with change over time. Specifically, achievement change that is observed in a series of annual analyses could be attributable to three possible sources: (1) an intervention, such as READ 180; (2) normal individual growth that would occur anyway; or (3) changes that occurred in the population at large. These three sources of change are not mutually exclusive, and are probably all operating simultaneously. What growth curve modeling does is help sort out the three sources of change, and thereby provide more confidence in results associated with an intervention. In particular, growth curve modeling allows us to measure the sources of change and include them in analyses

Growth curve modeling requires multiple, sequential measures of the same concept (here, reading achievement) using the same test over time, with the same set of individuals (students), while controlling for other measures. Conceptually, there is a distinction between two types of control measures: time-varying measures and non-time-varying measures. Some variables are inherent in the student – a good example is student gender. Such a variable will not change its value, no matter how many times it is measured. Other variables, in contrast, do vary from time point to time point. One example of a time-varying measure is a student's special education status. Across a series of testing time points, a student may move in and out of a special education program, and thus the value for this measure changes from time point to time point. A non-time varying measure, such as gender or race, does not change over time and its value remains constant for each student.

The distinction between time-varying and non-time-varying measures parallels the two levels of analysis in growth curve modeling. Growth curve modeling, as implemented for this study, has two levels of analysis – students *and* test points – which are modeled simultaneously. Modeling both levels simultaneously means we can include measures that are related to either level of analysis. In fact, measures related to the student level are the measures that are non-time-varying as discussed above. Measures related to the test point level are measures that are time-varying: their value may change from test point to test point.

A practical implication of this complex form of analysis is that having two levels of analysis means there are two sets of everything else as well: two sets of measures, as already

discussed; two sets of sample sizes; two sets of coefficients. The remainder of this section of the report describes the data that were available, the measures used in analyses, and the statistical models employed.

Data

The Des Moines Independent Community School District's Department of Student and Family Services provided the data for this report, and these are the same data used in the annual status reports (Palmer, 2003a, 2003b, 2004). The database includes data for all students who participated in READ 180 in Des Moines during one or more of the first five school years during which READ180 was implemented: 2000-01 through 2004-05. In order to protect student confidentiality, all student names and other identifiers were deleted before the data were transferred to PSA for analyses.

Information available for each student included basic demographic information as well as test scores on several achievement measures. For each student for each year[16] – starting with the year he or she first participated in READ 180 (or occasionally the prior year) and continuing beyond participation in READ 180 – the database includes information about grade level, level of special education services provided, status of READ 180 participation, school attended, and student's teacher assignment. The database also includes scores on one more of the five achievement measures discussed earlier: SDRT4 Comprehension, Vocabulary, and Total;[17] SRI, and ITBS/ITED Comprehension. Some of these items of information were used directly as variables in statistical analyses; others were used to construct new variables for use in analyses. We will return to this distinction later in describing measures.

Across the five years covered in this study, the annual number of students who enrolled in READ 180 for the first time ranged from just under 200 to almost 350. (See Exhibit B-1.) Cohort 1, which included students whose first enrollment was in 2000-01, was the largest READ 180 cohort, enrolling 110 sixth graders, 130 seventh graders, and 102 eighth graders – a roughly even distribution across the middle school grades which were the initial focus of READ 180 in Des Moines. Subsequent cohorts enrolled similar numbers of sixth graders, but most seventh and eighth graders in subsequent years were students continuing from previous cohorts, with many fewer seventh and eighth graders identified for first-time enrollment. Cohorts 2 through 5 thus are somewhat smaller than the first cohort, with Cohort 4 the smallest of all of the cohorts.

[16] Most students have incomplete data across the five-year span – that is, most students do not have information for every year, but only for some years.

[17] On the SDRT4, Reading comprehension and vocabulary are components of the Total Reading score, along with a third element, Scanning. For SDRT4 Scanning, analyses showed that the relationship with READ 180 could not be determined: Standard errors of model coefficients were unstable, possibly due to wide variation in scanning scores, suggesting that the coefficients were not reliable estimates of effects. SDRT4 Scanning results are thus not available and not discussed further.

Exhibit B-1
First-Time Enrollment in READ 180 by *Cohort Year* and *Starting Grade Level*

Grade	Cohort 1 2000-01	Cohort 2 2001-02	Cohort 3 2002-03	Cohort 4 2003-04	Cohort 5 2004-05
6	110	112	123	95	124
7	130	52	57	40	70
8	102	49	45	31	10
9		5	1	4	6
10		16	6	4	5
11		10	2	2	2
Totals	342	244	234	176	217

Exhibit reads: Cohort 1 included 110 sixth-graders who participated in READ 180 for the first time.

Source: Authors' tabulations from Des Moines school district database

For analyses, we pooled all 5 of the cohorts into a single, large sample of students. The rationale for pooling was to increase sample size and thereby improve our statistical power for detecting effects. Additionally, pooling the cohorts while also controlling for a measure of cohort is a statistical technique that allows the growth curve modeling to tease out the three possible sources of change. After pooling the cohorts, the total number of students in the database who participated in a year or more of READ 180 is 1,213 – the sum across the bottom row of Exhibit B-1.

