Effective Direct Instruction Implementations: The Impact of Administrative Decisions and Time

Technical Report 2016-1

Jean Stockard, Ph.D., Director of Research and Evaluation

April 3, 2016

Table of Contents

List of Tables

List of Figures

Executive Summary

While a large literature shows that Direct Instruction (DI) is more effective than other curricular programs, some DI schools and classrooms are more successful than others. A major reason for these differences is implementation fidelity – the extent to which teachers and schools administer the programs as they were designed. Teachers and schools that conform most closely to DI implementation protocols have the most success. Their students proceed through the programs at mastery more quickly and have higher achievement.

This report focuses on two general factors that influence the development of teachers' skills at implementing DI with fidelity: 1) the extent to which administrative practices and decisions support implementation of DI teaching practices and 2) the extent to which student performance improves as schools, teachers, and students have more experience with the program.

Data from different studies and school and community settings illustrate the effect of these factors. The data show that DI students make significantly more progress at mastery and have significantly higher achievement when

- their teachers implement the programs with greater fidelity
- their teachers have been trained for the specific programs they are teaching
- their teachers are given time and support to prepare lessons
- they are taught for the recommended time each week
- those deemed at risk are given extra instructional time ("double dosing")
- their teachers gain more experience in teaching the programs
- DI has been implemented at their schools for a longer period of time
- they start learning with DI in kindergarten

Effective Direct Instruction Implementations: The Impact of Administrative Decisions and Time[1]

A large literature has documented the effectiveness of Direct Instruction (DI) curricular programs. Students taught with DI have significantly higher achievement scores and greater growth in skills than those taught with other approaches. (See Coughlin, 2014 for a recent summary.) At the same time, not all teachers and schools have equivalent success with the programs. Gains are stronger in some schools and classrooms than in others. A major factor accounting for these differences appears to be variations in implementation fidelity – the extent to which teachers administer the programs as they were designed. All DI programs have extensive and detailed guidance regarding how the materials should be used. Teachers and schools that conform most closely to these implementation protocols have the most success.

Direct Instruction programs are based on mastery learning. The lessons are carefully designed and sequenced to ensure that students have mastered all prerequisite knowledge necessary for learning new things. This makes instruction more effective and more efficient. Students are better able to remember what they learn and to learn new material more quickly. A number of studies have shown that when students demonstrate mastery, in grade level DI programs, they perform better on standardized assessments. Those who are at or near grade level are much more likely than other students to score at or above the national mean on standardized achievement tests and at the proficient and advanced level on state assessments (Stockard, 2014).

The relationship of teachers' implementation fidelity, students' progress through the DI programs and student achievement can be conceptualized as a causal chain, illustrated in Figure 1. Teachers who are more skilled at implementing the program with fidelity have students who make greater progress through the curriculum at mastery. In turn, this greater progress at mastery results in higher achievement scores. In other words, when teachers are more skilled, their students master the material and progress through the programs more quickly. As a result, they have higher achievement scores.

[1] The author thanks Carrie Thomas Beck, Ashly Cupit, Gary Davis, Kurt Engelmann, Jerry Silbert, and Tim Wood for helpful comments on drafts of this paper. Any remaining errors and all opinions are the responsibility of the author.

Figure 1:
How Teachers' Skills Influence Student Achievement in DI Programs

Teachers' Skills and Fidelity → Students' Progress at Mastery → Higher Student Achievement

This report focuses on two general factors that influence the development of teachers' skills and fidelity. The first is the nature of the environment and, specifically, administrative practices that either hinder or support appropriate teaching practices. Building administrators, most often school principals, are generally key players in determining the nature of a school's teaching environment (Engelmann, 2014). The first part of this report summarizes data from a variety of settings that illustrate the way in which administrative practices can hinder or support appropriate teaching practices and student achievement.

The second general factor involves the extent to which the best results with DI come as schools, teachers, and students have more experience with the program. Teaching DI well is a technical and complex process. Like any other complicated skill, it requires training, practice and time to develop. Over time, with appropriate training and coaching, teachers become more skilled at delivering the programs, schools become more skilled at implementing the procedures that best support instruction, and students have higher levels of achievement. At the same time, students benefit when they have been taught with DI programs from the earliest grades. In short, the highest levels of student achievement appear when teachers have developed their skills in using Direct Instruction, when the recommended procedures are fully implemented into school routines, and when students have been exposed to these environments from the earliest grades (Engelmann & Engelmann, 2004). The second part of this report provides examples of these findings. Details on the data are provided in the Appendix. All school names are pseudonyms.

The Impact of Fidelity and Administrative Decisions on Student Achievement

Schools administrators head complex organizations and make many decisions that can affect the quality of the teaching that students receive and their achievement gains. This section provides four examples of how administrative decisions that counter implementation recommendations associated with Direct Instruction programs can significantly affect teachers' fidelity and student achievement.

Teacher Training and Assignment – Ash Elementary

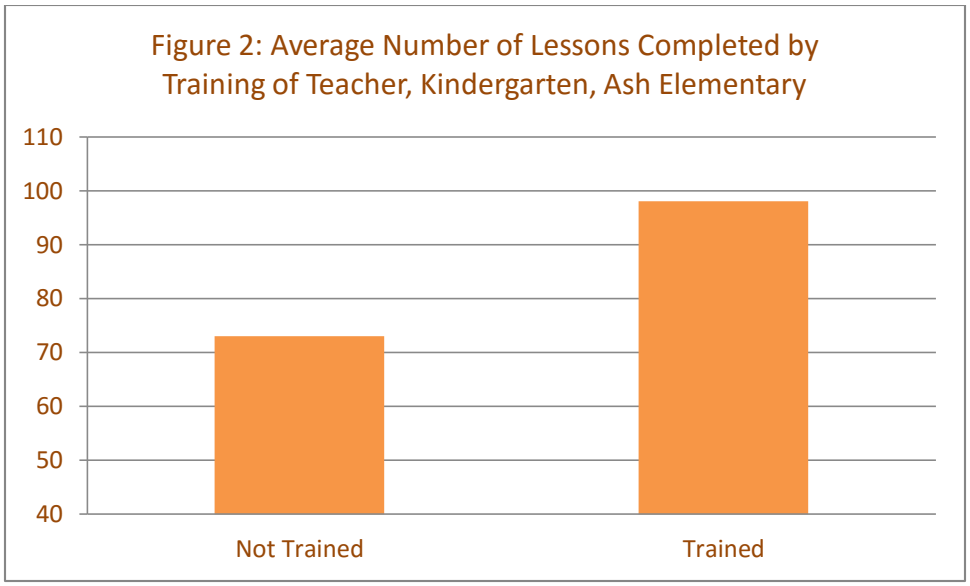
One of the most important decisions school principals can make is the assignment of teachers to grade levels and instructional groups. Given the highly technical nature of DI, training in the appropriate use of the programs is very important for promoting high fidelity. Much of this

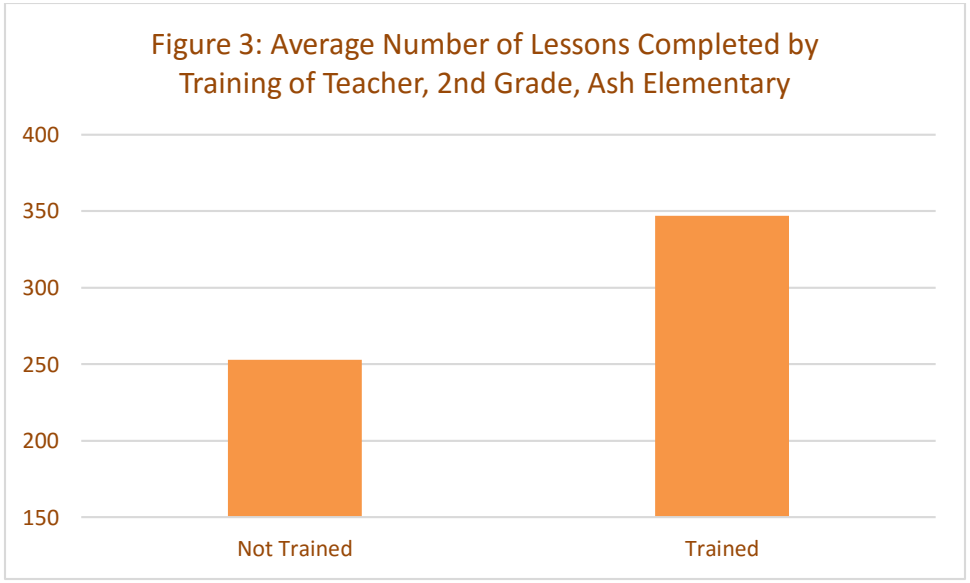
training is specifically oriented to the details involved with a given subject matter and grade level program (Engelmann, 2014). Data from a school in the southeastern United States illustrate the negative impact of assigning teachers to a level for which they have not been trained.

Ash Elementary is a very high poverty school with a high minority enrollment. It implemented the DI program *Reading Mastery Signature Edition* (*RMSE*) in some kindergarten and first grade classrooms in the fall of 2013. At both grade levels the students assigned to *RMSE* began the 2013-14 year with substantially lower scores than students in other classrooms at their grade level in the school. Yet, by the start of the next academic year (fall 2014), they had significantly higher scores (Tables A-1 and A-2). Based on these results the school opted to use *RMSE* for reading instruction in all K-2 classrooms in 2014-15. During that school year, students' scores continued, on average, to improve (Table A-3). Yet the rate of progress varied substantially from one classroom to another. And for some of the students the gains they had made in the previous year from their work with *RMSE* disappeared.

The major reason for these differences was teacher assignment. At the start of the 2014-15 school year, Ash Elementary's principal re-assigned teachers from one grade to another. The result was that one of the kindergarten teachers and one of the second grade teachers had no prior training in using *RMSE*. Understandably, their skills in implementing the program were markedly lower than those of other teachers. The impact of this decision on students' lesson progress is illustrated in Figures 2 and 3.

The figures compare the number of *RMSE* lessons students completed, at mastery, from fall to winter. At both grade levels students of the untrained teachers had much lower rates of progression at mastery. For instance, the benchmark for acceptable progress for kindergarten students from fall to winter is the completion of 80 lessons. Students with trained teachers completed, on average, 98 lessons at mastery by that point, almost 25 percent higher than the goal. In contrast, the kindergarten students with the untrained teacher only completed 73 lessons at mastery. Similar results appeared for the second grade students. Those with the trained teachers were much more likely to be near the established goal for their grade level. The significant results for the effect of teacher training held when students' levels of prior achievement were controlled (Table A-4).





Note: The number of lessons shown in the vertical axis represent the cumulative lessons completed within the RMSE program.

