Idaho Reading First 2006–2007 Evaluation

Conducted by—

BOISE STATE UNIVERSITY **CSi** PS

COLLEGE OF EDUCATION
CENTER FOR SCHOOL IMPROVEMENT & POLICY STUDIES

1910 University Dr.
Boise, ID 83725-1745

Phone: 208-426-1077
Fax: 208-426-3564
Roger A. Stewart

1

IDAHO READING FIRST EVALUATION
2006-2007 FINAL REPORT

TABLE OF CONTENTS

Executive Summary

EXECUTIVE SUMMARY

Idaho Reading First (IRF) is a federally-funded program focused on helping kindergarten through 3rd grade students at risk become successful early readers. The Idaho Reading First initiative began in the 2003-2004 school year with 21 schools in 11 districts. A second group of schools was awarded grants the second year of the initiative, 2004-2005, bringing the total number of schools to 31 representing 16 Idaho districts. One school withdrew from the initiative at the conclusion of the 2004-2005 school year, leaving 30 schools representing 16 districts. IRF schools are located in all geographic regions of the state and represent a diversity of size and demographic profiles. All are Title I schools, meaning at least 40% of their students qualify for free or reduced price meals.

This report presents the results of an evaluation conducted during the 2006-2007 academic year by Dr. Roger A. Stewart, a professor in the College of Education at Boise State University. This is the third year Dr. Stewart has conducted the evaluation so some findings are discussed in the context of this and previous evaluations. The evaluation design included all 30 IRF schools.

Data Sources for the 2006-2007 Evaluation

A variety of data were collected from IRF schools, including the following:

- Results from the Idaho Reading Indicator (IRI), a state-mandated early literacy screening instrument. Four years of data were longitudinally analyzed;

- Results from standardized instruments administered only in IRF schools (i.e., the Texas Primary Reading Inventory (TPRI) and the Iowa Test of Basic Skills (ITBS)). Four years of data were longitudinally analyzed;

- Extensive classroom observations of 29 high-performing IRF teachers followed by in-depth interviews with the teachers, their reading coaches and their principals. The 29 teachers were in 15 IRF schools and eight districts. The teachers were identified based on ITBS scores for their classrooms; and

- Data for teacher turnover, principal turnover, student mobility and special education referral rates over the past six years.

Findings from the 2006-2007 Evaluation

All data was collected during the spring of 2007 and analyzed and interpreted during the spring and summer of 2007. Results show positive benefits from Idaho Reading First, including:

- Average 3rd grade ITBS grade equivalent scores in IRF schools are higher than the state average from the last year of statewide ITBS testing (i.e., 2001). The average for all Idaho 3rd graders in 2001 was the 54th percentile, which equates to a grade equivalent score of 3.7. IRF 3rd graders averaged a 4.1 grade equivalent

score in 2004, the first year of IRF test data. This is substantially above the last available ITBS statewide average. This high relative performance by 3rd graders has been sustained for the duration of the IRF initiative with average grade equivalent scores holding relatively steady at 4.1 in 2005, 4.0 in 2006, and 4.0 in 2007;

- On average 94% of kindergarteners passed a screener on the Texas Primary Reading Inventory (TPRI) during the spring 2007 administration of the assessment;

- Hispanic kindergarteners are doing quite well. A three-year trend in scores reveals that on average about 90% pass a TPRI screener each spring. This percentage is quite close to the whole group (94%) and to White students (95%);

- In twelve IRF schools 100% of their Hispanic kindergarten students passed a screener in the spring of 2007. In 2005 ten schools achieved this and in 2006 13 schools did;

- On one or more of the three assessments given in IRF schools (i.e., ITBS, TPRI, and IRI) some grade levels within some schools have achieved sustained growth during the years they have participated in the initiative;

- Observation and interviews with the 29 high-performing teachers revealed common characteristics contributing to their success, including fidelity to the core program, extensive knowledge of the core program and its strengths and weaknesses, strong work ethic, high academic press in their classrooms, enhanced vocabulary instruction, high-quality workshop and intervention periods, high expectations for all students, excellent classroom management, effective use of data, and awareness of individual student's strengths and weaknesses;

- An important criteria used to identify high-performing teachers who were intensively studied this year was strong relative success with Hispanic students. Common characteristics of these teachers that were important to their success with Hispanic students include: (a) high expectations for all students and a work ethic to diligently pursue this each day, (b) an enhanced focus on vocabulary throughout the lesson cycle each day, (c) high-quality supplemental programming such as ELL/ESL programs, and (d) effective workshop/intervention periods where individual needs of students are consistently addressed; and

- Emerging yet tentative evidence reveals that special education referrals may be trending downward in some grade levels and IRF cohorts. The quality of this data remains suspect, however, due to inconsistent collection of data across IRF schools.

Areas needing additional attention/improvement:

- IRF schools continue to manifest a substantial amount of variability in test score performance both within and between schools. In other words, some grade levels

within schools are doing substantially better than other grade levels and some IRF schools are doing better than others;

- When all grade levels within IRF schools are examined in aggregate and when all IRF schools are examined in aggregate, individual schools and the whole group of schools are not consistently improving their test scores. Some grade levels within buildings have shown steady improvement but no school has achieved consistent growth at all grade levels on multiple assessments. Test score averages for the entire IRF network have remained relatively constant since the beginning of the initiative;

- The combined results of the three evaluations point to some salient variables important to student success. Most importantly, the quality of the teaching work force is of paramount importance. Continued staff development informed by the results of these evaluations and insights from other states that have addressed similar problems in their IRF networks are extremely important if gains are to be realized by the conclusion of the program in Idaho;

- With the exception of TPRI screener scores for kindergarteners, Hispanic student performance continues to lag that of White students and achievement gaps are not closing but instead remain relatively steady;

- Some high-performing teachers are experiencing work-related stress caused primarily by the rigidity with which they have to implement their core reading programs;

- Student mobility remains high in some IRF buildings. Eleven IRF schools experienced greater than 40% mobility in 2006-2007. Eleven others experienced 30-40% mobility. This is a difficult factor to control and one that IRF schools will have to continue to address through effective parent and community outreach and intervention programming; and

- Teacher turnover remains a significant but perhaps controllable factor in IRF buildings. Principal turnover is not as widespread as teacher turnover.

Recommendations for Future Program and Evaluation Activities

This evaluation accumulated a considerable amount of valuable data. Future evaluations should build on this database by examining other aspects of the IRF initiative to gain additional insight into key strengths and weaknesses. Specific recommendations include:

- A Web-based clearinghouse for best practices needs to be developed. The highest-performing teachers across the network have extensive knowledge about how to leverage the strengths and remediate the weaknesses of their core programs and the IRF literacy framework. This clearinghouse could provide ideas for lessons, teacher insights focused on curriculum and instruction, specific interventions that worked for particular types of children, ways to use data, etc.