The database includes pretest and posttest scores on at least one of the five achievement measures for all but one student. However, the number of students included in any one achievement analysis is smaller than the total number of students in the database. Specifically, scores for over 90 percent of the 1,213 students were available for the analyses of results on the SDRT4 (Comprehension, Vocabulary, Total) and the SRI, and scores for about 60 percent of these students were available for the ITBS/ITED analyses. The reason for this difference is that the SDRT4 and the SRI in Des Moines are specifically targeted for READ 180 students. Every effort is made to administer the tests to as many students as possible as they enter READ 180 and again at the end of each year in READ 180. The Department of Student and Family Services has been extremely successful in this endeavor. In contrast, the ITBS/ITED is a test administered district-wide on a yearly basis. The extent to which any one student will have test scores in the two relevant consecutive years (to serve as pretest and posttest) is much less common on a yearly testing cycle.[18]

As discussed earlier, the analyses conducted for this study use two levels of analysis: students and their test points. For growth curve modeling, there thus are multiple data points for each student, with each data point representing an individual test score. Consequently, the number of data points in the database is greater than the number of students in the database. Exhibit B-2 illustrates the number of data points that result from structuring the database in this way.

[18] Note that, for any single year of ITBS data, the participation rates are well over 90 percent.

Exhibit B-2
Numbers of Students Providing Test Data at Each Time Point

Testing Time Point	SDRT4 Comprehension	SDRT4 Vocabulary	SDRT4 Total	SRI	ITBS/ITED Comprehension
0 – Pretest (First test point)	1,163	1,161	1,140	1,168	990
1 – First posttest (Second test point)	1,175	1,161	1,139	1,122	888
2 – Second posttest (Third test point)	379	375	371	361	421
3 – Third posttest (Fourth test point)	75	75	74	74	45
4 – Fourth posttest (Fifth test point)	12	13	12	13	5
5 – Fifth posttest (Sixth test point)	2	3	2	2	3
Totals	2,806	2,788	2,738	2,740	2,352

Exhibit reads: One thousand one hundred sixty-three students provide test data from SDRT4 Comprehension at the pretest time point. Slightly more students (1,175) provide test data at the first posttest (that is, the second test point), and 379 provide test data at the second posttest (third test) time point. The total number of student-test points available for growth analysis for SDRT4 Comprehension sums to 2,806.

Note: The counts given in this exhibit reflect observations that have complete data on all analysis variables.

Source: Authors' tabulations from Des Moines school district database

In Exhibit B-2, we see that the number of students who have scores on SDRT4 Comprehension at pretest was 1,163. These students' test scores represent 1,163 data points in the database. At the time of the first posttest (the second overall test point), 1,175 students have scores on the SDRT4 Comprehension test; these students' test scores represent an *additional* 1,175 data points in the database, for a total of 2,338 data points after two testing time points. If we continue in this manner, adding the number of test scores available at each time point, we obtain a total of 2,806 students' test scores after the fifth posttest. Similar interpretations hold for the other achievement measures presented in Exhibit B-2. For SDRT4 Vocabulary, there are 2,788 test scores; for SDRT4 Total, 2,738 test scores; for SRI, 2,740 test scores; and for ITBS/ITED Comprehension, 2,352.

Measures

It is important to maintain the distinction among data, measures, and variables. Data are the raw information in the database, while measures are the conceptual factors we wish to study, and variables are the technical, procedural version of measures used in analyses. We discussed previously the information that was available in the data file from Des Moines. Some of the information in the data file was useful as-is, while other information was useful for creating new measures. Recall also that we are here interested in two levels of analysis: students and test points. Measures can be associated with either students or test points, but not both.

Analyses in this report rely on the outcome and predictor measures listed in Exhibit B-3. All five of the outcome measures are taken directly from the Des Moines database and are time-varying test-point-level measures. Of the six predictor measures, three are time-varying, test-point-level measures: *READ 180 status* and *special education status* are taken directly from the Des Moines database, and *time count* was created to designate the time between pretest and each posttest using the school year as the metric. The remaining three predictor measures are student-level non-time-varying measures: *cohort year, starting grade level*, and *participation pattern*.