As would be expected by the logic model shown in Figure 1, students of the teachers with greater skills not only had better lesson progress, but they also had higher achievement scores (Tables A-5 and A-6). Statistical analyses controlled for the skills that students had at the start of the school year. This control is important because it allows estimation of the extent to which having a trained teacher provided "added value" to students' achievement. The results were clear-cut. When students' initial scores were controlled, the influence of having a trained and more highly skilled teacher was strong and significant and more important than the influence of

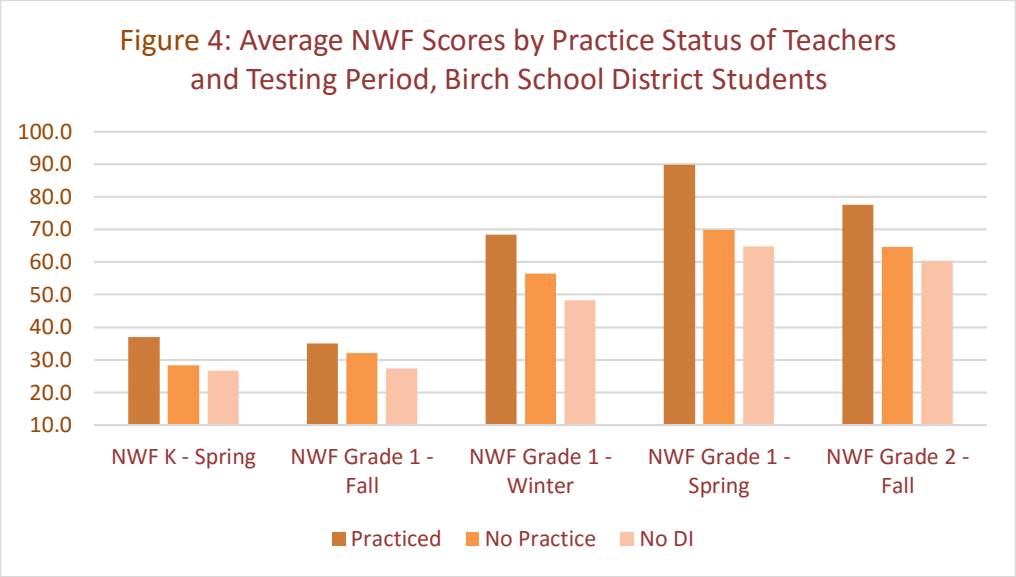
initial scores. The negative impact of being assigned to an untrained teacher was almost three times as strong as the positive impact of higher skills in the fall. In other words, if two students began the school year with equivalent achievement, the one assigned to an untrained teacher would, on average, make substantially less progress through the curriculum and have significantly lower scores than the other student (Table A-7).

Helping Teachers Develop Their Skills – Birch School District

Teaching Direct Instruction programs is technical and involved. It requires not only training, but practice and careful preparation for each lesson. One of the most important elements is regular rehearsal of lessons. Such practice helps teachers learn to present the material easily and fluently so that they can give full attention to their students during the lessons. Thus, one of the key elements of good implementations is providing time for teachers to practice their teaching formats.

The experience of Birch Schools (another pseudonym), located in the rural Midwest, shows how important this practice can be. For the first two years that the schools used the DI curriculum they did not provide time for teachers to practice their teaching formats. When they began to provide teachers practice time, the reading skills of their students increased markedly.

Figure 4 illustrates this change. It shows average scores of students on the Nonsense Word Fluency (NWF) measure in the Dynamic Indicators of Basic Literacy Skills (DIBELS) curriculum based measurement system. Results are given for three cohorts of students: 1) those who began kindergarten when the teachers were given time to practice (n=149), 2) those who began kindergarten when their teachers were using DI but were not given practice time (n=361), and those who began kindergarten before the school was using DI (n=216). In all of the comparisons the highest average scores were obtained by the students taught with DI and whose teachers were given practice time. The lowest scores were obtained by those who did not have instruction with DI programs. Most of the comparisons were statistically and educationally significant and appeared with other measures (Table A-8).



Note: "Practiced" refers to students that had DI and whose teachers were given practice time, "no practice" refers to students that had DI but their teachers did not have practice time, and "no DI" refers to students that did not have DI in Kindergarten.

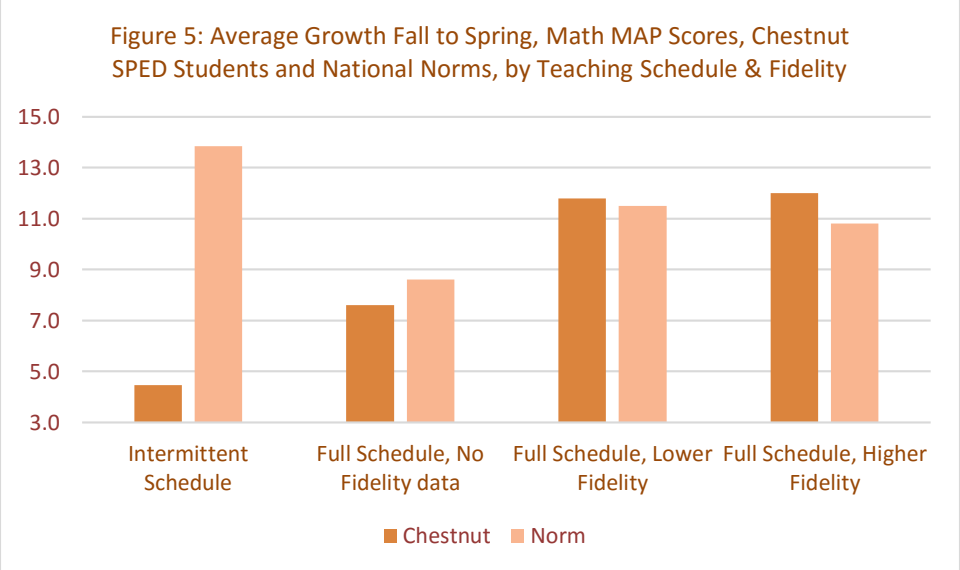
Scheduling Instructional Time – Chestnut Schools

In addition to determining teacher assignments and the time teachers have for preparation, administrators can control the time devoted to teaching. The implementation guidelines for DI programs stress the importance of maintaining a regular schedule of teaching. Students learn the most when they are regularly exposed to the material for the recommended amount of time. It is far easier to retain previous learning and to advance more quickly when students are exposed to the program each day of the school week. Data regarding special education students in 13 schools in the upper Midwest illustrate the importance of implementing programs thoroughly and completely within the school schedule. They also reinforce the importance of helping teachers implement the programs with fidelity.

Special education teachers in the set of schools we have termed Chestnut began using the Direct Instruction mathematics program *Connecting Math Concepts: Comprehensive Edition (CMCCE)* in the fall of 2014. Some schools fully implemented the program, with regular use all days of the week and for the specified amount of time. Others, however, partially implemented the program, using it only some days of the week and with a variable schedule. Mathematics achievement was assessed with the nationally normed Measures of Academic Progress (MAP) from the Northwest Evaluation Association (NWEA). Students were in kindergarten to grade six, although the majority were in grades two to four. Data were also available on the implementation fidelity of teachers in most of the schools. None of the teachers in the schools with only intermittent implementation had higher fidelity scores.

Figure 5 shows the average growth from fall to spring for special education students in four different groups of Chestnut students: 1) those in the schools that taught the program intermittently, 2) those in schools with regular implementation, but no fidelity data for their teachers, 3) students in schools with regular implementation and teachers with lower fidelity and 4) students in schools with regular implementation and teachers with high fidelity. The average growth of these students is compared to the average growth of students in their grade in the national normative sample. Note that the comparison group was not limited to special education students.

Students in the schools with intermittent use of the program had the lowest rate of growth, while those with a regular schedule and teachers who implemented with high fidelity had the highest rate of growth. The special education students in schools with a regular schedule of implementation had average growth rates that were slightly higher than that of the national normative group, while those in the schools with intermittent scheduling had rates of growth that were much lower than the norm. In other words, during the school year special education students in the schools that did not have a regular schedule of math instruction fell farther behind other students. In contrast, the special education students in the schools with regular schedules had a rate of growth that was close to or exceeded that of the national norming sample, a group comprised primarily of non-special education students (Table A-9).



Note: The norm group included all students, not just special education students.

Double Dosing for Students at Risk – Elm and Maple Schools

Schools often use Direct Instruction programs to help students with low skill levels catch up with their peers. The National Institute for Direct Instruction (NIFDI) recommends that students at risk in the lower grades have a second reading period to help them catch up with their peers, a practice often referred to as "double dosing." This section compares outcomes for at-risk students in a school that did not use double dosing (Elm Elementary) and a school that followed this recommendation (Maple Elementary). Both schools used a curriculum-based measure of Oral Reading Fluency (ORF). Each school also had data from a valid comparison group that did not have DI. For Elm Elementary the comparison was the next older cohort, and for Maple Elementary it was a nearby school. In both Elm and Maple elementary, the students began their work with DI in kindergarten. There were no significant differences in initial skills or demographic variables between the DI and non-DI students at that time.

Figure 6 shows ORF scores of first graders in Elm Elementary, which did not employ double dosing. Data are shown for two cohorts – one that had DI and one that did not – and, within each cohort, for students at high risk of learning problems and those with lower risk.[2] The bars on the left show data for the fall and the bars on the right show data for the spring. Within each time period the data for the high risk students are on the left. In the fall the DI students had slightly lower scores than those in the non-DI cohort in both "at risk" categories. By spring, the DI students had higher scores in both categories. However, the gap between those deemed at risk and the general population did not decline. In other words, the at-risk students at Elm Elementary were not catching up with their peers (Table A-10).

Figure 7 shows changes over a similar time period for students in Maple Elementary, which employed the recommended double dosing practice for students deemed at risk.[3] As with Figure 7, the bars on the left show data for the fall and the bars on the right show data for the spring of first grade; and, within each time period, the bars on the left give data for the at-risk students. In the fall, the lowest average ORF scores were obtained by the at-risk students in the DI school, a pattern similar to that at Elm Elementary. However, the results in the spring were very different. At the spring testing, the ORF scores of the at-risk students in Maple school were actually higher than those for the general education population in the non-DI group. This pattern of results, with the special education students in the DI school having reading skills equivalent to the general education students in the non-DI school, remained throughout the elementary years (Tables A-11 and A-12).

[2] For Elm Elementary "at risk" was defined as being in the lower third of the distribution at the start of the school year. Oral Reading Fluency was measured with the AIMSWeb system.

[3] At-risk students at Maple Elementary were those who were deemed as needing special education services. Thus, these students were, arguably, potentially even more "at-risk" than those at Elm Elementary.

Figure 6: AIMSWeb ORF Scores First Grade, by DI Group and At-Risk Status, Fall and Spring, Elm Elementary (No Double Dosing)

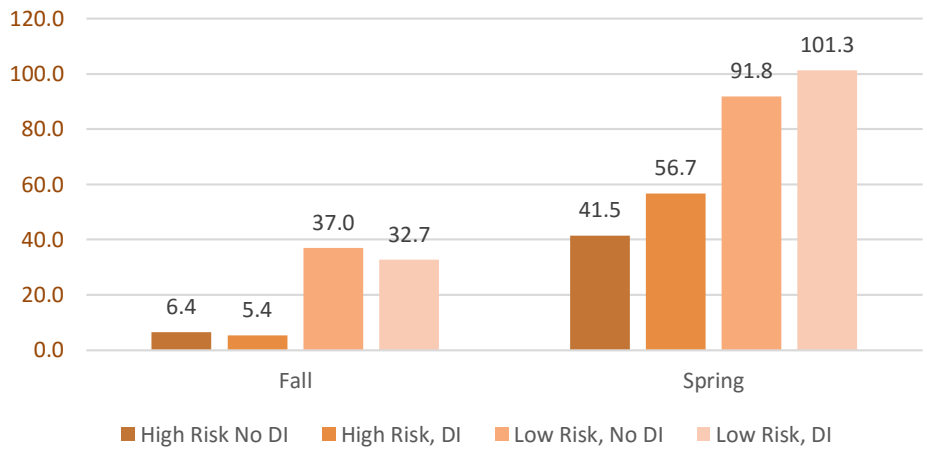
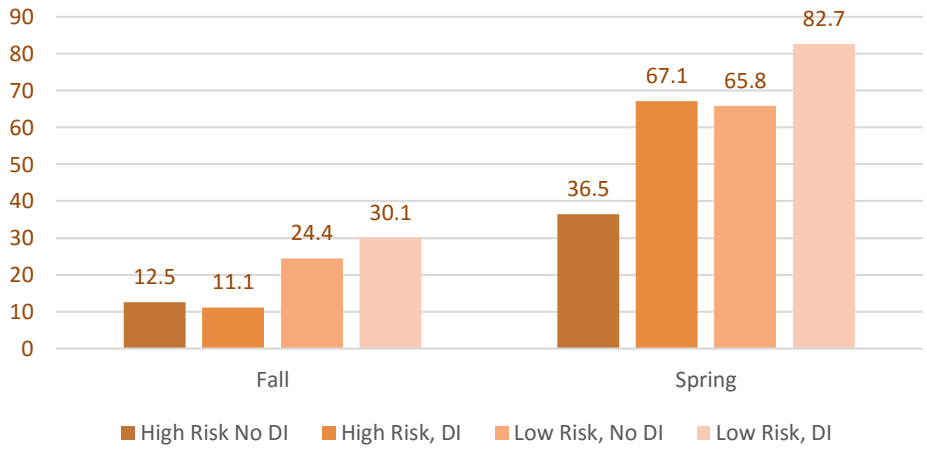


Figure 7: DIBELS ORF Scores, by DI Group and Risk Status, Fall and Spring, Maple Elementary (Used Double Dosing)



To summarize, in Elm Elementary, which did not use double dosing, the at-risk students fell further behind their general education peers over the course of the school year. In contrast, in Maple Elementary, which used the recommended practice, the scores of the at-risk students increased more rapidly than those of their general education peers. In the double dosing school the students deemed to be in special education at the start of the year had reading skills at the end of the school year that were indistinguishable from the general education students in an alternative reading program.