Specific areas that need to be showcased in the clearinghouse are vocabulary and fluency;

- Teachers should be included in calibration visits. The high-performing teachers expressed a strong desire to observe other teachers and share best practices and insights. The amount of institutional knowledge and expertise that has accumulated as a consequence of IRF is quite striking. Mechanisms need to be put in place whereby this body of knowledge can be shared. The clearinghouse mentioned above is one way, but face-to-face meetings and classroom visits between teachers were also emphasized by high-performing teachers. There are outstanding teachers all across the IRF network. Developing a cadre of these teachers to travel from school-to-school to do demonstration lessons and to discuss curricular and instructional issues with their IRF colleagues could serve as highly effective staff development;

- If cross-school sharing and collaboration are not possible, then facilitating sharing across grade levels in individual buildings could be an alternate possibility. As the test score analyses showed, some grade levels within individual IRF buildings have done quite well. The best practices within these grade levels should be systematically shared with the other grade levels in the building. Where this has already occurred and the best practices have not been incorporated at the other grade levels or the best practices have not resulted in higher student achievement, leadership needs to step in to ascertain the root cause of these outcomes;

- Perhaps the very best IRF teachers should be given permission to experiment, while maintaining fidelity to the core program, to see how they can impact student achievement. Their experimentation might result in increased test scores. If this occurs, then these best practices should be disseminated;

- IRF leaders should consider an expanded evaluation design for the remaining years of this initiative that includes both summative and formative components. This design might include continued, extensive classroom observations with individual teacher interviews, but it could also include focus groups of teachers, parents and students. Additionally, looking beyond the IRF network to successful schools outside of the network and perhaps even outside of Idaho might provide insights as to how to proceed and accelerate school growth; and

- A possible bright spot of significant importance is the impact on special education in IRF schools. There are hints in the current qualitative and quantitative data that referral rates are down and perhaps so too is the number of students who are placed in special education. Special educators in IRF schools and special education directors in IRF districts need to be systematically contacted to ascertain the most appropriate data to be collected that is widely available in IRF districts that will shed light on this outcome. Once this data is identified, then it needs to be collected from all IRF schools, and it should include at least three years of baseline data. Cost/benefit analyses should be conducted to show the impact of IRF, and special educators should be surveyed and interviewed for their insights about the impact IRF programming has had on special education.

This initiative, in its third/fourth year of implementation, continues to address the highly-challenging task of improvement for high-poverty learners, an area that requires time and effective intervention. Overall, Idaho Reading First has demonstrated success in a number of areas and is having a positive effect on student performance in many schools. Data, however, also indicate areas of continued concern, primarily the variability in test score performance within and between IRF schools. This report explores the strengths and continuing challenges of Idaho Reading First, provides initial explanations for the variance between IRF schools on test performance and provides recommendations for future action.

Introduction

INTRODUCTION

Idaho Reading First (IRF) is a federally-funded program focused on helping kindergarten through third grade students at greatest risk become successful early readers. It is a competitive grant process whereby schools meeting guidelines set by the State submit a proposal that is then peer reviewed and scored. Only those applicants who meet the guidelines and also submit high-quality proposals clearly delineating how funds will be used and how outcomes will be measured are awarded funding.

The Idaho Reading First initiative began in the 2003-2004 school year with 21 schools in 11 districts. A second group of schools was awarded grants the second year of the program, 2004-2005, bringing the total number of schools to 31 in 16 districts. During the third year of the grant, 2005-2006, one school withdrew from the initiative, leaving 30 schools in 16 districts. All geographic regions of the state and virtually all sizes and types of schools are represented. Reading First funds allow each participating school to:

- Purchase scientifically-based reading curricula and intervention programs;

- Hire a full-time reading coach who works in the Reading First school to provide technical support and assistance to the K-3 teachers as they implement the scientifically-based curricula and other interventions;

- Provide professional development to teachers, reading coaches and principals concerning scientifically-based early literacy curriculum, teaching and assessment;

- Participate in calibration visits where IRF school administrators, reading coaches and IRF personnel from the Idaho Department of Education travel in teams to participating IRF schools to observe classroom instruction and talk with teachers, the coach and the administrator of the building. The calibration visits serve as a mechanism for sharing ideas and information and exchanging best practices across the IRF initiative; and

- Receive technical assistance from the Idaho Reading First staff housed within the Idaho Department of Education.

Evaluation Designs: Past and Current

The designs of the annual evaluations of the Idaho Reading First initiative have changed each year in response to stakeholder needs. The first evaluation conducted during the 2004-2005 school year employed a comparison group design and analyzed one year of assessment data. Classroom observations and teacher, coach and principal surveys were conducted in both IRF and comparison schools. The evaluation's primary focus was to compare curriculum and instruction in IRF schools to that in comparison schools and to compare IRF school test scores to comparison school test scores. Although valuable information was derived from the comparison group design, considerable evaluation resources were applied to the comparison schools and only limited test score comparisons were possible since comparison schools only administered state-mandated tests (i.e., ISAT and IRI). When the substantial resources applied to comparison schools was reflected against the important finding from the 2004-2005 evaluation that there was

substantial variation in test score performance across IRF schools, it was decided to refocus efforts and resources on more intense study of IRF schools to try to explain this variation. If the variation could be understood, then remedial steps could be implemented to bring all schools to the point where test scores were consistently improving. The evaluation design was thus modified for 2005-2006.

The 2005-2006 evaluation focused more resources on looking at IRF schools in greater detail with the goal of explaining why some were making progress while others were not. Consequently, during 2005-2006, a non-comparative design was employed so that additional resources could be applied to more intense and sustained observations in IRF schools, more sophisticated longitudinal analyses of test score data to better identify high-performing and low-performing schools, and more sophisticated surveying of IRF teachers and coaches. Results from the evaluation provided important insights into the differential functioning of IRF schools. Specifically, the observations and survey information revealed important qualities of high-performing schools and to a lesser degree qualities of high-performing teachers within the schools. In other words, based on the data it became possible to predict which schools would be high performing and which would not, and to a lesser degree, which teachers would be high performing and which would not.

High-performing schools had excellent daily schedules that provided adequate time for workshop, intervention and teacher collaboration. These schedules were adhered to by teachers, and workshop/intervention time and collaboration time were wisely and effectively used to focus on students and their individual needs. These schools also had strong leadership and often excellent reading coaches. High-performing teachers within these schools had excellent classroom management, high levels of academic press in their classrooms and an infectious enthusiasm for their teaching and their students' learning. Again, much variation was found across both schools and teachers, but one thing surfaced of greatest importance: A school could have an excellent schedule and strong leadership, and the teachers could even teach the core reading program with fidelity, but if the teachers didn't teach with enthusiasm and a commitment to the success of all students, then results were less than optimal. In short, teachers were key to the success of IRF.

Based on these findings, it was deemed important that in the 2006-2007 evaluation the qualities of high-performing IRF teachers be explored in greater depth. To that end, the 2006-2007 evaluation focused on in-depth classroom observations and interviews with 29 high-performing teachers, their coaches and their principals, with the goal to create an Idaho Reading First best practices synthesis that could be used throughout the network of schools. Classroom observations have been a mainstay of all of the IRF evaluations, although they too have evolved as the evaluations have changed. This year classroom observations were conducted in 29 teachers' classrooms in 15 Reading First schools. These observations included the 90-minute literacy block and the workshop and intervention periods. Teachers were selected based on the ITBS test score performance of both their overall classroom populations and their Hispanic subpopulations. In other words, teachers were identified who had brought their overall class to a relatively high-level ITBS performance while also bringing their Hispanic students to high relative scores. Classrooms were identified where there were usually at least eight Hispanic

students enrolled, with preferably 10 or more. Hispanic student scores were focused upon because a key finding of the 2005-2006 IRF evaluation was that the proportion of Hispanic students in a school significantly and consistently predicted test scores with an inverse relationship, meaning the higher the percentage of Hispanics in a building, the lower the predicted test scores. Observed teachers spanned the continuum of IRF schools from rural to urban and from small to quite large. These observations were conducted by experienced literacy educators who had attended a one-day training session covering observation techniques, what to observe and how to set up and conduct observations. The observations were intensive and lengthy. Classrooms were observed for two hours or more. Observers took extensive notes while observing and also completed an observation guide that directed them to look for specific teaching behaviors. Many of the notes were verbatim transcriptions of what teachers and students said during the instruction. Additionally, the notes contained detailed descriptions of classroom instruction and materials along with a record of how long the teacher and students spent on various aspects of the lesson.