Exhibit B-3
Summary of Measures Used in Analyses, by Type

***Type*/Name of Measure**	**Description**
Outcome Measures	
SDRT4 Comprehension Subtest Score	A measure of how well students understand and analyze reading material
SDRT4 Vocabulary Subtest Score	A measure the range of words a student knows relative to grade-level norms
SDRT4 Total Reading Score	A composite measure of three reading subtests, scaled independently based on grade-level norms
SRI Lexile Score	A computer-adaptive measure of students' reading comprehension
ITBS/ITED Comprehension Subtest Score	A measure of students' ability to make inferences or generalizations about passages they have read
Predictor Measures – Time-Varying, Test Point Level	
READ 180 status	Student's participation status in READ 180 at each test point
Special education status	Student's participation status in special education services at each test point
Time count	A measure of the time between pretest and each testing point, in school-year increments
Predictor Measures – Non-Time-Varying, Student Level	
Cohort year	The school year during which a student first enrolled in READ 180
Starting grade level	Student's grade level assignment when first participated in READ 180
Participation pattern	The five-year pattern of entering, withdrawing, and re-entering READ 180

Cohort year and *starting grade level* are taken directly from the district database; *participation pattern* was created from information on student enrollment in READ 180 across the five years. The remainder of the of this section provides additional descriptions of each measure, starting with the achievement measures, moving to the time-varying predictor measures, and finishing with the non-time-varying predictor measures. Where relevant, a brief description is also provided of how measures were "operationalized" as variables in models.

SDRT4. The SDRT4 is a multiple-choice, group-administered, norm-referenced assessment of students' strengths and needs in reading, for grades 2-12. The SDRT4 offers eight levels and two forms; in Des Moines, the purple level (geared towards grade levels 4.5-6.5), Form J, was administered to all test-takers, regardless of current grade-level assignment, in its

paper-and-pencil form rather than online. The paper-and-pencil SDRT4, purple level, which covers literature, informational text, and functional material, provides subtest results as well as a composite score. The following three measures are treated separately in all analyses and discussions:

- ***SDRT4 Comprehension subtest*** measures how well students understand and analyze reading material, and includes some cloze questions and comprehension questions at informational, inferential, and prediction levels of analysis.

- ***SDRT4 Vocabulary subtest*** measures the range of words a student knows, relative to grade-level norms, and uses picture-word matching at the elementary level and synonym matching at higher levels.

- ***SDRT4 Total Reading*** is a composite of these two subtests plus a third, scanning, but is scaled independently based on grade-level norms rather than being a simple average of the three scores.

SRI. The SRI, which is an internal component of READ 180, is a computer-adaptive assessment for grades 1-12 designed to measure students' reading ability in terms of reading comprehension. Scores are reported as Lexile levels, or simply Lexiles, which range from 200 to 1800. Because the assessment is computer-adaptive, the difficulty of the questions that are asked is automatically adjusted up or down, depending on student responses, to assess each student's reading ability more accurately. This feature reduces the potential for floor or ceiling effects that are associated with assessments that are not adaptive. The SRI uses "authentic" text passages drawn from children's and young adults' literature, classics, and periodicals. It measures students' reading comprehension skills such as identifying details, cause and effect, and event sequence; drawing conclusions; and making comparisons and generalizations.

ITBS/ITED Comprehension Subtests. The ITBS Comprehension subtest is a two-part, group-administered assessment of students' ability to make inferences or generalizations about passages they have read. The passages range in length and can be of various types, including fiction, poetry, interviews, scientific information, and other non-fiction. In Des Moines, the ITBS was administered to sixth through eighth graders. In 2000, 2001, and 2002, some READ 180 students were assessed using a test level lower than their grade placement. Starting in 2003, all students took the test level corresponding to their assigned grade level at the time of testing. For all analyses in this report, we used scale scores, which allowed cross-level and cross-grade comparison of scores and investigation of growth over time.

The Iowa Tests of Educational Development (ITED) Comprehension Subtest is the high school correlate of the ITBS. The ITED Comprehension subtest is a group-administered assessment of high school students' ability in reference to factual understanding, inferential understanding, and analytic understanding of passages they have read. The passages range in length and can be of various types, including fiction, poetry, expository non-fiction, memoirs, essays, biographical sketches, and editorials. The ITED was administered to ninth through eleventh graders in Des Moines. In 2000, 2001, and 2002, some READ 180 students were assessed using a test level lower than their grade placement. Starting in 2003, students took the

test level corresponding to their assigned grade level at the time of testing. For all analyses, we used scale scores, which allowed cross-level and cross-grade comparison of scores and investigation of growth over time.

In addition to student achievement results on the assessment measures just discussed, the statistical analyses included six predictor measures (as listed in Exhibit B-3). The measures are all taken directly from the district database or are created based on other information available in the database. Three of the six predictor measures are time-varying test-point-level measures: *READ 180 status, special education status,* and *time count.* The other three are the three non-time-varying student-level measures (*cohort year, starting grade level,* and *participation pattern*).

READ 180 status. Student participation in READ 180 at the time of each posttest was the measure of most interest in this study because the effect associated with this measure provides an estimate of the effectiveness of the READ 180 program for this sample of students. For the pretest and the first posttest, all students were enrolled in READ 180. However, due to different participation patterns over the years, students may or may not have been participating in READ 180 at the time of subsequent posttests. For analytic purposes, we assign this variable a value of 1 for testing time points at which students had been enrolled in READ 180 for a full year and a value of 0 for testing time points at which they were not enrolled or were enrolled for less than a full year. For testing time points at which students had a value of 0 for *participation status,* they served as comparison students. Thus, although all students included in the analyses had been in READ 180 for at least one year, individual students served as comparisons only for the time points at which they were no longer enrolled in READ 180. The effects associated with this variable in the analyses are the effects attributed to the READ 180 program for this sample of students, after controlling for other variables in the analysis.