The Best Results Come with Time

While many of the examples above and in the general literature involve results after exposure to DI programs for a year or even less, it is important to remember that it takes time for schools to fully incorporate all of the procedures associated with a good implementation. The best results occur when DI is institutionalized within the school setting, fully familiar to students and teachers, and in place long enough that a student who began at the school in kindergarten would have experienced the model throughout the elementary years. Based on decades of experience and data from numerous implementations, Engelmann and Engelmann suggest that while a site may implement Direct Instruction in as little as two years the full potential of the model to improve student performance does not occur "until about the sixth year of implementation" (2004, p. 117). The sections below provide data that show how student achievement improves as DI implementations become stabilized, how teacher fidelity and the achievement of students increases over time, and the importance of students experiencing Direct Instruction throughout their elementary career. The highest rates of achievement occur for students in schools that have stabilized DI implementations, whose teachers have multiple years of experience teaching the program, and who begin learning with DI programs in kindergarten.

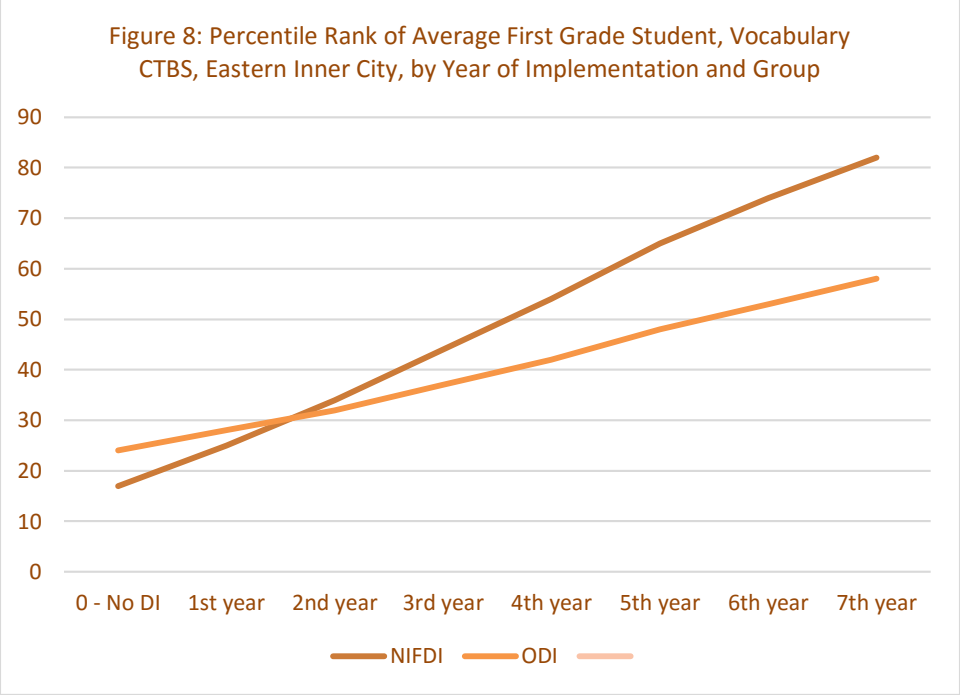
Student Achievement Increases Over Time

Data from two sites, one urban and one rural, illustrate the extent to which student achievement improves over time. While both sites experienced gains in student achievement in the first few years of implementation, the strongest gains only appeared as the implementation became stabilized within the schools.

A Large Urban District. The first example comes from a large eastern U.S. urban school district (Stockard, 2011a). While results from this site illustrate the way in which achievement increases over time, they also show the importance of high fidelity. The district has a very high rate of poverty and a high minority enrollment. As part of a major school reform initiative, a charter school system within the district implemented Direct Instruction. Some of the DI schools worked with the National Institute for Direct Instruction (NIFDI) and followed all of the implementation guidelines, while others did not fully abide by the suggestions. Reading achievement was measured with the Comprehensive Test of Basic Skills (CTBS).

Figure 8 shows the percentile rank associated with the average CTBS vocabulary score of first grade students from before the implementations began to the seventh year of implementation for these two groups of schools. It shows how the highest levels of achievement appeared in the latest years of implementation. The figure also shows the importance of strong implementation fidelity. The change over time was stronger for the NIFDI supported schools,

with high levels of implementation fidelity, than for the group of schools that used DI programs but had less rigorous implementations. Most notably before implementing DI the NIFDI schools had lower levels of average achievement than the other group, but after seven years of implementation had much higher levels. Similar results occurred with the CTBS measure of reading comprehension (Stockard, 2011a and Table A-13).

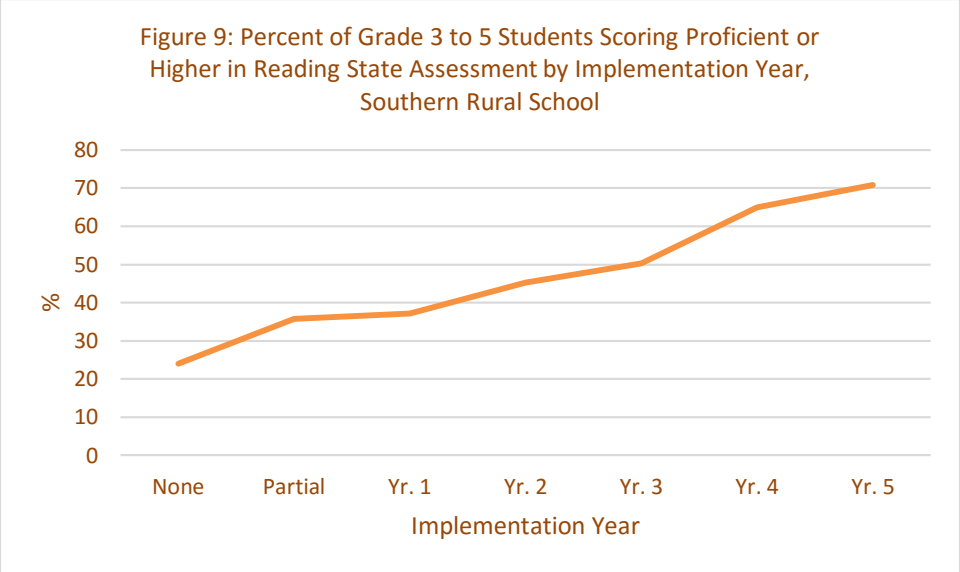


Note: ODI refers to "Other Direct Instruction" schools, which used DI programs but did not necessarily follow all of the recommended implementation procedures.

A Small Rural Community. Another example comes from an elementary school in the rural south, studied by Michael Vitale and Beverly Joseph (2008). Ninety-eight percent of the students were minorities and 98 percent received free or reduced price lunch. The school implemented Direct Instruction midway through a school year after being identified, from its performance on the state-administered accountability tests, as one of the lowest-performing schools in the state. In the following year DI was fully implemented. Figure 9 reports the percentage of third to fifth grade students who were rated as proficient or above on the state assessment from the year before implementation to the fifth full year of implementation. While the graph shows gradual improvement over the time period, the highest levels of achievement only appeared in the fifth year of implementation (Table A-14). Vitale and Joseph commented on the time that it took to achieve these high levels of achievement. As they put it,

> it is important to note that the actual trend in improved student achievement required a 3 to 4 year time period to emerge….With this emerging achievement pattern, an evaluation of the effectiveness of DI based on the initial years of implementation would have reached a different conclusion than an evaluation conducted over the entire implementation period (Vitale & Joseph, 2008, p. 8).

The same comment could, of course, be made about the trend in the inner city schools shown in Figure 8.



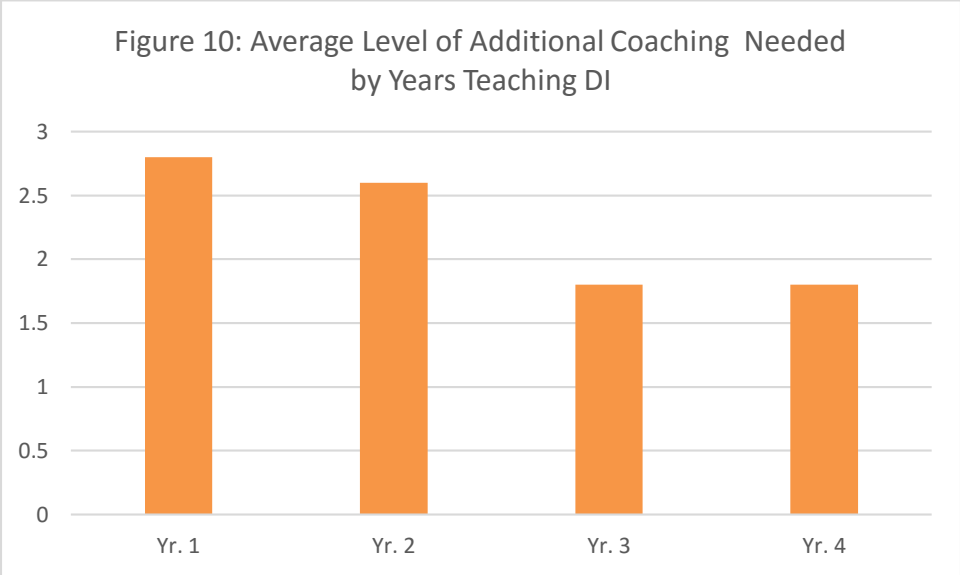
Note: In the year labeled "partial," DI programs were implemented for only part of the year. Year one is the first year in which the programs were implemented throughout the year.

Teachers' Skills and Student Achievement Improve over Time

A major reason that student achievement improves over time in well implemented DI schools is that teachers become more comfortable and effective in teaching the programs as they receive appropriate training and coaching. Again, examples from two settings, one urban and one rural, illustrate this phenomenon.

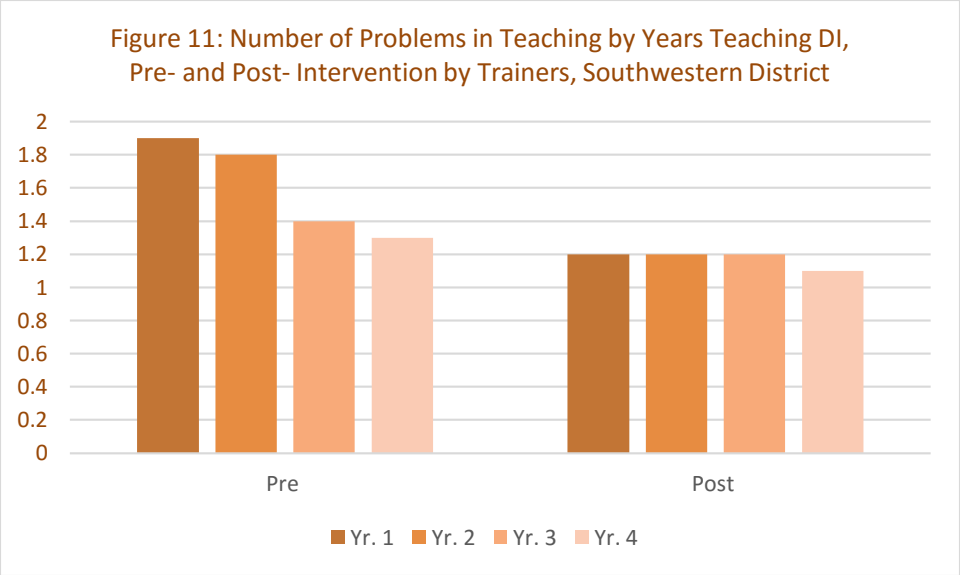
A Large Southwestern City. Coleen Carlson and David Francis's (2002) report of a large scale implementation of DI programs in a southwestern city documented the way in which in-classroom coaching can improve teachers' instructional practices and how, as time goes on, less coaching and intervention by trainers is needed. Approximately 300 teachers were involved in the project, all of whom received initial training and coaching as well as additional coaching during the school year that included attention to specific elements of teaching the individual programs.