In the past, classroom observers have always had informal conversations with the teachers they observed, the reading coach in the building and the building principal, but during this year's evaluation more intense formal interviews also were conducted with those individuals. The purpose of the interviews was to collect additional information about what makes these teachers so successful with their students. Individual interviews were conducted with all 29 of the teachers who were observed. The person who observed the teacher also conducted the interview with that teacher, her coach and her principal. All interviews were conducted after the observation had been completed. This provided the interviewer with more context within which to discuss the teacher's practice. These interviews lasted at least an hour and were conducted after school hours so the teachers could participate with minimal interruption. These 29 teachers were located in 15 different IRF schools. Therefore 15 coaches and 15 principals were also interviewed. These interviews lasted 30-60 minutes and most were conducted during the school day since coaches and principals usually have more flexible school-day schedules.

All interviews were semi-structured and followed the same interview protocol of questions, meaning that a list of questions was to be closely followed but the interviewer could diverge from the list when opportunity arose or if circumstances called for it. Interviews were not electronically recorded (See Appendix A for the interview protocols). Instead interviewers took notes and then immediately after completing the interview wrote down additional comments and insights before they were lost from memory.

The mass of data collected was substantial. The notes from the observations and the interviews totaled over 300 pages. A qualitative data analysis process was used to reduce and synthesize this data into the findings and recommendations that will be reported in this document.

In addition to the in-depth exploration of high-performing teachers, the 2006-2007 evaluation had two other important components. First, the longitudinal data analysis was extended another year. This look at school test scores over all of the years the schools have been in the initiative was first done during the 2005-2006 evaluation. It was continued in the 2006-2007 evaluation, since each additional year of data allows for

better trend analysis and more valid conclusions. During 2006-2007 the first cohort of IRF schools had four years of test score data available, thus making trends more readily identifiable. Iowa Test of Basic Skills (ITBS), Idaho Reading Indicator (IRI), and Texas Primary Reading Inventory (TPRI) scores were longitudinally analyzed.

Second, data on school background variables was collected. This data collection was started in 2005-2006 and continued in 2006-2007. The data collected included principal turnover data, teacher turnover data, special education referral data and student mobility data. These variables provide important information about school characteristics that can influence test score performance. If the teaching staff is in constant flux or if leadership is changing rapidly, the school will probably experience greater challenges coming together to focus on student literacy learning. High student mobility can also be a negative for test score performance. And special education referral rates can be a good indicator of the outcomes of high-quality curriculum and instruction and can thus be a measure of school success.

Teacher and principal turnover data was collected for two years prior to the start of IRF in the particular school and then for the years of participation. The two prior years were deemed important to establish a baseline of turnover for the school to see if turnover accelerated or decelerated with the advent of IRF. IRF coaches in each building provided this data. Student mobility data was collected from principals, school staff or central office personnel during 2005-2006, but for 2006-2007 the Idaho Department of Education provided this data. Mobility was defined as the percent of students who withdrew or enrolled after the beginning of school. Anecdotal information received from Idaho Department of Education personnel who work with IRF schools and anecdotal information received by the evaluators as they conducted observations in the schools during 2005-2006 supported the hypothesis that referral rates were dropping in IRF schools. To substantiate this anecdotal information more systematic data was collected.

In the remainder of this report, additional details about all facets of the evaluation will be provided along with the results from the various instruments and surveys.

Classroom Observations and Interviews: An Exploration of Best Practices

CLASSROOM OBSERVATIONS AND INTERVIEWS: AN EXPLORATION OF BEST PRACTICES

Following are some quotations from teachers, principals and coaches that capture the essence of what these 29 teachers were like and the best practices that they exemplify. The characteristics demonstrated in these quotations plus others will be more thoroughly developed and discussed below. These initial quotations are provided as an introduction and overview for the reader:

> *"Be very explicit with examples, review all strategies and connect everything to the real world. Keep expectations high and practice good classroom management."* (Teacher response to a question about what teachers should do to be successful.)

> *"She doesn't waste instructional time. She is pro-active in preparation and plans. She focuses on a goal and goes for it.*" (Principal response when asked about outstanding characteristics of a teacher.)

> *"She is a spokesperson for all students and really taps into ELL student needs. She advocates for students and does home visits when needed."* (Principal response when asked why the teacher had been successful with Hispanic students.)

> *"She keeps her data and uses it. She has strong classroom management skills. Her students are engaged with the instruction. She takes time to get them focused. When they do not understand something, she takes them back to where they do understand and brings them forward."* (Coach response when asked about outstanding characteristics of a teacher.)

> *"She knows Open Court. She's quick to catch on, grasps new materials like during in-service. She's responsive to new ideas, very reflective in the way she teaches. She's open-minded about trying new things. She started using graphic organizers before anyone else. She teaches deep comprehension. She leads teachers in blending board [techniques]—you know, the green band."* (Coach response when asked about outstanding characteristics of a teacher.)

These few quotes reveal many of the important characteristics of these 29 successful teachers: explicit instruction, full use of classroom time, connections to the real world, high expectations for all students, strong classroom management, dedication, hard work, willingness to spend extra time, thorough planning, attention to individual children, high student engagement, teacher reflection and teaching for higher order thinking. This preliminary list underscores an important point — that successful teaching in IRF classrooms is both an art and a science. As you will see below, these teachers have strong technical command of their core programs. This is the science of teaching. But they also teach with great passion and a moral commitment to help all children become readers and writers. This is the art of teaching. We now turn to a description of the teachers.

Background to the teachers: A total of 29 teachers were observed and interviewed. All were female. Twenty-eight of these teachers were identified by their ITBS scores; they had achieved high relative ITBS scores for both their overall classes and their Hispanic students during 2005-2006. The 29th teacher was a special services teacher who was praised highly by her colleagues during interviews as an integral part of their success. She happened by the classroom of one of the 28 teachers while an interview was being conducted so she was asked if she would be willing to be interviewed. She agreed. She had been observed while she was working in one of the identified teacher's classrooms.

The 28 identified teachers were located in 15 different IRF schools in eight districts. Six schools had one identified teacher, four schools had two teachers, four schools had three teachers, and one school had four teachers. This sums to 30 identified teachers but only 28 were observed and interviewed. The two missing teachers had left their IRF schools and were unavailable. There were seven teachers at each of the grade levels (i.e., K, 1, 2, and 3).

Years of teaching experience of the 29 teachers ranged from 3 to 37. The average was 17.6 years with a standard deviation of 11.9. Eight teachers had less than 10 years of experience. Teachers had been at the grade level in which they were identified as being high performing for an average of 9.7 years with a standard deviation of 6.2. Their years teaching at this grade level ranged from a low of two to a high of 22.

To summarize, except for two teachers with only three years of experience and one teacher with four, all others were quite experienced educators. Additionally, all but seven teachers had five or more years experience at the grade level in which their high performance had been identified. This was an experienced pool of teachers who had spent a number of years teaching at the grade level in which they were identified. All but one teacher had been teaching at their IRF schools for three or more years.

Because of their experience, these teachers have grown into the IRF program. As they have used the core program and have worked with their colleagues to develop the daily schedules and procedures, they have come to know all of it well, both strengths and weaknesses, and have therefore evolved to the point where they can devote energy and time to making the program better.

Reliability and validity of data and data analyses: Observers were directed to describe what they saw in classrooms in sufficient detail so that a reader of the notes taken during the observation could vicariously experience the instruction in the classroom and the style of the teacher. This resulted in much verbatim classroom dialogue and teacher talk being captured in the notes along with much description of what was occurring in the classroom, including teacher and student behaviors and curriculum materials being used.

The person who conducted the classroom observation of a teacher also conducted the interview with that teacher, and usually the coach and the principal. Interviewers were directed to take notes during the interviews instead of electronically recording. Notes are

not as intrusive or threatening during an interview. Interviewers were directed to capture interview responses in sufficient detail so again a reader could vicariously experience the interview. This, too, resulted in the capturing of numerous verbatim responses.