Special education status. *Special education status* refers to the categories of special education services provided by the district, and it is a measure whose value could vary from one time point to the next. For the present study, at each time point, a student's special education status was measured as one of four levels: Up to 2 hours per day (referred to as 'resource' level); Special class 1-3 hours per day; Self-contained most of the day; and Self-contained all day.[19] The measure of special education was operationalized as a series of four "dummy" variables, one for each category of special education, each coded 1 for students in that category and 0 for students not in the category. For analyses, one of the dummy variables must be omitted to serve as the reference category. In the final models for this study, the lowest level, Up to 2 hours per day, served as the omitted reference category.

Time count. The longitudinal growth curve analyses necessary for this study require a measure of the passage of time. For this, we created a simple counting variable, *time count,* with assigned values of 0 at a student's pretest time point, 1 at the first posttest time point, 2 at the second, and so on, up to a maximum possible of 5, representing the fifth posttest time point, for six (0-5) possible test points overall. The length of time between each posttest is approximately one school year. In the longitudinal analyses, the effect associated with this variable represents

[19] These are the levels of special education services as defined by the State of Iowa. Iowa identifies students for special education services as "entitled individuals" rather than assigning disability labels.

the rate of change in achievement over time, or *growth,* which represents the second source of change that growth curve modeling sorts out – the change associated with normal individual growth that occurs naturally.

The remaining three measures listed in Exhibit B-3 were student-level non-time varying measures:

Cohort year. *Cohort year* refers to the school years and cohorts discussed previously, and has five possible values: 1, 2, 3, 4, or 5. *Cohort year* is an important measure to control for in any longitudinal analysis because each cohort is subject to slightly different historical conditions, such as changes in district policy, that affect students in different grades or different cohorts in different ways. Since we are unable to control for the countless separate factors that contribute to the unevenness of historical conditions, controlling for *cohort year* is the best way to even out the uneven reality of local circumstances. This is one facet of how growth curve modeling teases out the three sources of change: *cohort year* represents the third source of change discussed previously, that related to changes that occur in the population at large.

Starting grade level. Starting grade level ranged from sixth to eleventh grades. *Starting grade level* is an important factor to control for because intervention programs often have different impacts at different grade levels. In fact, previous research in Des Moines indicated that there may be differential impacts for READ 180, in particular for seventh versus eighth graders who are in their second year of READ 180 (Palmer, 2003a and 2003b).

Participation pattern. *Participation pattern* refers to students' histories of entering, leaving, and re-entering the READ 180 program over the years of the study. There were seven distinct patterns of participation, defined in terms of the number of years of participation and the extent to which participation was consecutive, but several of these had very few students that followed them – too few, in fact, to allow meaningful analysis. Exhibit B-4 displays the seven patterns and the numbers of students for each pattern.

Exhibit B-4
Seven READ 180 Participation Patterns

Participation Pattern	Participating Students	
	Number	Percent
One year only	831	68.51
Two consecutive years only	308	25.39
Three consecutive years only	53	4.37
Two non-consecutive years	14	1.15
Three non-consecutive years	2	0.16
Four consecutive years	4	0.33
Five consecutive years	1	0.08
Totals	1213	100.00

Source: Authors' tabulations from Des Moines school district database

From Exhibit B-4, we can see that the two most common participation patterns accounted for 94 percent of all students: just over two thirds of students participated in READ 180 for a single year without ever returning to the program during the five years of the study; and another quarter participated for two consecutive years, leaving after those two years and not returning

during the five-year span of the study. In addition, just over 4 percent participated for three consecutive years only. For analyses, we operationalized *participation pattern* as a series of four dummy variables: one year only, two consecutive years, three consecutive years, and other (for the latter, we collapsed the four smaller categories into one catch-all category). Recall that a series of dummy variables must omit one category in analyses, to serve as a reference category. In this study, three consecutive years served as the omitted reference category in analyses because it was the category that was most different from the others (i.e., had the most statistically significant contrasts with the other categories).

Exhibit B-5 presents ranges, means, and standard deviations for the predictor measures as operationalized for analysis. (See Appendix A for descriptive statistics for the outcome measures.)