Trainers reported the level of additional coaching that they gave to teachers during the year, and Figure 10 reports the average level of coaching received by teachers with different levels of experience teaching DI. Higher values indicate that more coaching was required. Data in the Figure show that teachers in their first year of teaching required the most additional help , while those with three or four years of experience required substantially less. In other words, over time teachers needed less help in implementing the programs with fidelity (Table A-15).



Source: Carlson & Francis, 2002, p. 161

Observations of the teachers confirmed that their skills had improved with training and intervention. Data in figure 11 illustrate the number of problems in teaching the material that were observed by the trainers. The bars on the left are ratings before extra coaching. The bars on the right are the ratings given to the teachers after help from the trainers. It can be seen that the teachers had more problems in their presentation skills before intervention, but fewer after intervention. In addition, the changes were strongest for those in their first and second years of teaching the program (the first two bars within each sub-set). In fact, with practice and support, by the end of the year the trainers reported relatively few problems for any of the teachers. In other words, training made a difference in teachers' skill levels. And, as teachers had more experience with the program they had similar levels of skill.

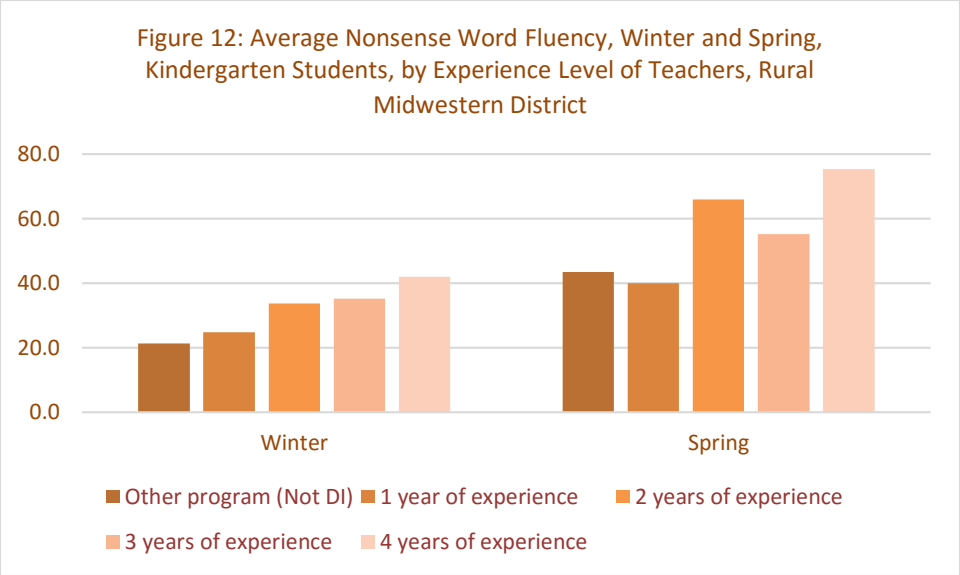


Source: Carlson & Francis, 2002, p. 161. Data are number of problems reported by trainers. Similar results appeared when teaching was assessed by independent raters.

Using complex statistical analyses, Carlson and Francis found that students' achievement was significantly related to the extent to which teachers improved their skills. This result appeared at all three grade levels examined and with measures of both word reading and comprehension (Table A-16).

A Rural Midwestern District. The association between coaching, teachers' improved skills and student achievement has appeared in a number of settings other than that studied by Carlson and Francis. The results in Figure 12 illustrate this process using data from a district in the rural Midwest. Five kindergarten teachers worked in the district before DI was implemented and continued to work there for up to four more years. During that time they received training and coaching from the National Institute for Direct Instruction. The graph shows the average DIBELS Nonsense Word Fluency (NWF) scores of their students from the year before they began using DI to the time when they had four years of experience with the program. The highest levels of student achievement appeared after the teachers had become fully experienced with the program (Table A-17).

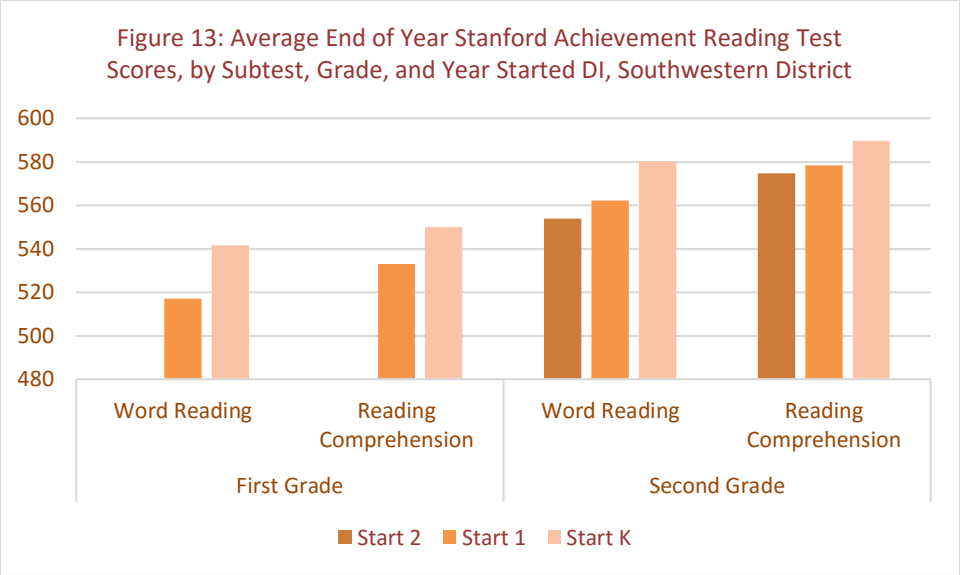
It is important to remember that these data are from the same teachers – both before and after training. Thus, the analysis controls for a variety of individual and school factors that can affect success. It provides important data to support the contention that, with training, practice, and support, all teachers can improve their skills and thus help their students have higher levels of achievement.



Students Have the Best Outcomes When They Start DI in Kindergarten

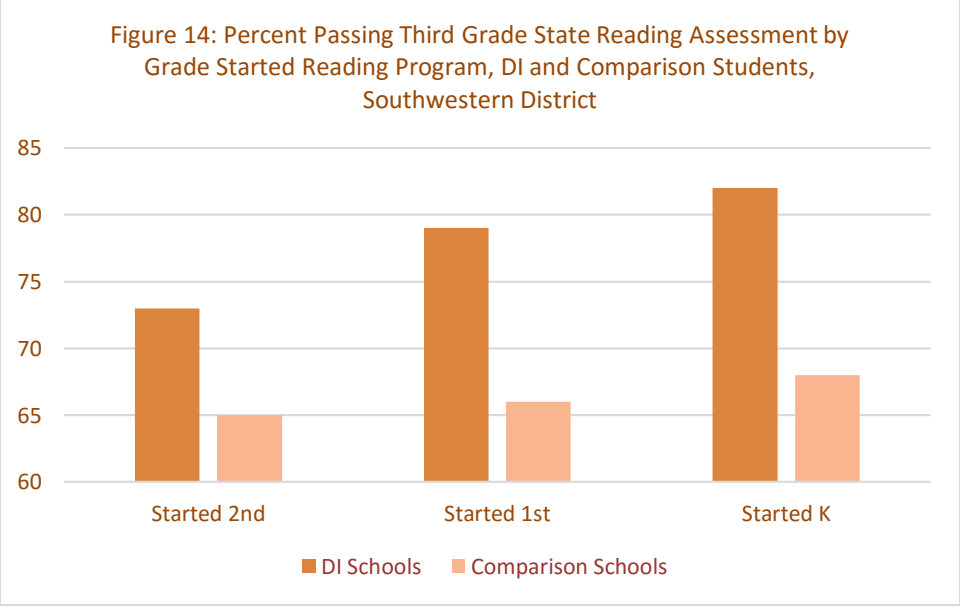
The previous sections have shown how improving teacher skills is a very important reason that schools have higher achievement after DI has been implemented for a number years. Yet, students' experience with DI is also important. Data from a variety of sources show that students who begin their experience with DI in kindergarten have significantly higher rates of achievement than those who start at later grades. Examples from two sites are given below.

Data from a Large Southwestern District. One example of this effect comes from the work of Carlson and Francis (2002) in the large southwestern district described above. The project lasted for three years and targeted students in grades K-2. Thus, they were able to compare achievement scores of students who began DI in kindergarten with those who started the program at later grades. Figure 13 shows the average Stanford Achievement Test (SAT) Scale Scores for first and second grade students on measures of word reading and reading comprehension. Data are given for those who started *Reading Mastery (RM*) in kindergarten, in grade 1 and (for the 2nd grade data) in grade 2. For each comparison the highest scores appear for those who began the program in K (the far right hand bar in each comparison), followed by those who started in grade 1 and then by those who started in grade 2 (Table A-18).



Source: Carlson and Francis, 2002

Similar results appeared on the students' scores on the state-wide reading assessment given in third grade. Figure 14 shows the percentage of students who scored at the proficient level or higher by the grade at which they began their work with DI. The figure also includes the percentage for students in comparison schools (matched on prior achievement and various

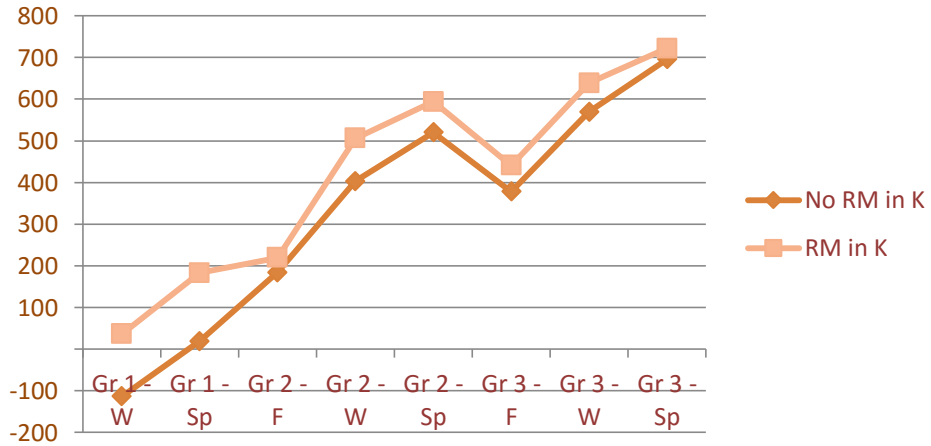


demographic factors). As in other studies, the DI students had significantly higher state assessment scores than those in comparison schools that did not have the program. However the differences with the comparison students were significantly larger when the students began

DI at an earlier grade. The students who began DI in kindergarten had significantly higher scores than those who began the program in higher grades. Similarly, those who started in grade 1 had higher scores than those who began in grade 2. There were no differences in the comparison schools between students who started the program at different grade levels (Table A-19).