The team compiled more than 300 pages of notes from the observations and interviews. The lead evaluator read all of the notes and annotated them for themes that emerged about what was common among these high-performing teachers. One of the experienced observers who had also conducted observations for the 2005-2006 evaluation was retained during the summer of 2007 to review and critique the lead evaluator's conclusions about the data.

To further enhance the reliability and validity of the data analyses, instances of triangulation were noted as supporting evidence for assertions. Triangulation occurs in a number of different ways. One form of triangulation occurs when there is corroboration between research participants. An example of this is when something observed in a teacher's classroom or said by a teacher during an interview is also mentioned by a reading coach and/or principal during their interviews. For example, an observer might note that the teacher goes out of her way to emphasize the relevance to students of what is being learned. The teacher then mentions in an interview that she emphasizes relevance in her lessons to help students stay engaged and motivated to learn. The teacher relaying this information is corroboration for the observation and it represents the start of a triangulated assertion. The supporting evidence becomes all the more defensible when either the principal or the reading coach or both also mention that the teacher is good about making things relevant for the students. As these occurrences accrue across numerous teachers, coaches and principals, the data becomes deeply triangulated and the assertions derived accrue adequate support.

Triangulation was not difficult to achieve with this data set. What the teachers said about themselves was often exactly what the coach and principal said about them and vice versa. Furthermore, what observers witnessed in the classrooms was often articulated by the teachers, coaches and principals. For example, an observer witnessed a focus on vocabulary and higher order questioning in the classroom. Then, during the interview, the teacher was asked what qualities of her teaching are important to her success with her students and she said that she focuses on vocabulary and also higher order questioning. If the coach and/or the principal also mentioned a key characteristic of the teacher's success is a focus on vocabulary and higher order questioning, then additional triangulation occurred. These teachers knew what they did in their classroom and why they were doing it. They were very aware of their practice, and their instruction was very clear and concise so observers could readily pick out salient qualities. Additionally, the teachers were quite articulate in describing their teaching. Coaches and principals were also quite knowledgeable of what was occurring in the teachers' classrooms, so there were many instances of triangulation since teachers were self-aware and articulate and coaches and principals were astute observers.

Findings from the best practices exploration: The findings will be discussed in the following section. The characteristics outlined in the opening section will be expanded

and additional characteristics will be introduced and discussed. But before doing so, three terms need to be defined. The first is core program. This is the core reading program that is used in the school. IRF schools adopted either Open Court Reading (OCR) or Nation's Choice from Houghton Mifflin (HM). OCR will be mentioned numerous times in the findings, but HM will not be. This is because only two IRF schools adopted HM and thus mentioning it in the context of the findings would compromise the anonymity of the teachers who were observed and interviewed. But OCR can be mentioned since so many IRF schools adopted it. The second is workshop and intervention time. These are blocks of time devoted to skill development and remediation. In the high-performing teachers' classrooms each of these blocks of time was usually 30 minutes or more so that the combination of workshop and intervention was an hour or more each day. These blocks of time are mentioned because some teachers had almost all of their students for both workshop and intervention while other teachers had all of their students for workshop but only a portion of their students for intervention. There were other combinations also that were observed, so in the following account of the findings it is important for readers to keep in mind that some teachers don't have all of their students for both workshop and intervention whereas others have all or most of their students for both. The third term is supplemental services. These include Title I, special education, and English language learner (ELL) programs.

We now turn to the findings. The findings represent the themes that emerged from the analysis of the observation and interview documents. The themes are synonymous with the qualities or characteristics that these teachers had in common. There is some repetition across the qualities, but this is important since readers will build an increasingly detailed picture of these teachers and their classrooms as the qualities are introduced and discussed. The presentation is similar in structure to the core programs adopted by IRF schools. The core programs spiral, which means there is some repetition of old concepts and skills as new content is introduced and developed. Similarly, the discussion of the teachers' qualities spirals, providing room for repetition as new information is provided. The result is the construction of a multi-layered description of these teachers and what they do in their classrooms.

Fidelity to the core program coupled with extensive knowledge and awareness of its strengths and weaknesses. The teachers implemented their core programs with fidelity and knew why they were doing what they were doing because they had deeply reflected on their teaching, the core program, their students' performance, and their prior experience. One coach described a teacher in the following way: "She understands the purpose for everything that she does." The teachers' reflections, which had spurred them to do professional reading, attend workshops or conduct Internet research on ways to better teach and use the core program, helped these teachers build a professionally relevant foundation under their current practice. In short, they spoke with authority concerning the core program and why and how they were using it. All but a few of the teachers emphasized the importance of fidelity to the program. In the following quotation a teacher talks about the importance of fidelity to her success, "I think it has a lot to do with that I'm faithful to the core elements of the program. If I'm to teach reading, then that's what I do. Fidelity to the program, I think, is critical." But fidelity

for these teachers was not blindly following the prescribed core program and supplemental programs. Instead, fidelity was more complex and will be discussed and developed more immediately below and then throughout the remaining discussion of the findings.

These were very professional teachers who took their jobs seriously, spent considerable time thinking about and working on their literacy instruction and felt a moral imperative to teach **all** children how to read and to read well. They weren't blindly following the core program because they had been told to do so, but instead they approached the core program with a spirit of critical inquiry. All of the teachers thought their core program was good or even excellent. They appreciated the systematic instruction and the completeness of the materials. They also believed the programs to be well grounded in current research about early literacy learning. But it is important to note that all of these teachers were acutely aware of both strengths and weaknesses in their core program and were proactive in a variety of ways in capitalizing on the strengths and remediating the weaknesses (See Table S2-1 for a list of strengths and weaknesses.)

Table S2-1: Strengths and Weaknesses of OCR Identified by Teachers, Coaches, and Principals

OCR Strengths	OCR Weaknesses
Systematic instruction.	Vocabulary — both the type of words targeted by the program and the type of instruction and instructional materials built around the targeted words.
Sequential instruction.	Does not address the needs of high learners or low learners but instead focuses on middle level learners.
Completeness of materials.	Some stories are not well written or lack engaging content.
Having all of the materials (the funds were greatly appreciated so that all of the materials could be purchased instead of only bits and pieces of the program).	Some worksheets are poorly constructed.
Grounded in current literacy research.	Recommended time allotments for lessons are sometimes too short or too long.
Green band provides a solid foundation in word attack skills and is well planned.	Assessments don't match how the concept or skill is taught.
Strong emphasis on phonemic awareness and phonics.	Assessments sometimes emphasize concepts or skills more or less than what is found in the curriculum and instruction.
Supplemental ELL materials are effective.	Anthologies are too difficult for low readers and ELL students.
Provides common instructional frame work for all teachers and grade levels.	Fluency — not enough materials, specifically a lack of leveled readers for low readers to practice with.

	Not enough re-teaching for lower level readers and ELL students.
	Challenge materials are "really just busy work."
	Writing — not enough focus on this and not enough practice with real, extended writing.

Their proactiveness took a variety of forms that will be more fully described in subsequent paragraphs because a large part of their excellence was due to this shared characteristic. But one illustrative example will be provided here. Although these teachers used the core program with fidelity, they sprinkled their own ideas throughout their literacy instruction. For example, one teacher gathered information, activities and instructional ideas from another well-known and currently popular core program she was familiar with from a previous teaching position. She integrated that into her presentation, while continuing to research ways to effectively work within her school's adopted core program. The important thing is that these teachers thoroughly knew the core program and how their students were or were not served by it. Furthermore, they took steps to adjust the program to better fit their students' needs. There was no evidence that they adjusted the program for their personal benefit, such as adjusting it to better fit their personal teaching style or to make their teaching easier or less time consuming. Instead, they supplemented and adjusted the core program with a focus on student learning and how best to address their students' needs. Importantly, their supplements and adjustments were data driven and in aggregate were very time intensive.