Exhibit B-5
Descriptive Statistics for Predictor Variables

Measure	Range	Mean	SD
Student-Level Measures (Non-time-varying)			
READ 180 status	0 or 1	0.50	0.50
Special education status			
Level 1 (reference category)	0 or 1	0.30	0.45
Level 2	0 or 1	0.10	0.30
Level 3	0 or 1	0.52	0.50
Level 4	0 or 1	0.08	0.28
Time count	0 to 5	1.02	0.98
Time-Point-Level Measures (Time-varying)			
Cohort year			
Cohort 1 (reference category)	0 or 1	0.28	0.45
Cohort 2	0 or 1	0.20	0.40
Cohort 3	0 or 1	0.19	0.39
Cohort 4	0 or 1	0.15	0.35
Cohort 5	0 or 1	0.18	0.38
Starting grade level			
Sixth (reference category)	0 or 1	0.46	0.50
Seventh	0 or 1	0.29	0.45
Eighth	0 or 1	0.20	0.40
Ninth	0 or 1	0.01	0.11
Tenth	0 or 1	0.03	0.16
Eleventh	0 or 1	0.01	0.11
Participation pattern			
One year only	0 or 1	0.69	0.46
Two consecutive years	0 or 1	0.25	0.44
Three consecutive years (reference cat.)	0 or 1	0.04	0.20
Other	0 or 1	0.02	0.13

Exhibit reads: The variable measuring *READ 180 status* had possible values of 0 or 1, with a sample mean of 0.50 and standard deviation of 0.50.

Models

In order to investigate growth trends over time, we used two-level hierarchical linear modeling where students were the Level 2 data points and testing time points nested within students were the Level 1 data points. A separate series of statistical analyses was conducted for each of the achievement measures available in the database, resulting in five sets of analyses.

In each analysis, we included five independent variables. At Level 1, we included the *time count* variable, which was allowed to vary randomly in all models. This means that the effect of this variable could vary across students and that this varying effect could be predicted by additional variables if included in the model.

At Level 1 of all models, we also included an indicator of students' *READ 180 status* at each testing time point. Initial investigation of *READ 180 status* indicated that it should be treated as a fixed effect, which means its effect does not vary across students, but has a constant effect on achievement for all students. The evidence for this was generated from models where *READ 180 status* was allowed to vary randomly; in these models, the reliability estimate for the *READ 180 status* coefficient was less than 0.1, the customary cut-off for indicating a variable should be allowed to vary randomly; *and* the p-value for the hypothesized random effect of *READ 180 status* was not statistically significant, leading us to reject the hypothesis that *READ 180 status* has a randomly varying effect. *READ 180 status* was thus entered into models as a fixed effect.

The third variable included in all analyses was the student's *special education status* at the time each test score was obtained. As with READ 180 status, this variable could change from test point to test point. This measure was operationalized as a series of three dummy variables, with resource level serving as the omitted reference category.

Following standard model-building procedures, we investigated the effects and statistical significance of two sets of control factors, both of which were Level 2 (student-level) factors. The first was *starting grade level,* operationalized as a series of five dummy variables coded 1 or 0 and representing seventh through eleventh grades, with sixth grade serving as the omitted reference category. The second control factor was student's *participation pattern* in READ 180 across the five years as a whole. Year-to-year status in READ 180 can change, but the five-year participation pattern is always the same for any given student. This was operationalized as a series of three dummy variables coded 1 or 0 representing: one year only, two consecutive years, or other.. The three consecutive years pattern served as the omitted reference category.

These two control factors were tested in preliminary models for effects on initial achievement (the intercept) and for effects on the growth rate in achievement (the slope on the time variable). In general, *starting grade level* had a statistically significant positive effect on initial achievement, and somewhat less frequently had a statistically significant negative effect on rate of achievement growth. *Participation pattern* was only occasionally statistically significant. Where effects were statistically significant, the factors were included in final models; where not statistically significant, they were omitted.

In addition to the five independent variables noted above, which were all included in all analyses, one other variable was included in some analyses, where the effect of this variable was discovered to have a statistically significant relationship with achievement. That variable is the *cohort year* during which a student first participated in READ 180. *Cohort year* was only a statistically significant predictor in analyses involving SDRT4 Total and SRI score. Again, this was operationalized as a series of dummy variables coded 1 or 0 representing Cohorts 2-5, with Cohort 1 as the omitted reference category.

Each of the five measures of achievement – SDRT4 Comprehension, SDRT4 Vocabulary, SDRT4 Total, SRI, and ITBS Comprehension – served as the dependent variable in a separate series of preliminary and final statistical models. For each achievement measure, we used the scale score in all statistical analyses. We used the HLM Version 6.0 statistical software and all variables were entered into models without any group- or grand-mean centering. A typical model is depicted in Exhibit B-6.

Exhibit B-6
Example HLM Model: Predicting SDRT4 Comprehension Scores

Level 1 Model

$$Y = B0 + B1*(TIME) + B2*(READ180) + B3*(LVLSVC2) + B4*(LVLSVC3) + B5*(LVLSVC4) + R$$

Level 2 Model

$$B0 = G00 + G01*(GRADE7) + G02*(GRADE8) + G03*(GRADE9) + G04*(GRADE10) + G05*(GRADE11) + G06*(ONEYEAR) + G07*(TWOYEAR) + G08*(OTHPATT) + U0$$
$$B1 = G10 + G11*(ONEYEAR) + G12*(TWOYEAR) + G13*(OTHPATT) + U1$$
$$B2 = G20$$
$$B3 = G30$$
$$B4 = G40$$
$$B5 = G50$$

Exhibit reads: Test score (Y) is a function of time point, *READ 180 status, special education status*, a random error term (R), and a constant term (B0). The constant term is a function of *starting grade level, participation pattern*, and a random error term. The effect associated with time (B1) is a function of *participation pattern* and a random error term.