Rural Midwestern District. While the data in Figures 13 and 14 show comparisons at just one time point, the effect of starting DI in kindergarten is long-lasting and cumulative. Figure 15 shows the change in Oral Reading Fluency Lexile scores of students in a Midwestern district. (Lexiles are a developmental scale of reading that ranges from less than zero for those who are just beginning to read to above 1700 for advanced readers. Thus they adjust for the different content used in the ORF at each grade level (MetaMetrics 2009).) One line in Figure 16 shows the growth of students who started *Reading Mastery (RM)* in Kindergarten while the other shows the growth of those who began the program in first grade. As would be expected, at the middle of first grade the students who had *RM* in kindergarten had significantly stronger reading skills. While the difference declined at the start of grade 2, the gap widened at later testing periods and did not decline significantly until the end of Grade 3 (Table A-20). Thus, the students who began DI in kindergarten had advantages at each grade level.

Figure 15: Growth in ORF Lexiles Mid Grade 1 to End Grade 3 by Having DI in K

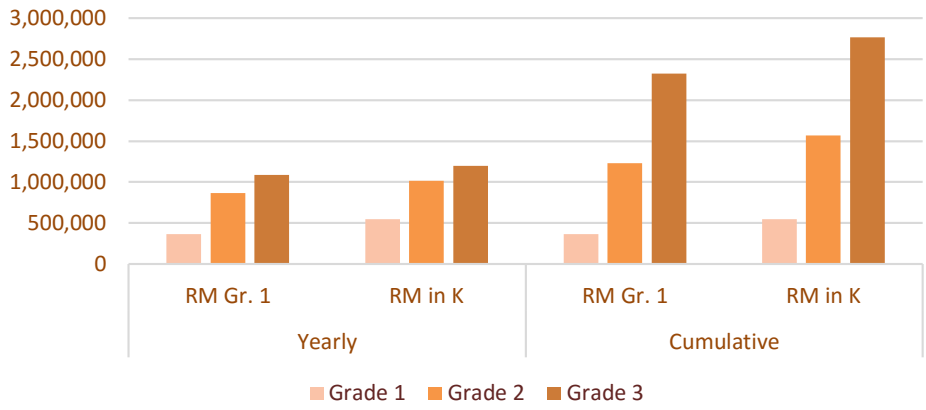


The Cumulative Impact of Starting Early with DI. It is important to remember that the advantage in reading skills of the students who began their work with *RM* in kindergarten is cumulative in nature. A number of authors have noted the importance of "reading volume," cumulative exposure to the written word, for future academic success and intellectual development. Because students with early exposure to *Reading Mastery* have more advanced

reading skills at younger ages they are exposed to more material throughout the primary years. Thus, their total reading volume would be expected to be substantially higher.

Figure 16 shows estimates of the average yearly and cumulative reading volume of students in the Midwestern district through the primary years. By the end of first grade the students who started *RM* in first grade would have been exposed to about two-thirds of the written material of those who began the program in kindergarten. While the gap in yearly volume gradually narrowed in the second and third grade, it remained throughout the primary years. Most important, the cumulative impact continued to increase over second and third grade. The final estimate of differences in reading volume is far from trivial. By the end of third grade the average student who began work with *RM* in kindergarten would have read over 400,000 more words than the student who started the program in grade 1 (Table A-21 and Stockard & Engelmann, 2010).

Figure 16: Estimated Reading Volume, Yearly and Cumulative, by Grade and Year Started DI, Midwestern District



Note: Some data taken from Stockard and Engelmann, 2010

Summary

The results described above have shown how administrators' decisions regarding the implementation of Direct Instruction programs can have a strong influence on students' achievement. Administrators are the ones who can ensure that teachers are assigned to grades and programs for which they have been trained, that teachers receive ongoing opportunities to practice so they can implement the programs with high fidelity, and that all students have the appropriate amount of instructional time. At the same time, the data show the importance of patience and persistence. While student achievement will almost always increase from previous levels within a short time after beginning work with Direct Instruction, it takes years for the full

benefits to appear. Learning to teach well involves extensive training, coaching and practice. Like any skill teachers become better over time if they have proper training and support. Without such support teachers' skills can stagnate. Even worse, they can develop practices that are detrimental to their students. In addition, students benefit much more when they are in schools where DI implementations have been in place long enough for all staff and students to be familiar with them. Those who benefit the most are the students who are exposed to the program from Kindergarten, attend schools that have used DI for multiple years and who have teachers with multiple years of experience teaching Direct Instruction. These students have higher achievement at each grade level and a greater accumulation of knowledge and skills than their peers.

Appendix

This appendix provides detailed results for the findings summarized in the text. The order parallels the order of topics covered in the body of the report. Supporting tables are at the end.

Teacher Training and Assignment – Ash Elementary

Ash Elementary implemented the DI program *Reading Mastery Signature Edition* (*RMSE*) in some of their kindergarten and first grade classrooms in the fall of 2013. At that time reading skills were measured by a district constructed measure for both the kindergarten and first grade students and by the Iowa Test of Basic Skills for the first graders. In the spring of 2014 and fall of 2015 the I-Ready measurement system was used as well as students' grade level placement within *RMSE*.

Changes in Achievement with Implementation of RMSE. Table A-1 compares scores on all available assessments for those who had *RMSE* in 2013-14 and those who did not from the fall 2013 to the fall 2014 testing periods. Data in the first panel are for the cohort entering kindergarten in 2013-14 and data in the second panel are for those who entered first grade in that year. The first two columns give means and standard deviations for each variable for the students who did not have *RMSE* in the 2013-14 year and the next two columns give the results for those who did have *RMSE*. The fifth column of data gives the t-ratio testing the difference between the two means, the sixth column gives the associated probability, and the final column reports the effect size, Cohen's d, associated with the difference. Among kindergarten students (data shown in the first panel of Table A-1), those exposed to *RMSE* had lower scores in the fall before starting the program (the first line of data), but higher scores in the spring and subsequent fall. All but one of the associated effect sizes surpassed the .25 level typically used to denote educational significance. Yet, only one of the results was statistically significant, no doubt because of the relatively small sample size. For the first grade students (data in the second panel), those exposed to *RMSE* had lower scores on all three of the fall measures as well as the spring measure. However, they had higher scores on the fall 2014 measures and substantially less loss over the summer months.

Table A-2 reports effect sizes associated with the difference between the *RMSE* and non-*RMSE* groups in the change from the fall 2013 testing to the spring and fall testing periods in 2014. In other words, it states in standard deviation terms the extent to which the gap between the *RMSE* group and the non-*RMSE* students changed over the given time period. A positive value indicates that the *RMSE* students had a greater increase than the non-*RMSE* students. For the 2013 kindergarten cohort all of the effect sizes were positive and substantially larger than .25. Three of the four comparisons were statistically significant. For the first grade cohort the effect sizes with three of the four measures (9 of the 12 effects) were positive. However, the only

effects associated with the changes that were statistically significant were those involving changes to the fall 2014 assessments.

The I-Ready assessment places students in three tiers, with placement in Tier 1 indicating that they are at or above grade level. Table A-3 reports the change in tier placement, as indicated by the I-Ready assessment, over the 2014-15 school year, when all students were exposed to *RMSE*. Over time, the percentage of students at Tier 1 increased while those at Tier 3, and in need of intensive assistance, decreased. Paired t-tests were used to compare students' placement at Tier 1 in one time period with the next. In all but one comparison the upward movement was significant. The exception was for the change from winter to spring for students in Grade 2, where the percentage at Tier 1 remained unchanged. There was however a drop in the percentage of students at Tier 3 for that grade and time period.

Teacher Training, Lesson Progress and Student Achievement. As noted in the text, in 2014-15 untrained teachers were assigned to grades K and 2. Table A-4 reports the average lesson placement at mastery in *RMSE* in winter 2015 for students assigned to trained and untrained teachers in these grades. Mean values and standard deviations are given as well as t-tests and Cohen's d, a standard measure of effect size. In both comparisons the students who had trained teachers had greater lesson progress than those who did not. Both t-ratios were significant and the effect sizes were large and educationally significant.

Table A-5 gives the results of testing the model displayed in Figure 1 using multivariate analyses and data on kindergarten and first grade students from the 2014-15 school year. (While all of the first grade teachers had been trained, they, according to skilled observers, differed in their implementation skills. The measure of first grade teachers' skills is based on the observations.) Students' spring 2015 I-Ready scale scores were regressed on their I-Ready scores in fall 2014, the skills of their teachers, and their point within the *RMSE* curriculum in spring 2015. Three models were tested: Model 1 predicts spring scores from fall scores and thus controls for reading achievement at the start of the school year. Model 2 adds teacher skills, and model 3 adds students' placement in *RMSE* in the spring. Recall that the causal chain depicted in Figure One suggests that the impact of teacher skills on student achievement occurs because students of more highly skilled teachers make greater progress at mastery through the programs. If this diagram were supported, the regression results would show that the impact of teachers' skills on achievement scores in Model 2 would decline markedly when lesson progress at mastery is added in Model 3 and that there would be strong and significant influences of lesson progress. As noted in the text it is important to include students' fall I-Ready scores as a control variable because it allows estimation of the extent to which the other two variables – teachers' skills and progress through the curriculum – provide "added value" to estimates of students'

achievement. For each model the unstandardized regression coefficients (b), the associated t-ratio, and the standardized coefficient (beta) are given. The unstandardized coefficients can be compared across models; the standardized coefficients can be compared within a given model (from one variable to another).

For both grade levels the results support the model depicted in Figure 1. Results in Models 1 and 2 show that higher prior achievement (as measured by fall scores) and having a more highly skilled teacher significantly influence students test scores. When lesson progress at mastery is added as a predictor in Model 3, the influence of fall scores declines markedly and the influence of teachers' skills declines to insignificance. In other words, the influence of having a skilled teacher can be explained by students' lesson progress at mastery. This result appears in both the analyses for kindergarten and first grade students (in the two panels). The standardized regression coefficients can be used to compare results within a given model and are equivalent to effect sizes. For instance, the value of .48 associated with lesson placement for kindergarten students indicates that, when compared to students with the same fall I-Ready scores and the same teacher, those who had a lesson placement that was one standard deviation higher would have spring I-Ready scores that were .48 of a standard deviation higher.[4]

Figure A-1 and Tables A-6 and A-7 provide data that illustrate the extent to which placement with an untrained teacher can counteract earlier achievement gains, using data for second graders in Ash Elementary in 2014-15. Figure A-1 illustrates the relationships tested: looking at the influence of fall placement in *RMSE*, having *RMSE* in first grade, and the training of the second grade teacher on students' winter placement in *RMSE*. The curved lines on the left are the correlations among the three predictors. They show that students who had *RMSE* in Grade 1 were placed at higher levels at the start of the year (r=.21), but were more likely to be placed with the untrained teacher (r = .25). At the same time, students with lower fall placements (the students with more needs) were also more likely to be placed with the untrained teacher (r = -.17)

The straight lines (single arrow points) leading from these three variables to students' winter placement indicate the "net" effect of each of the three variables on winter placement. (These are obtained from the regressions summarized in Table A-7.) The most important influence on students' placement in winter was having an untrained teacher. Students in this classroom

[4] The correlation between fall and spring scores was.56 for K, .65 for grade 1, and.86 for grade 2. The correlation of teachers' skills and spring scores was .16 for K, .42 for grade 1, and .28 for grade 2. The correlation of progress through the curriculum and spring scores was .66 for K, .79 for grade 1, and .53 for grade 2. The correlation of progress through the curriculum and teachers' skills was .25 for K and .57 for grade 1. The correlation could not be computed for grade 2 because the lower rated teacher stopped use of the program before the end of the school year.

were placed .66 of a standard deviation lower than those who were in other classes when their fall placement scores were controlled. Students with higher fall placements had higher placements in winter (.22 of a standard deviation), but this effect was a third of the size of the effect of the teacher. The positive impact of having *RMSE* in first grade on student achievement in second grade was washed out by the impact of the teacher.