It is important to make it clear what supplements were used and what adjustments were made. Supplements included adopted programs such as Read Naturally, but all of these teachers supplemented the core program with teacher-made materials and activities or materials and activities found online or from other resources. These supplements were strategic in that they addressed an identified weakness in the core program or an area of weakness in students where the core program did not provide enough practice or instruction. A quite common area that was supplemented was vocabulary. According to the teachers, the lesson scripts and the worksheets and activities in OCR don't develop the words thoroughly enough due to an absence of multiple examples and visual referents for the words. This, they said, is especially true for their ELL learners. Teachers altered the lesson presentation to incorporate more discussion of the words, including examples and connections to other words and to the students' lives. Teachers also either altered existing OCR worksheets, made completely new ones or found worksheets from other sources that had adequate visuals attending the vocabulary words.

Time was another variable that teachers altered. None reduced the 90-minute reading block, but some adjusted how much time they spent on various core lesson components when the time recommended by the publisher was either too much or too little. These adjustments were usually guided by notes the teachers had made when teaching the lessons in previous years. This underscores the high degree of knowledge these teachers

had about the programs and their teaching, along with the high degree of critical reflection that was part of their practice. But not all teachers adjusted time. Others were conflicted about the time issue in that they knew time adjustments were needed but they couldn't make the adjustments because it would be seen as a lack of fidelity to the program. This rigidity was mentioned by all but a few of the high-performing teachers, and the few who didn't mention it were in situations where they were given more freedom to interpret the core program. All expressed the desire, ranging from minimal to acute, to have more flexibility in the delivery of the core program, and they truly believed that given this flexibility they could serve their students even better. They did not mean by having greater flexibility that they would significantly diverge from the adopted core program. They liked their core program and praised it for some significant and important strengths, and all but a few teachers said that staying close to the program — that is teaching it with fidelity — was important to their success. What they meant was having the ability to make the adjustments that they felt would better serve their students. For example, several 3rd grade teachers mentioned in interviews that some of the units in 3rd grade are too long for what is learned by the students whereas others may be too short. They would like the "freedom" to make time adjustments at both the unit and daily lesson levels. Some are doing this already "under the radar," but most are not since they feel compelled to follow the program with strict fidelity. An observer made reference to this issue in the following note:

> She [the teacher] made it very clear that she likes Open Court, but doesn't like the inflexibility of Reading First. She said it "leads to missed opportunities to teach to the moment or to cover things students are interested in and need to know about to be motivated to learn." She went on to say, "Even without Reading First, I would still be consistent with the program because it works. I just want some flexibility to make good choices for my students." I [the interviewer] asked if she thought that other teachers would still follow Open Court without Reading First pressure and she admitted that she did not know and she said, "Maybe not."

Another teacher said, "The rules I break are because my students don't need something, or the inflexible time requirements do not give me adequate time." These excerpts bring to the surface the current dilemma that IRF faces. The high-performing teachers who are feeling constrained should probably be given more flexibility with the program to see what results they can achieve, but not all teachers are probably ready for this freedom since some might not be able to use it to the advantage of their students' learning. Furthermore, they might use the freedom, as the teacher suggested above, to move away from the core program, which would undermine the IRF initiative. It is thus important for leadership and oversight structures to be put in place in the buildings so that strategic flexibility can be built into some classrooms and monitored for its effects on student learning while other teachers are provided continued support to improve their teaching while learning how to better utilize the core program.

One principal said, "Reading First can stifle creative teachers. On the other hand, it lays a foundation of understanding." What is important to understand at this point is that for some teachers the foundation has been laid and now it is time to see what they build on it.

But for others the foundation is still being put in place. Schools will have to differentiate their approaches to their teachers the same way teachers have to differentiate their instruction for their students. As one highly-rated coach remarked in an interview, "She [the identified teacher in the building] could do just as well with her kids without Reading First." Whether or not this statement is true — and most of the identified teachers would probably disagree with it since they are quick to emphasize how much they have gained from IRF — is not as important as what it implies in this particular school. All teachers are being treated the same. All teachers are charged with and monitored for strict adherence to the core program because without this discipline some teachers would stray from the program and their students might not receive the type or quality of instruction they need. But treating all teachers the same because of the possible "misbehavior" of some will not serve IRF in the long term as new problems arise and new solutions need to be explored.

The importance of building flexible approaches to teachers and their instructional programs can not be over-emphasized. A few of the high-performing teachers were under a great degree of stress over this issue. One was interviewing for a teaching position in another school the evening of the day the classroom observation and interview were conducted. She was open about her wanting to leave and she was at the very top of the group of high-performing teachers. She was simply outstanding and her coach and principal strongly corroborated this assertion. Another teacher said during an interview, "I don't know how much longer I can go on." This was in reference to her feeling so constrained by IRF. It is important to note that her comment had nothing to do with the heavy workload found in IRF schools. She loved her work and spent hours outside of school preparing her lessons, but she longed for more freedom so that she could better address her students' needs. And, importantly, she had not been teaching that long. IRF needs these high-performing teachers to stay in these schools because they can become an excellent source of knowledge and inspiration for all of the teachers across the network. They are also the teachers who have the skills to attack new problems as they arise.

It is important and interesting to note that these teachers did not always want time to slow down in the core program, but instead they wanted the freedom to move more rapidly through some of the core program materials and units. OCR is quite fast-paced so this desire is perhaps somewhat counterintuitive. For example, one 2nd grade teacher said, "The curriculum is too long in places and wastes instructional time." Several 3rd grade teachers voiced concerns similar to this. One said about some of the units in their core program, "The first three are long and include phonics, etc. and then with Unit 4 the phonics stuff is dropped and there really isn't enough to fill the 90 minutes. I have to spend the 90 minutes but with some flexibility I could probably get to Unit 6, which I never do now." A principal said about her high-performing teacher, "Sometimes the redundancy isn't good for her. An excellent teacher will make it their own, where another teacher might need to go through it several times. She could probably cover it in less time than what is required."

In summary, it is important to discuss this theme of frustration that came through, although it is also important to emphasize that the frustration manifested itself in varying degrees depending on the teacher and her personality and situation. Several expressed frustration with their core program because they could see places where it was lacking and they struggled with the dilemma of 1) using the core program with fidelity and not fully addressing their students' needs or 2) **"**illegally" supplementing or altering the core program and putting their school at risk for losing funding. This was a source of quite severe stress for several of the teachers and less so for others. But all felt some level of constraint in their implementation of the core program because of IRF requirements. IRF needs to look at this issue carefully since some of their best teachers in the network feel overly constrained and unable to adequately meet all students' needs as a consequence of the way the core programs are being implemented. The issue of fidelity to the page versus fidelity to the program arises here. These are fine teachers who like the adopted core program. Specifically, they like its overall structure and its emphasis on phonemic awareness and phonics. They also like the structure of the school day that provides them time to work with individual students and small groups. Every teacher liked the green and red bands and especially emphasized the importance and effectiveness of the green band in OCR for most of their students, including their ELL. But it is important to qualify this statement. Teachers felt quite positive about the green and red bands but they also were quick to point out that those bands are seldom appropriately leveled for their lowest-performing readers and their highest-performing readers. In summary, these teachers are not looking for an excuse to jettison fidelity, but instead they are looking for ways to make the program better and more effective for students. Timothy Shanahan, former president of the International Reading Association, and a strong proponent of the basal series in use in IRF classrooms has said, " … teachers will still need to make modifications to even the best of programs" (*Reading Today*, April, 2007). This is exactly what we found happening in high performing IRF classrooms.