Results

Overall student achievement results suggest that READ 180 had a positive, statistically significant effect on three of the achievement measures while having no measurable effect on the other three achievement measures. In particular, for each additional year that students participated in READ 180, students scores increased by approximately 30 Lexiles on the SRI, 6 scale-score points on SDRT4 Comprehension, and 4 scale-score points on SDRT4 Total – above and beyond the observed yearly growth in achievement. For ITBS/ITED Comprehension, and for SDRT4 Vocabulary, the relationship with READ 180 was positive but not statistically significant and deemed "no effect." For SDRT4 Scanning, the relationship with READ 180

could not be determined: Standard errors of model coefficients were unstable, possibly due to wide variation in scanning scores, suggesting that the coefficients were not reliable estimates of effects. SDRT4 Scanning growth curve results are thus not available and not discussed further. Exhibit B-7 presents the results of the final HLM models for the five other achievement measures.

For the HLM results presented in Exhibit B-7, the Level-1 intercept is interpretable as the mean baseline achievement score for the group of students with values of zero on all of the variables included in the relevant model. For SDRT4 Comprehension, for example, this group would be sixth-graders who had a three-consecutive-years pattern of participation (these are the students who would have zeros for the series of grade 7-11 dummy variables and for the series of pattern dummy variables). For SDRT4 Total, as a contrasting example, the reference group is sixth-graders from Cohort 1 with a three-consecutive-years participation pattern. The level-1 coefficients on the *starting grade level* dummy variables indicate that each successively higher grade had a mean pretest score somewhat higher than the mean pretest score for sixth-graders. The level-1 coefficients on the cohort dummy variables have a similar interpretation.

The positive coefficients on the *participation pattern* dummy variables indicate that each of these patterns start with a higher pretest score than the omitted reference category, which here is three-consecutive-years participation pattern. Thus, for example, students in the one year only participation pattern started READ 180 with a SDRT4 Comprehension pretest score that was 16.04 points higher than students in the three consecutive years participation pattern (Exhibit B-7). A similar interpretation applies to the other two pattern dummy variables, relative to the omitted three-consecutive-years participation pattern.

In Exhibit B-7, an aspect of the results that deserves attention is the relationship between achievement and time – that is, growth in achievement. The results confirm that there is a linear growth trend such that each year, a student's achievement score increases an average number of points. This held true for all five achievement measures and was statistically significant in all five analyses. For SDRT4 Comprehension, the number of points associated with each year's increment of achievement growth was 15.27; for SDRT4 Vocabulary the increment was 17.21 points per year increase; for SDRT4 Total, 20.59; for SRI, 74.01; and for ITBS/ITED Comprehension, 4.58. Additionally, the coefficients for the time slope indicate associations between each variable and the rate of growth. The negative coefficients on the grade-level associations with time indicate that the *rate* of growth was slower, but still positive, for higher grade levels – that is, relative to sixth-graders, seventh-, eighth- and tenth-graders had slower rates of growth in achievement for SDRT4 Vocabulary and for SDRT4 Total.

The *special education status* variables in Exhibit B-7 are Level 1 variables which potentially change value from testing point to testing point for any given student. They are dummy variables with service level one as the omitted reference category (resource level), and their coefficients describe the relationship between that variable and achievement. So, for example, students with a service level of two have lower achievement than students with a

Exhibit B-7
Results of HLM Growth Models: Final Models by Achievement Measure

	SDRT4 Comprehension	SDRT4 Vocabulary	SDRT4 Total	SRI Score	ITBS/ITED Comprehension
For Intercept-Model for Pretest					
Intercept	585.95 ***	585.36 ***	586.47 ***	332.64 ***	185.35 ***
Cohort2			5.82 *	57.13 **	
Cohort3			4.60	58.78 **	
Cohort4			-1.99	41.00 *	
Cohort5			4.81	33.01	
Grade7	8.37 ***	15.85 ***	12.21 ***	79.73 ***	3.32 **
Grade8	16.41 ***	32.59 ***	22.01 ***	137.97 ***	10.68 ***
Grade9	22.05 **	54.32 ***	29.56 **	179.91 **	17.63 **
Grade10	35.56 ***	60.87 ***	43.72 ***	221.47 ***	25.67 ***
Grade11	33.81 ***	62.88 ***	41.46 ***	218.33 ***	15.71
One-Year	16.04 **		6.94 **		
Two-Year	8.68		-3.23		
Other	16.74 *		2.02		
For TIME-Ach Slope					
Intercept	15.27 ***	17.21 ***	20.59 ***	74.01 ***	4.58 **
Cohort2			-0.23		
Cohort3			-0.25		
Cohort4			6.19 **		
Cohort5			-8.97 **		
Grade7		-6.94 ***	-4.44 **		
Grade8		-8.30 **	-8.48 ***		
Grade9		-14.18	-2.38		
Grade10		-17.40 **	-12.58 *		
Grade11		-9.60	-2.18		
One-Year	-3.34	12.19 ***	-0.53	43.99 **	4.14 **
Two-Year	2.43	9.67 ***	-4.72 *	60.59 ***	4.06 **
Other	-8.65 **	5.93	-5.70 *	-24.03	0.69
For READ180-Ach slope					
Intercept	6.11 **	0.72	4.72 *	28.16 *	0.31
For LevSvc2-Ach slope					
Intercept	-10.37 ***	-10.41 **	-10.44 ***	-39.83 *	-6.60 ***
For LevSvc3-Ach slope					
Intercept	-15.13 ***	-11.00 ***	-12.95 ***	-64.56 ***	-9.74 ***
For LevSvc4-Ach slope					
Intercept	-12.01 ***	-5.64	-9.48 **	-17.83	-4.53 *
Proportion of Variance Explained	0.20	0.18	0.26	0.16	0.34
No. of students	1,204	1,204	1,203	1,204	1,159
No. of test points	2,806	2,788	2,738	2,740	2,352