Tables A-6 and A-7 provide the information that was used to construct Figure A-1. Table A-6 reports the correlations among the variables in the figure. Table A-7 reports the results of the regression used as the basis for the path diagram. The values in the first column of data in Table A-7 are standardized regression coefficients. (These are the coefficients on the straight lines in Figure A-1). They indicate the change in winter placement that would be predicted from a one standard deviation change in each of the variables listed when the other variables are constant or controlled. For instance, students who had fall placements one standard deviation above the mean in the fall would be expected to have winter placements .22 of a standard deviation above average if all had the same teacher and the same experience with *RMSE* in first grade. Similarly, those who had an untrained teacher would have scores that were .66 of a standard deviation below the average no matter what their fall placement was or their first grade experience with *RMSE*. The second column of data gives the correlation coefficients (taken from Table A-6). These are also in standard deviation units and can be directly compared to the first column. We can conceive of the correlations as indicating the "total effect" or relationship between two variables, the first column tells us the "direct effect." The third column is the difference between these two values or what is often called an "indirect effect." It shows the change that occurs in the predictions when you control for other effects. The most notable result here is the decline in the influence of fall placement in *RMSE* from .50 as a total effect to only .22 for a direct effect. The impact of higher placement in the fall was more than halved by the influence of having a less skilled teacher.

Helping Teachers Develop Their Skills – Birch Schools

Table A-8 gives the data for Birch School. The top panel of the table gives the data shown in Figure 4, showing the average value on NWF scores for each cohort and the corresponding analysis of variance results. The second panel provides information on pairwise comparisons among these three groups: Scheffe probabilities and effect sizes. The bottom panel expands this analysis by comparing ORF Lexile scores in first and second grade for four cohorts: 1) those who had DI beginning in grade K and whose teachers had practice time in both grades; 2) those that had DI in grade K and grade 1, but whose teachers only had practice time in grade 1; 3) those that had DI in both grade K and grade 1, but whose teachers did not have practice time in either grade; and 4) those that did not have DI in K, had DI in grade 1, but whose teachers were

not given practice time. In all comparisons the scores were highest for the students in group 1, followed by those in group 2.

Scheduling Instructional Time – Chestnut Schools

Table A-9 gives the data summarized in Figure 5 as well as data for the total group of schools that implemented *CMCCE* on a daily basis (combining the three fidelity groups). RIT scores, which are interval and cumulative in nature, were used in all analyses of the MAP. The norms were based on samples of over 20,000 students in each grade. The publishers of the MAP assessment provide national data (norms) on the scores at each testing period and the average growth of students from one testing period to the next for each grade.[5] These data were used to compute the extent to which each student's scores and growth varied from that for the national sample.

The first two rows of Table A-9 report the average MAP score of students in the groups at the fall and spring testing periods, and the third row reports the average change from fall to spring. The next three lines give the average scores for students in the national norming population within the same grade as those in each Chestnut school group. The next three lines report the average effect size within the groups. The effect size was calculated for each student, dividing the difference of an individual student's score (X) and the average score for the norm group at their age (μ) by the standard deviation of the norm group (б), a standard formula for a z-score:

$$Z = (X - \mu)/б.$$

These values were then averaged across the students within each subgroup. A value of 0 indicates no difference, on average, from the national sample, negative values indicate lower scores or less growth, and positive values indicate higher scores and more growth than would be expected given national norms. The final line in the table reports the number of students in each group.

Double Dosing for Students Needing Extra Help – Elm and Maple Schools

Table A-10 gives the mean and standard deviation of fall, winter, and spring AIMSWeb Oral Reading Fluency (ORF) scores for the two cohorts of first grade students in Elm School, one that had *RMSE* for reading instruction and one that did not. Results are given separately for the students with scores in the lower third and higher two-thirds of the distribution at the fall testing. The DI students at Elm school had DI in both kindergarten and grade one. The students in the comparison cohort did not have DI in either grade.

[5] Normative data were obtained from Northwest Evaluation Association (2011).

Table A-11 gives descriptive statistics for the special education and general education students at the two Maple Schools from the beginning of grade 1 to the end of grade 3. Table A-11 gives results for all the groups, with results for special education in the top panel and those for general education in the bottom panel. The first three columns give results for the control (non-DI) school and the next three columns give results for the DI (*RMSE*) school. The final column reports the effect size (Cohen's d) comparing the values of the two means at each testing period using the formula of the difference between the means divided by the common standard deviation. The students in the DI Maple School had the program beginning in grade K until Grade 3, the students in the comparison school did not have DI at any of these grades.

Table A-12 is a modified version of Tables A-10 and A-11 and has the comparative data that were used to construct Figures 6 and 7.

Improvements in Student Achievement over Time

Table A-13 shows the data from the urban schools used to construct Figure 8 and the corresponding data for the measure of comprehension. The values were obtained from mixed model regressions that controlled for the poverty level of the schools. All calculations involved NCE scores, which were then converted to percentiles. Values for the other DI schools (ODI) for year zero (before implementation) were predicted from the results of the regressions. (See Stockard, 2011a for details.)

Table A-14 shows the data for the rural southern school that was used to construct Figure 9. It also shows data for each of the grades included in the graph. (See Vitale & Joseph, 2008 for details.)

Teacher Training, Teacher Skills, and Student Achievement

Table A-15 has the data used to construct Figures 10 and 11. Carlson and Francis reported that "for all teachers, the more intervention provided, the more problematic behaviors decreased in the areas of classroom management, organization, and disciplinary technique or behavior management ($r=-.59$, $p< .0001$)" (p. 160). The measure of teacher skills in Table A-15 and Figure 11 was obtained from trainer ratings. However Carlson and Francis reported similar results when a measure of teacher skills was obtained from independent classroom observers ($r=-.29$, $p< .01$).

Table A-16 is taken from Carlson and Francis (2002, Table 12, p. 163). It gives the results of two-way analyses of variance testing the relationship of teachers' correction procedures (the manner in which they correct students' errors) to literacy skills. The main effects of fall and spring levels of correction examine the relationship of teachers' skills in the fall and spring to

students' skills. The interaction effects test the extent to which changes over time in these skills influence students' outcomes. All results were significant. Teachers who less often used appropriate correction procedures had students with lower levels of performance.

Table A-17 shows the data used to construct Figure 12. Five kindergarten teachers were included in the comparison: All had data for the year in which DI was not used. One had data for all four of the subsequent years, two had data for three of the subsequent years, and two had data for the first two subsequent years. A more extensive statistical examination, using linear growth models, can be found in Stockard (2011b). This analysis found the strong impact of Direct Instruction on students' reading skills persisted with controls for individual teacher.

Table A-18 gives data from Carlson and Francis (2002) that compares the Stanford Achievement Test scores of students who began their work with DI at different grade levels and was used to construct Figure 13. It also includes data from comparison students. The DI students had significantly higher scores than the other students in all comparisons, but the differences were largest for students who began their DI work in kindergarten. Carlson and Francis report extensive multivariate analyses that support the findings shown in the table. Table A-19 gives the data on state assessment scores that is depicted in Figure 14.

Table A-20 shows the data used to construct Figure 15 regarding the growth in ORF Lexile scores for students in the Midwestern district who began their work with *RMSE* in K or in Grade 1. Effect sizes that indicate the magnitude of the difference are also included.

Table A-21 shows the data used to construct Figure 16 regarding differences in cumulative reading volume for students beginning DI in K and grade 1. The calculations for daily volume were derived by multiplying the mid-year ORF score (the number of words read correctly in one minute) by 60, assuming that students would read for approximately 60 minutes each school day. To obtain the yearly estimate, the daily number was multiplied by 180, the number of days in a typical school year. Thus, calculations are estimates for the average student, assuming one hour of reading per day and limited to the 180 day school year. (See Stockard and Engelmann, 2010, p. 16).

Figure A-1

Path Diagram of Relationship of Winter Placement in RMSE with Fall Starting Lesson, Having RMSE in Grade 1, and Training of 2nd Grade Teacher, 2nd Graders, Ash Elementary

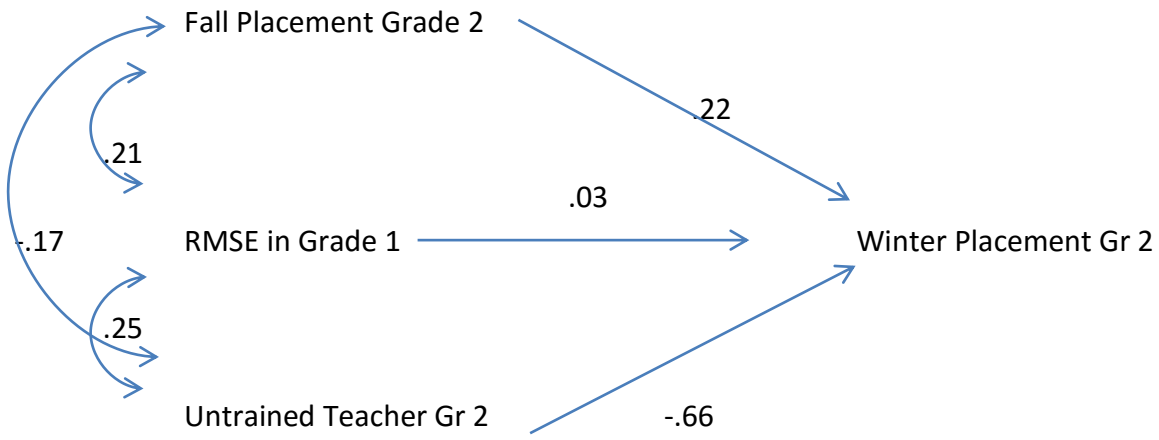


Table A-1

Reading Achievement Scores, Ash School, Fall 2013 to Fall 2014, by Group and Grade

District Measure Scores, I-Ready Scores, and Fall Placement RMSE, 2013-14 Kindergarten Cohort, by Group

	No *RMSE*		*RMSE*				
	Mean	SD	Mean	SD	t	p	d
District Measure Fall 2013	11.5	3.2	10.4	3.2	-1.02	0.84	-0.34
District Measure Spring 2014	16.3	2.4	17.5	3.2	1.22	0.11	0.40
IR Spring 2014	394.1	28.4	401.8	45.3	0.56	0.29	0.19
RMSE Placement, Fall 2014	43.3	52.6	177	53.9	7.56	<.0001	1.64
IR Fall 2014	388.5	26	404.5	42.4	1.26	0.11	0.41

Iowa Scores, District Measure Scores, I-Ready Scores, and Fall Placement RMSE, 2013-14 First Grade Cohort, by Group

	No *RMSE*		*RMSE*				
Measure	Mean	SD	Mean	SD	t	p	d
Iowa Vocab, Fall 2013	38.2	16.2	35.2	24.2	-0.36	0.64	-0.14
Iowa Read, Fall 2013	45.9	29.6	33.9	29.3	-0.14	0.87	-0.41
District Measure Fall 2013	13.5	2.3	11.9	3.6	-1.40	0.91	-0.49
District Measure Spring 2014	20.2	1.8	19.9	3.3	-0.27	0.61	-0.10
IR Spring 2014	465.5	33.2	444.9	31.3	-1.76	0.96	-0.63
RMSE Placement, Fall 2014	219.2	80.7	250.2	50	1.47	0.07	0.51
IR Fall 2014	449.5	35.1	462.1	37.1	0.98	0.17	0.35

Note: The sample is limited to students with data from fall 2013-14 and fall 2014. For the cohort of students who entered kindergarten in fall 2013, the *RMSE* group had 31 students on all measures except the Spring I-Ready, where there were 30. For the non-*RMSE* group in this cohort there were 13 students and they had data on all measures.For students who entered first grade in 2013-14 the *RMSE* group had 23 students with scores on the Spring I-Ready, 26 students with scores on the Iowa Reading, 28 students with scores on the Iowa Vocabulary, and 29 students on the other measures. For the non-*RMSE* students in the first grade cohort there were 11 students with data on all measures except the Iowa vocabulary measure, where there were 10 students.