Supplementing the core program. In addition to supplemental programs like Read Naturally, all teachers supplement their core reading program with other materials they develop or locate from other sources and all teachers adjust the instruction in the teacher's manual. This supplementing and adjusting occurs in the 90-minute block, workshop and intervention. A coach said of one teacher, "She's determined and conscientious. She finds a lot of things to add to the program. She's a lifetime learner who embraces a lot of different ideas."

Teachers are adjusting the instruction that is prescribed in the teacher's manual and they are making and/or locating worksheets and activities to supplement the core program. Some of these adjustments are not unique to the teacher, but are occurring throughout the school. For example, one school has moved away from transparencies that OCR requires and uses charts and pictures so the kids have visual referents that remain after the formal instruction. When interviewed the teacher, coach and principal all talked about why they no longer use transparencies. With a transparency, once it is removed from the overhead projector or the projector is turned off, the visual referent is gone. This particular school found this to be a significant weakness in the OCR program and developed a school-wide solution to the problem, namely the use of charts and pictures in place of the

transparencies. The teacher observed in this school was using the charts and pictures and had been part of their development and implementation. Another teacher in a different school talked about using Read Naturally in a slightly different way that aligned better with OCR and now three of her colleagues are doing the same since it appears to work so well. But there were many more adjustments to both curriculum and instruction that only the high-performing teachers were making in their classrooms. This was, for the most part, a distinguishing characteristic of these teachers.

Their adjustments of the prescribed instruction in the teacher's manual will be discussed more in subsequent sections. What will be discussed here in greater detail are the changes and additions they make to instructional materials. This characteristic was discussed some above, but more detail is needed to underscore its importance. These teachers were very motivated to make materials or find materials that would better serve their students' needs. As one teacher said, "I work hard and I'm dedicated. I don't consider myself dead yet. I don't like to get stuck in a rut. I've been in so many programs, that I pull the strengths from each of them. I've been in a lot of districts."

Observers/interviewers were shown publisher-made materials from sources such as Modern Curriculum Press and Frank Shaffer. But teachers also made large amounts of materials themselves, and the time they spent doing this was truly exemplary. For example, one teacher wrote additional stories that highlighted skills and concepts taught in OCR since she believed, like many of the other teachers, that students didn't read enough in OCR and needed more practice and reinforcement of the skills. Another teacher said that she made "rapid word charts of high-frequency words, flash cards, word walls, games, word sorts, centers for writing and word study, and worksheets which revisit and reinforce concepts from the basal." She said that with the OCR anthology, "There are worksheets – usually there are 2 to do. If they're weak and I don't think they will accomplish anything, then I do supplement with teacher-made worksheets." Another teacher showed the observer numerous board games and computer-based games that she had developed to help her better target her instruction during workshop and intervention time. The amount of thought and time that she had put into these materials was striking. She jokingly said, "I want to get an audience with an Open Court trainer to see if I can sell some of these to them." Another teacher said that she adjusts time for various sections of the program. She looks at the unit tests before the unit starts to see what is being asked and then she supplements the program to make sure the kids get enough instruction and practice. She added that the unit tests are not always well written and the program doesn't "always hit the stuff on the test enough or well enough for the kids to do well on the tests." She also mentioned that "A unit test will have two questions over a skill and if the kid gets one wrong they aren't proficient. These tests have to be turned in!" Like most all of these teachers, however, she said she likes OCR and thinks that "it is a very good program."

To summarize, these teachers knew their students in great individual detail. Because of this knowledge, they adjusted their curriculum and instruction throughout the day and used workshop and intervention time (if they had the students for intervention time) to further customize curriculum and instruction for each child. They took advantage of

workshop and intervention time to provide additional instruction and curricular materials tailored to each child's needs. Teachers made a variety of games (board-based, manipulative-based, computer-based, etc.), worksheets and hands-on activities that provided additional practice and application of important concepts and skills. Games were the most common type of activity the teachers talked about developing. In some instances individual teachers made all of these materials and some of them spent a considerable amount of time outside of the school preparing them. In other cases, teachers within a grade level collaborated to develop ideas and materials. This team effort also took considerable time, but at least it was spread among several instead of an individual teacher taking full responsibility. One thing is important to mention here. All of the high-performing teachers said that preparing for workshop was very time intensive but worth the effort because that is where they get the opportunity to focus on the individual child. Workshop will be more fully discussed below.

Excellent classroom management and warm, safe classroom environments. The teachers created classrooms that were warm and friendly places where students were comfortable and willing to take risks. But warmth and friendliness were only a part of the environment and don't adequately reflect the complexity of these teachers' classroom management skills. There were many other intersecting and intertwined attributes of these classrooms that combined to create very positive spaces for student learning. For example, routine was important. It was immediately apparent to observers that the students were thoroughly steeped in classroom routines. They knew where and when to go and when and they knew how to behave once they got there. A coach remarked about routine and risk taking in one teacher's classroom: "Just her consistency so that her routine is there day in and day out. Very friendly and open. Displays that environment for her kids that risks are ok and they're comfortable."

These teachers also held students to high standards of conduct. Children were treated fairly and consistently but they were also expected to attend to the lesson, their work and the good of the group. There were classrooms where hardly a child was reprimanded for inattentiveness or misbehavior during the entire time the observer was present which was sometimes longer than two hours. These were amazing classrooms where the teacher and children were so deeply engaged in learning that inattentiveness was rare. A principal captured the essence of these almost magical teachers in the following comment: "Energy level and personality are the biggest differences. She has a rapport with children and knows how to connect to them – to read their heart." In short, these teachers were charismatic educators who felt so strongly about all children learning that their classrooms became hubs of activity centered around learning how to read or learning how to be a better reader. These teachers focused on helping their students become attentive learners who cared about their own learning and that of others. They not only focused on the immediate literacy needs of their students but also the long term needs such as how to be a good person who will be successful now and in the future. They oftentimes accomplished this by showing students the personal relevance of what they were learning. Other teachers reprimanded children but they did so in a kind way, and didn't have to do so very often.

Teachers were aware of each child in their classroom and whether that child was attending and learning at any given moment. Of course, there was a range of this "with-it-ness," but overall these teachers were quite attuned to their students at an individual level and whether they were learning or not. As soon as they detected a child losing attention or falling behind in the lesson, they were calling out the child's name, directing a question or comment at them about the lesson, and/or moving next to the child to get them back on track.

In short, these classrooms were well-managed places where everyone knew what to expect, where everyone was valued and where reciprocal trust and respect were key components of the classroom culture. One teacher said when she was asked about what makes her successful, "I love what I do and my kids know that. I am honest and trusting of my students and that helps them in their daily work." Instructions were clearly stated to the children, as were expectations. These teachers directed the children but in such a way that the children didn't feel like they were being forced or brow beaten into submission. Instead the children felt like they were part of something exciting and important. Of course, enthusiasm waxed and waned over the course of up to three hours of literacy instruction, but these teachers had wonderful capacity to read their students and adjust their instruction and their energy levels to maximize engagement.

However, each of these classrooms was quite different. Some were more militaristic where there were very clear procedures that the children were instructed to follow. In these classrooms there was a place for everything and children followed very specific patterns of behavior. For example, there were special chairs where the children would sit while they waited for their turn to talk with the teacher or there were different groups for vocabulary development that were given colors which denoted what materials they were to use and where they were to form their group in the classroom. In other classrooms there, of course, was still plenty of routine and organization but they didn't come to the fore as much, and consequently children's movements weren't choreographed so tightly. In short, the ends were the same for all of these teachers — namely superbly managed classrooms where all children could learn, feel good about their learning and take the necessary risks to do so — but the means to get there varied. Some teachers had excellent hand and verbal signals whereas others were not as concerned about this and used their enthusiasm and excellent questioning skills to keep the students engaged. Some teachers had a quite rapid pace during the 90 minute block. One observer wrote in her notes, "I can't keep up. This is just too fast for me, but the kids are fully engaged." Whereas other teachers had more relaxed paces but were still able to cover all the material and keep students engaged.