Exhibit reads: For SDRT4 Comprehension, sixth-graders have a pretest scale score of 585.95 just prior to their first year of participation in READ 180. Students in higher grade levels begin with higher initial scores. Students experience average yearly growth of 15.27 scale-score points, and READ 180 students additionally attain 6.11 scale-score points for each additional year of participation in READ 180.

Notes: * p<0.05; ** p<0.01; *** p<0.001. For the series of dummy variables, the omitted reference categories are as follows: For starting grade level, sixth-graders; for *cohort year*, Cohort 1; for *participation pattern*, three consecutive years; for *special education status*, Level 1.

Source: Authors' tabulations from Des Moines school district database

service level of one (the latter being the omitted reference category), by approximately 10 points on SDRT4 Comprehension.

The final item of interest in Exhibit B-7 is the coefficient on the READ 180 indicator variable, the key explanatory variable in these models. In three of the five models, the coefficient on READ 180 – that is, the *effect of READ 180* – is positive and statistically significant. Each additional year of participation in READ 180 is associated with additional points on each of those three achievement measures. For SDRT4 Comprehension, each additional year in READ 180 affords 6.11 additional scale-score points. With multiple years of additional participation, the additional scale-score points add up in multiples of 6.11. This leads to a divergence in scores for students who remain in READ 180 versus students who participate for only one year.

Exhibit B-8 depicts the divergence in achievement scores associated with participation, or not, in READ 180, using as an example achievement on SDRT4

Exhibit B-8
Effect of READ 180 on Achievement, SDRT4 Comprehension, for Students Starting READ 180 as Seventh-Graders