Table A-2

Effect Size Associated with Difference between RMSE and non-RMSE Groups in Change, Fall to Spring by Grade and Cohort, Ash Elementary

2013-14 Kindergarten Cohort, Difference between Groups in Change from Fall 2013 District Measure						
Change from Fall 2013 to:	Effect Size		t-ratio			
Spring District Measure	0.74		2.23*			
Spring I-Ready	0.53		1.60			
Fall 2014 *RMSE* Placement	1.98		5.96***			
Fall 2014 I-Ready	0.75		2.26*			

1st Grade Cohort Difference between Groups in Change From Fall Measures						
	Fall, ITBS Vocabulary		Fall, ITBS Reading		Fall, District Measure	
Change from Fall 2013 to:	Size	t-ratio	Size	t-ratio	Size	t-ratio
Spring District Measure	0.04	0.11	0.31	0.85	0.39	1.06
Spring I-Ready	-0.49	-1.34	-0.22	-0.61	-0.22	-0.6
Fall 2014 *RMSE* Placement	0.65	1.77	0.92	2.51*	1.00	2.73*
Fall 2014 I-Ready	0.49	1.34	0.76	2.07*	0.84	2.29*

Note: The effect size (d) associated with the change was calculated by subtracting the effect size associated with the difference between the groups in the fall from the effect size associated with the difference between the groups in the spring. The t-ratio associated with the change was calculated by multiplying the effect size by the square root of the product of the sample size of the two groups (see note in Table A-1 for sample sizes) divided by the sum of the size of the two groups ($t = dV[(n1*n2)/(n1+n2)]$, n1 and n2 are the size of the two groups. Probability levels were adjusted for the degrees of freedom in the comparisons. *=p<.05, **=p<.01, ***=p<.001.

Table A-3

Percentage of Students at Tier 1 and Tier 3 from I-Ready Assessment, by Testing Time and Grade and Associated t-tests, Ash Elementary

	Kindergarten		Grade 1		Grade 2	
Test Time	Tier 1	Tier 3	Tier 1	Tier 3	Tier 1	Tier 3
Fall	18	0	18.6	6.8	29	18
Winter	47	0	39.0	3.4	42	11
Spring	62	0	52.5	1.7	42	4

Test of Change in Tier 1 Placement (Paired t-tests)

	Kindergarten		Grade 1		Grade 2	
Comparison	t	prob	t	prob.	t	prob.
Fall to Winter	4.95	<.0001	3.85	0.0001	2.44	<.0001
Fall to Spring	6.87	<.0001	5.08	<.0001	1.99	0.03
Winter to Spring	2.80	0.007	2.66	0.005	0.00	1.00

Note: Analysis was limited to students with data at all three time points: 68 students for K, 59 for grade 1 and 55 for grade 2.

Table A-4

Average RMSE Lesson Reached at Mastery, Winter 2015, by Training of Teacher, K and Grade 2, Ash Elementary

Kindergarten		
	Trained	Not Trained
Mean	98	73
SD	25	22
N	38	33
t-ratio	4.48	
Probability	<.001	
Cohen's d	0.93	

Grade 2		
	Trained	Not Trained
Mean	347.21	253.16
SD	29.57	77.49
N	58	19
t-ratio	7.75	
Probability	<.0001	
Cohen's d	1.54	

Table A-5

Regressions of Spring I-Ready Scale Scores on Fall Scores, Teacher Skills, and Lesson Progress, K and Grade 1, Ash Elementary

Kindergarten (N=68)									
	Model 1			Model 2			Model 3		
	b	t	beta	b	t	beta	b	t	beta
I-Ready Fall	0.66	5.42***	0.56	0.7	5.54***	0.57	0.36	2.60**	0.29
Good Teacher	----	----	----	12.67	1.87[a]	0.19	3.43	0.53	0.05
Cum. Spring *RMSE* Lesson	----	----	----	----	----	----	0.31	4.13***	0.48
Constant	177.61	4.31***	----	158.9	3.72***	----	231.58	5.48***	----
Adjusted R^2	0.31***			.32***			.49***		
Grade One (N=59)									
	Model 1			Model 2			Model 3		direct
	b	t	beta	b	t	beta	b	t	beta
I-Ready Fall	0.62	6.47***	0.65	0.55	5.65***	0.58	0.25	2.73**	0.27
Good Teacher	----	----	----	19.8	2.33*	0.24	-2.98	-0.38	-0.04
Cum. Spring *RMSE* Lesson	----	----	----	----	----	----	0.36	5.68***	0.65
Constant	206.6	5.39***	----	223.8	5.95***	----	256.8	8.37***	----
Adjusted R^2	.41***			.46***			.65***		

Note: *RMSE* utilizes a mastery learning approach in which each lesson builds on the previous lessons. Thus, students progress through the program can be calculated as a continuous measure starting with 1 at the start of the kindergarten program, proceeding through 160 at the end of K, beginning with 161 at the start of first grade, etc.

*, p< .05; **, p < .01; ***, p < .001; a, p < .10

Table A-6

Correlations of Winter 2015 Placement in RMSE, Fall 2014 Placement in RMSE, Having RMSE in Grade 1, and Training of Grade 2 Teacher, Grade 2 Only, Ash Elementary

	Winter Placement	Fall Placement	*RM* in Grade 1	Untrained Teacher
Winter Placement	1			
Fall Placement	.50***	1		
RM in Grade 1	-0.01	0.21	1	
Untrained Teacher	-.76***	-0.17	0.25	1

Note: All measures are Pearson Product Moment Correlations, *** = p<.001

Table A-7

Regression of Winter RMSE Placement on Fall RMSE Placement, Having RMSE in Grade 1, and Training of Grade 2 Teacher, Ash Elementary

	Standardized Coefficient (Direct Effect)	Correlation (Total Effect)	Indirect Effect
Fall Placement	0.22[a]	0.5	0.28
RMSE in Grade 1	0.03	-0.01	-0.04
Untrained Teacher	-0.66***	-0.76	-0.11

Note: a, p < .10; * p < .05, *** p < .001.

Table A-8

DIBELS Nonsense Word Fluency and Oral Word Fluency Lexile Scores, by Cohort and Testing Period, Birch School District

Comparing Cohorts with Partial or Full Implementation in Kindergarten

		Means		Anova results		
	Full	Partial	No DI in K	F	df	prob.
NWF K - Spring	37.0	28.3	26.6	14.99	2, 458	<.0001
NWF Grade 1 - Fall	35.0	32.0	27.3	4.24	2, 455	0.015
NWF Grade 1 - Winter	68.3	56.4	48.2	15.89	2, 449	<.0001
NWF Grade 1 - Spring	89.9	69.9	64.8	20.95	2, 453	<.0001
NWF Grade 2 - Fall	77.5	64.6	60.4	9.93	2, 452	<.0001

Pair-Wise Comparisons (Figure 4 Data)

	Scheffe Probabilities			Effect Sizes		
	Full v. Part	Full v. None	Part v. None	Full v. Part	Full v. None	Part v. None
NWF K - Spring	<.001	<.001	0.66	0.53	0.63	0.10
NWF Grade 1 - Fall	0.45	0.02	0.12	0.15	0.38	0.23
NWF Grade 1 - Winter	0.001	<.001	0.03	0.43	0.73	0.30
NWF Grade 1 - Spring	<.001	<.001	0.38	0.61	0.77	0.15
NWF Grade 2 - Fall	0.001	<.001	0.47	0.42	0.56	0.14

Comparing Cohorts by Exposure to DI in Kindergarten and First Grade - Supplemental Data

	Cohort				Anova results	
	1	2	3	4	F	prob.
ORF Lexile Grade 1 - Winter	-15.5	-15.4	-97.7	-39.3	2.02	0.11
ORF Lexile Grade 1 - Spring	137.4	132.1	57.9	105.9	1.55	0.20
ORF Lexile Grade 2 - Fall	261.4	196.4	184.0	183.8	2.98	0.03

Note: For the bottom panel, cohorts are defined as follows: cohort 1) those who had DI beginning in grade K and whose teachers had practice time in both grades; cohort 2) those that had DI in grade K and grade 1, but whose teachers only had practice time in grade 1; cohort 3) those that had DI in both grade K and grade 1, both whose teachers did not have practice time in either grade; and cohort 4) those that did not have DI in K and had DI in grade 1, but whose teachers were not given practice time. Degrees of freedom for those comparisons ranged from 3,445 to 3, 452.

Table A-9

Average NWEA Measures of Academic Progress and Effect Sizes, Fall, Spring, and Growth, Chestnut Students and National Norm Group, by Scheduling Status and Teacher Fidelity

	Full Schedule, No Fidelity data	Full Schedule, Lower Fidelity	Full Schedule, Higher Fidelity	Full Schedule, Total Group	Intermittent Schedule
MAP Fall 14	185	189	184	187	166
MAP Spring 15	193	201	196	198	171
Change MAP Fall to Spring	7.6	11.8	12	10.9	4.5
MAP Fall, Norm Group	207	188	192	193	168
MAP Spring, Norm Group	215	199	202	204	182
Change Fall to Spring, Norm Group	8.6	11.5	10.8	10.7	13.9
Average Effect Size, Fall	-1.54	0.08	-0.57	-0.43	-0.18
Average Effect Size, Spring	-1.55	0.09	-0.47	-0.4	-0.9
Average Effect Size, Growth	-0.16	0.07	0.17	0.04	-1.27
N	32	71	20	83	40

Note: Higher fidelity was defined as having an average rating of 90 percent or higher across all observations; lower fidelity as less than 90 percent. None of the teachers in the less well implemented schools had high fidelity. The averages for the norm group were calculated using data for students of the same grade.

Table A-10

AIMSWeb ORF Scores, First Grade, Elm Students, Fall, Winter, and Spring, by DI Group and At-Risk Status

	Fall	Winter	Spring
No DI, Higher Risk			
Mean	6.4	37.4	41.5
SD	2.8	30.7	23.6
No DI, Lower Risk			
Mean	37.0	64.6	91.8
SD	23.5	34.8	34.2
No DI, Total Group			
Mean	27.7	56.4	76.6
SD	24.2	35.8	39.0
DI, Higher Risk			
Mean	5.4	30.5	56.7
SD	3.3	17.2	26.9
DI, Lower Risk			
Mean	32.7	73.3	101.3
SD	26.6	28.2	31.9
DI, Total Group			
Mean	23.9	59.5	86.9
SD	25.4	32.2	36.8

Note: Oral Reading Fluency is based on the words read correct per minute. For the non-DI group there were 46 students in the higher risk group (lower one-third of the distribution) and 106 in the lower risk group. For the DI group there 43 students in the higher risk group and 90 students in the lower risk group.