The following quotes from a teacher and a coach capture the essence of classroom climate and management. The first quote is from a teacher talking about her management:

> I really feel like I won't teach if they're not all listening. The students in my class can verbalize why they're here and why they need to learn. When I'm here I don't have discipline problems, but when I have a sub, the kids aren't as good. One of the things that helps my kids is that I give them a high level of attention always. I have a

higher level of focused attention from the kids. I nip it in the bud. I don't let them go at all. Once they get off task, it's like 6 years getting them back.

This first quotation exemplifies how these teachers command student attention but they have rapport with their students so that attention is not seen as submission to some over-lording power but instead as a constant in the classroom that is expected of everyone. The teacher quoted above also emphasizes the importance of relevance to her students and to her success when she says that her students know why they are in school and why they need to learn. This motivates them to attend and behave since they understand the importance of the work they are doing. Finally, she emphasizes that she gives a high level of attention to her students at all times and they reciprocate by giving her their attention. When they don't, as all children will do at one time or another, she nips it in the bud.

In the following quote, a coach emphasizes the high level of engagement that comes as a consequence of pace, rapport and tight management:

Her children are highly engaged. There is never any downtime. The pacing is very rapid fire. Children who are slower are also with her and it pulls them along. They love her so much, they want to be right with her. Her feedback is immediate and excellent. She does a lot of modeling/correction. As soon as there is a mistake, she models it and the children repeat. So they understand it immediately, but they never feel badly about it because it's always fun, it's always a game. Children do not get away with anything. If they don't get their homework done, they know the consequences and what to do to solve it. Excellent classroom management.

Other important attributes are mentioned in this excerpt. Modeling and immediate feedback are characteristics of these teachers. These teachers are in tune with their students' learning at all times. No teachable moment is wasted. When a mistake is made the teacher catches it and models the correct answer for the students. These teachers will also model incorrect responses and let the students "catch" them and correct them. This "catch the teacher" ploy is a good example of the give and take interchange that occurs in these classrooms. More discussion of the amount and quality of the dialogue in these classrooms will be discussed below.

So where did these excellent management skills come from? Answering this question gets at the heart of the art and science of teaching. These teachers had very objective behaviors that contributed to their excellent management. For example, they had very clear expectations for their students, and they provided very clear directions for academic tasks and instruction. Teachers also provided clear guidelines for personal behavior — how students were to conduct themselves while they performed their tasks, how they were to interact with each other and how they were supposed to transition from one task to another. But they also had intangible qualities that made them engaging educators. Of course it is hard to pinpoint a single causative agent. Management in these classrooms was highly complex and stemmed from the teacher's knowledge of their children, concern for each child, their planning and organization, and their personalities. The following quote from a coach alludes to this tight interplay between personality and

technical knowledge of the program and children: "Both teachers have strong relationships with their students. They are also 'skillful' with Open Court so they can attend to students' needs." In short, strong relationships and skillfulness with OCR are both integral to good management. The strong relationships are the product of skilled use of the program so student needs can be addressed, and in turn the skillfulness is a product of having strong relationships with students so the teacher is motivated to use the program to its fullest extent. Without being overly simplistic and overly reductive, since management and motivation in classrooms is such a complex phenomenon that depends on such intangibles as teacher personality, one thing did stand out: These teachers had passion for teaching and deep concern for their students. Following are some excerpts showing teachers' thoughts about what they bring to teaching that influences their success:

> I was more accepting of OC in emphasis on fidelity at the beginning. I bring extra effort and passion to my teaching.

> I think the passion I bring to teaching is communicated to students and so they become passionate about learning.

> I am more passionate about teaching and my students than the other teachers. I often work with my struggling students during lunch — my colleagues aren't willing to do that.

In summary, classroom climate and management were intimately intertwined in these teachers' classrooms. The teachers had excellent management skills but a lot of their success in management came as a consequence of their successfully creating and sustaining warm, inviting and exciting classroom environments. Management is more than just getting your class under control. It has to do with the energy, enthusiasm, drive and commitment of the teacher. In short, high expectations from the teacher transfer to the students. But high expectations come from genuine feelings of care, trust and resolve that stem from a deeply-held belief that all children can learn and that a teacher can dramatically influence this.

Focus on the individual child and high expectations for all. A focus on the individual child and high expectations for all children were universal in these classrooms and occurred throughout both whole group and small group instruction. These attributes have been discussed above, but more elaboration is in order. These teachers didn't see a class of 25 students. They saw 25 individuals in a class. Of course, during the 90- minute block they focused on whole class direct instruction, but during this time they also monitored each individual in their classroom for understanding and engagement. If a child wandered, they called out his or her name to ask a question or to prod him or her back into attentiveness. If they sensed a child not understanding, they stopped and asked a question or gave another example. They were constantly aware of the state of learning in their classrooms and this awareness was focused on the individual child as well as the group. A hallmark of these teachers is that they stopped whole class, small group or one-on-one instruction when the response was not at the level they expected. As one

principal remarked about a teacher, "She stops instruction if the kids aren't performing at the level she expects. Not all teachers do this." We have witnessed this in our observations over the past two years. High-performing teachers stop instruction as much as necessary to get students to rise to their expectations. When they stop instruction, they provide very specific feedback to the students as to why instruction stopped, they sometimes model appropriate responses for the children, or they ask questions of the children to prompt their thinking and reflection on their performance. A teacher might ask, "How do good readers read a sentence like this?" The sentence might have an exclamation point at the end and the children read it without emphasis. In short, teachers intensely work with the children to bring them to the expected level of performance.

During workshop and intervention, these teachers excelled. They often spoke of workshop and intervention time being the time when they could "really work" with the students on literacy. They looked forward each day to workshop and intervention because they represented yet more blocks of precious time when they could continue to focus on the individual needs of the children. These teachers were instructional engineers. They were engineering workshop and intervention time so that each child in their charge learned to read. They took pride in this and felt the focus on the individual student was key to their success. One teacher when comparing herself to her colleagues said, "I've tried to differentiate in my classroom, more so than in the other classrooms." And the primary vehicle for this differentiation was workshop and intervention time. When they talked about the strengths of Reading First and their overall literacy programs in their schools, a common response was the importance of the small group and one-on-one attention children received during workshop and intervention time. Of course the teachers felt their time with the children in small groups or one-on-one was important, but most of them also felt that the time the children spent in small group or one-on-one instruction with aids or supplemental services teachers was also important. This will be discussed more in a subsequent section.

These teachers truly felt in their hearts that every child in their classroom, no matter the ethnicity or socio-economic level of the child, would learn. There was no apologizing for a child's under-performance because of home life, SES or ethnicity. Instead, these teachers treated all children the same when it came to expectations for learning. A few of the teachers knew some Spanish and used it regularly to help build bridges between the English the children were learning and the Spanish they knew. But most were English-only speakers who reached out to Hispanic children by showing equally high expectations for them and lending them a helping hand customized to their particular needs so that they too could rise to those expectations. The teachers talked about spending lunch hours with children helping them or spending time before or after school. Some talked about home visits or attending soccer games. But not all of the teachers did these things, many just created a dynamic classroom environment where children were comfortable taking risks and the teacher was always there to help them step to the next level.

During interviews teachers continually mentioned high expectations as something that they possessed to a great degree, and coaches and principals mentioned it regularly as an

important characteristic of the teachers. These high expectations and the drive for children to reach them underpinned much of the hard work these teachers devoted to their teaching. For example, when they saw ELL students struggling with the vocabulary in the core program, they knew that without additional intervention the students would not be able to meet expectations. So the teachers took the time to figure out how they could emphasize vocabulary throughout the day. Thus, we observed teachers having the children quickly act out the vocabulary words during the 90-minute block and we observed a variety of vocabulary activities during workshop or intervention time, including teacher-made worksheets that had more pictures on them to associate with the words.