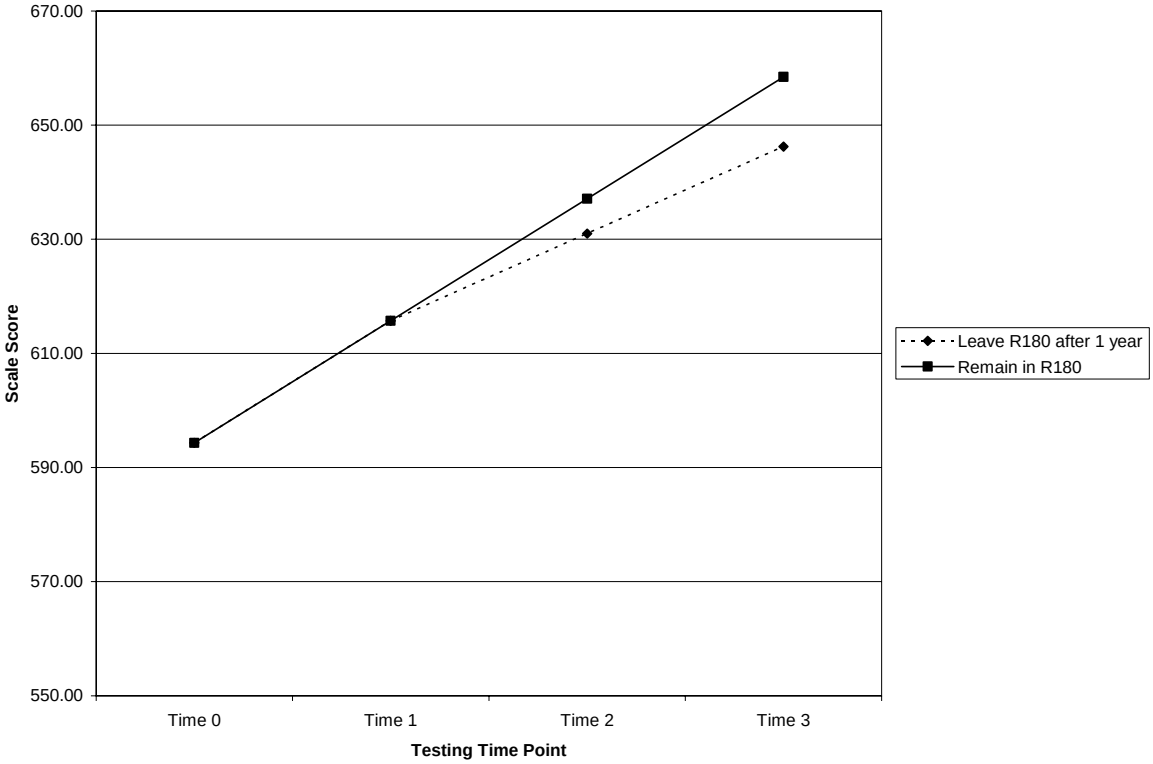

Exhibit reads: READ 180 students who remain in the program beyond the first year have higher mean scale scores on SDRT4 Comprehension than students who leave the program after only one year of participation.

Source: Authors' tabulations from Des Moines school district database, via growth curve analysis where n=1,204 students and 2,806 test points.

Comprehension for students who began READ 180 as seventh graders. In this exhibit, Time 0 represents the pretest for seventh graders, administered in the fall just prior to their participation in READ 180. Time 1 represents their first posttest, administered the spring after their first year of participation. At Time 2, we see the scores begin to diverge into two groups: students who remain in READ 180 for another year have slightly higher scores than students who no longer participate in the program. Scores for the two groups continue to diverge in subsequent years such that each additional year of participation in READ 180 is associated with additional increase in the gap between the two groups' scores. After three years of divergence, there is a gap of 12 scale-score points between the two groups: 646 versus 658. Since a student's annual expected gain in SDRT4 scores is 15 points (Scholastic, 2002), this 12-point gap represents a difference of four-fifths of a year's worth of growth. The same pattern is observed for the other grade levels on SDRT4 Comprehension, with different starting and end points, but the same 12-point advantage for students who continue in READ 180.[20]

The results for SDRT4 Total and SRI indicate a pattern of results similar to that depicted in Exhibit B-8, with somewhat different start and end points and rates of divergence. The rate of divergence on SDRT4 Total is slightly lower than for the SDRT4 Comprehension subtest, with 4.72 additional scale-score points associated with each additional year of participation in READ 180, such that after 3 years of posttests, the gap between participants and non-participants is 9.4 points, representing a READ 180 advantage of 0.63 years of growth. For the SRI, each additional year of READ 180 participation is associated with 28.16 additional points. This seems much higher than the SDRT4 results, but the expected annual gain is also much higher (116 versus 15), so it translates into the same range of advantage: after three years of posttests, the READ 180 advantage on SRI is 56.32 points, or approximately 0.49 years of growth.

It is very important to bear in mind that these observed effects of READ 180 occur *on top* of other growth. So, while 0.49 years of advantage may seem small when considered that it is spread over 2 years, it is growth *in addition to* other growth students are experiencing. Furthermore, these results are conservative estimates of the effect of READ 180. Due to the nature of the data available, the comparison against which READ 180 is judged to have an effect involved students who have had some READ 180 already (at least one year). If we were able to compare READ 180 students to never-READ 180 students who were otherwise similar, we could expect to observe even larger effects of participation in READ 180.

On the other hand, the students who served as comparison students in the SDRT4 and SRI analyses were also more mobile by definition, and therefore more disadvantaged than the READ 180 students in those analyses. This difference arises because only students who had spent a full year or a portion of a year in READ 180 were administered the SDRT4 and the SRI., with full-year students designated as READ 180 participants, and less-than-full-year as comparison students. However, the fact of receiving only a portion of a year of READ 180 indicates that the student had changed schools, possibly to a school that could not accommodate the mobile student into their READ 180 program (N. Palmer, personal communication). That is, we know with certainty that the less-than-full-year-READ 180 students (the comparison

[20] It should be noted that this is a prediction, based on the HLM results and the assumption of a linear relationship between time and achievement score. There are very few students for whom we have measured five years of achievement scores. Refer to Exhibit 6 for numbers of students at each posttest.

students) were students who experienced mobility during the school year. For the full-year READ 180 students, we do not have enough information to know the extent of mobility – it is possible that some full-year READ 180 students were also mobile, but that the school they moved into was able to accommodate them into their READ 180 program.

It is thus possible, and even likely, that the comparison students (less-than-full-year READ 180) for the SDRT4 and SRI analyses were more disadvantaged than the READ 180 students. Since we do not have the same information for full-year READ 180 students, we cannot control for the effects of mobility. This means that any effects observed for the SDRT4 and SRI tests may be larger than we would observe if we could control for mobility.[21] Despite these limitations, having a flawed comparison group is better than having none at all, especially since we *are* able to control for other important variables, including pretest score. However, any interpretation of these results must consider these caveats.

[21] This caution does not apply to the ITBS/ITED tests, which are administered district-wide to all students, not just READ 180 students.