Table A-11

Means and Standard Deviations, DIBELS Oral Reading Fluency Measures, by School, Grade, and Special Education Status, Maple Schools

Special Education Students							
	Control School			DI School			
	Mean	S.D.	N	Mean	S.D.	N	Cohen's D
1st grade, first of year	12.5	11.8	8	11.1	9.0	7	-0.13
1st grade, mid year	21.1	23.1	16	35.7	26.6	16	0.59
1st grade, end of year	36.5	25.7	16	67.1	39.2	16	0.94
2nd grade, first of year	36.6	24.9	16	60.9	32.1	16	0.85
2nd grade, mid year	57.6	31.0	16	90.6	47.8	16	0.84
2nd grade, end of year	78.6	29.2	16	107.4	42.3	17	0.81
3rd grade, first of year	65.7	29.1	15	92.9	41.8	17	0.77
3rd grade, mid year	83.7	31.5	15	112.9	42.1	17	0.79
3rd grade, end of year	100.9	36.7	15	130.4	44.5	17	0.73
General Education Students							
	Control School			DI School			
	Mean	S.D.	N	Mean	S.D.	N	Cohen's D
1st grade, first of year	24.4	30.1	52	30.1	31.9	49	0.19
1st grade, mid-year	37.4	31.7	68	52.2	34.4	67	0.45
1st grade, end of year	65.8	37.4	68	82.7	37.4	67	0.45
2nd grade, first of year	62.7	34.1	69	75.9	35.4	67	0.38
2nd grade, mid-year	94.0	35.3	68	101.3	33.9	67	0.21
2nd grade, end of year	111.7	30.7	68	121.3	32.5	66	0.30

3rd grade, first of year	99.0	31.6	66	110.5	36.6	62	0.34
3rd grade, mid-year	117.4	31.9	65	130.6	37.2	62	0.38
3rd grade, end of year	135.1	30.7	63	147.3	35.5	61	0.37

Table A-12

Oral Reading Fluency (ORF) Scores, Grade 1, by Testing Period, School, At-Risk Status, and RMSE Status. Elm and Maple Schools

Elm School (No Double Dosing)

	At-Risk		Not-At Risk	
Testing Period	Control	*RMSE*	Control	*RMSE*
Fall	6	5	37	33
Winter	37	30	65	73
Sprng	41	57	92	101

Maple Schools (Double Dosing for RMSE Students)

	At-Risk		Not-At Risk	
	Control	*RMSE*	Control	*RMSE*
Fall	13	11	24	30
Winter	21	36	37	52
Sprng	37	67	66	83

Note: Elm School used AIMSWeb measures of ORF and Maple Schools used DIBELS.

Table A-13

Percentile of Average First Grade Student, Reading CTBS, Inner City Schools, by Year of Implementation and Group

Vocabulary		
Year of Implementation	NIFDI	ODI
0 - No DI	17	24
1st year	25	28
2nd year	34	32
3rd year	44	37
4th year	54	42
5th year	65	48
6th year	74	53
7th year	82	58
Comprehension		
Year of Implementation	NIFDI	ODI
0 - No DI	17	24
1st year	25	28
2nd year	34	32
3rd year	44	37
4th year	54	42
5th year	65	48
6th year	74	53
7th year	82	58

Note: All scores were adjusted for poverty level of school using the results of mixed model regressions. Calculations used NCE scores and these were converted to percentiles. Values for the ODI schools for year 0 were predicted from the results of mixed model regressions (See Stockard, 2011 for details).

Table A-14

Percentage of Students At or Above Grade Level Proficiency in Reading on State Assessment, Grades 3 to 5, Rural Southern School

Implementation Status	Grade 3	Grade 4	Grade 5	Mean
None	28.6	20	23.4	24
Partial	44.7	23.9	38.9	35.8
Yr. 1	32.8	44.4	34.6	37.2
Yr. 2	38.2	31	66.7	45.3
Yr. 3	50	39.6	61.5	50.3
Yr. 4	70.6	58.4	66.1	65
Yr. 5	66.1	76.6	69.8	70.8

School-wide Direct Instruction was started in mid-year 1997-98. 1998-99 was the first full year of implementation. Data taken from Vitale and Joseph, 2008, Table 2, page 8.

Table A-15

Number of Trainer Reported Problems, Pre- and Post- Intervention and Average Level of Training Intervention by Type of Intervention and Years Teaching DI, Southwestern District

	Number of Years of Teaching DI			
	Yr. 1	Yr. 2	Yr. 3	Yr. 4
Classroom management				
Pre-intervention problems	2.0	1.8	1.2	1.3
Post-intervention problems	1.4	1.3	1.2	1.3
Level of intervention	2.4	2.2	1.6	1.8
Programmatic teaching				
Pre-intervention problems	1.9	1.8	1.4	1.3
Post-intervention problems	1.2	1.2	1.2	1.1
Level of intervention	2.8	2.6	1.8	1.8

Table A-16

Student Performance Predicted from Fall and Spring Teacher Corrections

Performance Measure	Predictor	F	Probability
Kindergarten			
Word Reading	Fall corrections	6.13	0.01
	Spring corrections	5.68	0.02
	Interaction	6.01	0.01
Grade 1			
Word Reading	Fall corrections	15.70	0.0001
	Spring corrections	20.60	0.0001
	Interaction	28.04	0.0001
Comprehension	Fall corrections	28.44	0.0001
	Spring corrections	20.03	0.0001
	Interaction	35.28	0.0001
Grade 2			
Word Reading	Fall corrections	14.20	0.0001
	Spring corrections	19.80	0.0001
	Interaction	16.76	0.0001
Comprehension	Fall corrections	28.89	0.0001
	Spring corrections	19.68	0.0001
	Interaction	27.54	0.0001

Note: Taken from Carlson and Francis, 2002, Table 12, p. 163; Kindergarten analysis, df=3, 1,1459; First grade analyses, df = 3, 1, 646; Second grade analyses, df = 3, 1, 877.

Table A-17

Average Nonsense Word Fluency Scores, Winter and Spring, Kindergarten by Teachers' Years of Experience Teaching DI, Midwestern District

	Winter	Spring
Other program (Not DI)	21.4	43.5
1 year of experience	24.8	40.0
2 years of experience	33.8	65.9
3 years of experience	35.3	55.2
4 years of experience	42.0	75.4

Note: N=115 students for the year with the non-DI program, 97 for the year when teachers had 1 year of experience, 94 for 2 years, 61 for three years and 22 for 4 years.

Table A-18

End of Year Stanford Achievement Test Means and Percentile Rank of Average Student, By Year Started Program, Grade and Exposure to DI

First Grade

	DI Schools		Comparison Schools			
	Started K	Started 1st	Started K	Started 1st		
Word Reading						
Mean	541.51	516.98	509.83	506.79		
SD	55.48	52.87	51.79	51.79		
% Below 25th percentile	8	20	26	23		
% Above 50th percentile	73	52	48	47		
Reading Comprehension						
Mean	549.86	533.12	518.31	518.31		
SD	49.37	47.97	47.71	45.18		
% Below 25th percentile	11	19	26	24		
% Above 50th percentile	78	60	55	55		

Second Grade

	DI Schools			Comparison Schools		
	Started K	Started 1st	Started 2nd	Started K	Started 1st	Started 2nd
Word Reading						
Mean	580.3	562.27	553.98	555.21	553.98	551.01
SD	43.65	46.05	41.44	42.76	44.53	41.67
% Below 25th percentile	16	27	33	33	31	32
% Above 50th percentile	61	44	36	38	36	37
Reading Comprehension						
Mean	589.75	578.32	574.64	571.01	569.32	569.57
SD	30.86	34.99	32.91	31.56	35.23	33.59
% Below 25th percentile	12	24	29	34	34	32
% Above 50th percentile	66	51	43	39	38	36

Note: Data taken from Table 5, p. 154, Carlson & Francis, 2002. All analyses used multi-level modeling techniques, controlling for the classroom, or teacher, effects. Comparison schools were matched on rate of the receipt of free or reduced price lunch, ethnic minority, limited English proficient, and meeting the minimum state-mandaded reading performance requirement. n = 6,984 for first grade and 6,142 for second grade.

Table A-19

Average Scale Score and Percentage Passing Third Grade State Assessment by Year Started DI and Comparison Schools, Southwestern District

DI Schools			
	Started K	Started 1st	Started 2nd
Mean	79.86	78.05	75.04
SD	16.44	18.26	21.22
% Passing	82	79	73
Comparison Schools			
	Started K	Started 1st	Started 2nd
Mean	72.47	71.89	71.08
SD	22.66	24.02	23.46
% Passing	68	66	65

Note: Data taken from Table 7, page 157, Carlson & Francis, 2002

Table A-20

ORF Lexiles, Midwestern District, By When Started DI and Testing Period

	Started DI in First Grade			Started DI in K			
Testing Period	Mean	SD	N	Mean	SD	N	Effect Size
Gr 1 - Winter	-113.6	281.7	104	36.1	283.9	214	0.53
Gr 1 - Spring	18.8	289.5	104	182.8	274.8	214	0.58
Gr 2 - Fall	184.0	229.6	104	219.6	223.7	214	0.16
Gr 2 - Winter	403.3	252.6	104	506.0	268.4	214	0.39
Gr 2 - Spring	520.5	258.3	104	594.0	275.2	214	0.28
Gr 3 - Fall	378.0	243.9	97	440.6	254.1	199	0.25
Gr 3 - Winter	569.6	263.3	96	638.4	268.6	88	0.26
Gr 3 - Spring	695.6	268.9	94	722.1	268.7	86	0.10

Note: Data also reported in Stockard and Engelmann, 2010

Table A-21

Estimated Differences in Words Read of Students Starting DI in K and Those Starting Grade 1, By Grade, Yearly and Cumulative, Midwestern District

Mid-Year ORF and Estimated Daily Reading Volume

	Mid-Year ORF		Est. Daily Reading Volume	
	Gr. 1 DI start	K DI Start	Gr. 1 DI start	K DI Start
First Grade	34	51	2,040	3,060
Second Grade	80	94	4,800	5,640
Third Grade	101	111	6,060	6,660

Estimated Yearly Reading Volume and Differences by Group

	Gr. 1 DI start		K DI Start	
	Yearly	Cumulative	Yearly	Cumulative
First Grade	367,200	367,200	550,800	550800
Second Grade	864,000	1,231,200	1,015,200	1,414,800
Third Grade	1,090,800	2,322,000	1,198,800	2,505,600

Note: Some data taken from Stockard & Engelmann, 2010.

References

Carlson, C. D., & Francis, D. J. (2002). Increasing the reading achievement of at-risk children through Direct Instruction: Evaluation of the Rodeo Institute for Teacher Excellence (RITE). *Journal of Education for Students Placed at Risk*, *7*(2), 141-166.

Coughlin, C. (2014). Outcomes of Engelmann's Direct Instruction: Research Syntheses. In J. Stockard (Ed.), *The Science and Success of Engelmann's Direct Instruction* (pp. 25-54). Eugene, OR: NIFDI Press.

Engelmann, K. (2014). Creating effective schools with Direct Instruction. In J. Stockard (Ed.), *The Science and Success of Engelmann's Direct Instruction* (pp. 99-122). Eugene, OR: NIFDI Press.

Engelmann, S. E., & Engelmann, K. E. (2004). Impediments to scaling up effective comprehensive school reform models. In T. K. Glennan, Jr., S. J. Bodilly, J. R. Galegher, & K. A. Kerr (Eds.), *Expanding the reach of education reforms: Perspectives from leaders in the scale-up of educational interventions* (pp. 107-133). Santa Monica, CA: Rand.

MetaMetrics. (2009). *Linking DIBELS oral reading fluency with the Lexile framework for reading.* Durham, NC: MetaMetrics, Inc.

Northwest Evaluation Association. (2011). *RIT scale norms: For use with measures of academic progress (MAP) and MAP for primary grades*. Portland, OR: Author

Stockard, J., & Engelmann, K. (2010). The development of early academic success: The impact of Direct Instruction's *Reading Mastery*. *Journal of Behavioral Assessment and Intervention for Children*, *1*(1), 2-24.

Stockard, J. (2011a). Direct Instruction and first grade reading achievement: The role of technical support and time of implementation. *Journal of Direct Instruction*, *11*(1), 31-50.

Stockard, J. (2011b). Increasing reading skills in rural areas: An analysis of three school districts. *Journal of Research in Rural Education*, *26*(8), 1-19.

Stockard, J. (2014). *The relationship between lesson progress in Direct Instruction programs and student test performance. NIFDI Technical Report 2014-1*. Eugene, OR: National Institute for Direct Instruction.

Vitale, M. R., & Joseph, B. L. (2008). Broadening the institutional value of Direct Instruction implemented in a low-SES elementary school: Implications for scale-up and school reform. *Journal of Direct Instruction*, *8*(1), 1-18.