When one teacher was asked how she differed from her colleagues, she responded, "Oh my…. I think my expectations are higher than most. I think I expect a lot more, and I get a lot more. And the discipline. I don't let bad behavior get in the way of learning and I see that a lot in other classrooms." Notice again how expectations and discipline have a reciprocal relationship. The teacher has high expectations for student learning and that in turn creates high expectations for behavior and vice versa. Another teacher provided some wonderful recommendations for other teachers while also alluding to her having high expectations:

> Interviewer: What does Reading First need to do so that all teachers attain the level of success that you have?
> Teacher: I would think just more staff training. I guess they consider me a successful teacher. Maybe others could come and learn from others. I think I have something good going in my workshop time; I'd like to share that with others. The expectations are more of a personality thing. It's how much you're willing to put into it.

This teacher also articulated the important reciprocal relationship between high expectations and the amount of effort a teacher is willing to put into her teaching. If a teacher truly has high expectations for her students, then she will invest the effort and time to bring her students to that level. Conversely, it may be that a teacher who is willing to invest the time and effort into teaching well has high expectations for her students. Either way, the result is students meeting those expectations. As one teacher said, "You take what you're given and you make it work. Have fun with it. I like the level of material that's involved. They have to reach to get it." This teacher enjoyed the high level of the core reading materials because it facilitated her having high expectations for her students and in turn her teaching.

High levels of academic press and full use of time. Academic press was a construct developed and explored in the 2005-2006 evaluation. Research has shown it to be a strong correlate with student achievement. Stone et. al. (2005) define academic press in the following way:

> Academic press refers to several dimensions (see Middleton and Midgley 2002; Phillips 1997). One dimension relates to overall academic norms (e.g., attendance,

27

homework completion) and standards (McDill, Natriello, and Pallas 1986). This dimension also refers to the extent to which instruction is organized, focused and goal-oriented (Phillips 1997). Another dimension focuses on the degree to which students are pushed for high performance, usually in terms of grades and achievement tests. A third dimension is concerned with mastery and understanding and taps the extent to which students are expected to continually put forth high levels of effort, are deeply questioned for understanding of material and are given progressively more challenging tasks. (pg. 4)

Academic press occurs when a classroom is efficiently run and all parties, including teacher and students, get along well together, work hard together, enjoy their work and focus on both immediate and long-term outcomes. Academic press is closely related to the high expectations discussed above. High expectations create an environment conducive to academic press and vice versa. These attributes have been previously discussed and illustrated but additional details are provided here. Stone, S., Engel, M., Nagaoka, J., & Roderick, M. (2005). Getting It the Second Time Around: Student Classroom Experience in Chicago's Summer Bridge Program. Teachers College Record, 107(5), 935-957.

Classrooms with a high degree of academic press are places where focused work is by far the norm not the exception, time is not wasted, student success is the norm and student achievement reflects this. The high-performing IRF teachers had high academic press classrooms but they achieved the academic press in quite different ways, resulting in a range of "classroom personalities." This range of classroom personalities has already been revealed in some of the discussion thus far and will be further illuminated in subsequent sections. What is important to address here is the superb use of time that these teachers exhibited. These teachers used every available moment in the classroom to be teaching. They started lessons on time according to the daily schedules. They brought the children to attention in kind, humorous or fun ways and crisply began the lesson for the day with clear directions, expectations and outcomes. They also made connections between what was learned during previous lessons and what was about to be learned. Transitions were oftentimes masterfully executed in these classrooms, with even this brief amount of time put into the service of learning. For example, during one transition an observer wrote, "When the kids were getting ready for recess the teacher asked each group to act out a vocabulary word before they could leave. She said, 'Ok, Josh, give me the action for survive.'" During other transitions observed in other classrooms, students were directed to sing the sounds they had just learned or reviewed while they pulled a different book or worksheet from their desks or moved from one location in the classroom to another.

This efficient and full use of time extended to how lessons were conducted. Teachers packed as much as possible into the time they had for literacy instruction. More examples of this will be provided when the emphasis on vocabulary is discussed below, but an example will also be provided here. In the following excerpt from an interview, the teacher talks about using the blending board to also cover concepts found on the Idaho Standards Achievement Test (ISAT):

"With the Open Court (reading block), we do blending which lasts about 30 minutes. One thing about blending board, I add a lot to it. That's where I find I can get in a lot of those skills that are going to be taught on ISATs, so we talk about the vocabulary on the ISATs and I also write it on the board. Making sure that if it fits with that blending board, then I use it. It's a good spot to teach in context."

These teachers were thinking about their instruction, curriculum and assessments all of the time. They knew all of them in such intimate detail that they could leverage any component at any given time to maximize student learning. This teacher's use of the blending board to also cover ISAT skills is just one example of this.

High level of daily preparation. None of these teachers came to class unprepared or under-prepared. Many of them talked about the considerable amount of preparation time they spend getting ready for each day of school. One 30-year veteran teacher said, "You can't come in here and wing it any more. You have to be ready every day." These high-performing teachers were ready every day. Oftentimes they made the lessons look easy, as if they could do them in their sleep, but underlying this apparent facility was intense preparation so that the lesson would be paced properly, weaknesses in the lesson could be addressed, strengths could be capitalized upon, and workshop and intervention time could be focused on each child's needs. These teachers were hard-working, dedicated professionals who saw their work as absolutely important. They felt strongly that every child in their room would learn and it was their job to keep working until each child was doing so. When teachers were asked how they differed from their colleagues, 12 said that they were willing to put in more time, energy and thought to their literacy instruction. They didn't say this in haughty or superior ways. Instead, they said it in matter of fact ways without judging their colleagues. These teachers were no-nonsense educators who came to school each day ready to teach and committed to making incremental progress with each of their students each day. The following excerpt from a teacher interview is typical of what these teachers said:

> Interviewer: Is there anything else that you've thought of as we have been talking that would answer the question "Why do you think you are doing so well with your students?"
> Teacher: I spend a lot of extra time outside of regular hours. I would think that contributes to my success.

All of the teachers spent time outside of contract time preparing. We believe that a substantial portion of their success boils down to their having high expectations for all students and then spending the time to build the program to match those expectations. That means they find extra materials, new teaching ideas, games, etc. to make their program work for all students. It's a simple formula but very hard to make universal.

Data driven workshop and intervention periods that focus on the individual student's needs. Data was a very interesting variable among these teachers. There were data converts who started out intensely disliking assessment data, but who had come to love it

and heavily rely on it. There were those who loved data from the beginning, and there were two or three who saw the utility and importance of assessment data, but still were not that enamored of it and continued to use their experience and informal observation for most of their instructional decision-making. Following are some excerpts illustrating teachers' use of data and its important role in workshop and intervention. A coach said about two identified teachers:

> We provided ways to organize data. Sally and Jane (pseudonyms) have come up with their own ways to track data. It's all plotted out and they keep track and look at it to see consistent patterns for kids who are struggling. They use the workshop time for addressing these needs. One thing about the three of them [speaking about the grade level team of three teachers]… they know that the data is important but they also know how to balance that with what their teacher instincts and knowledge tell them. Some people in RF have lost that, but these three have never lost that.

In the following excerpt a teacher reflects on her use of data and how it has impacted her:

> We make spreadsheets. This has been a positive thing for me also. We list the skills covered and we color code them as to high, medium, low and decide who needs the work and who doesn't. They form the groups for the workshop and intervention this way.

The excerpts reveal that teachers use data to drive their workshops. They also use their professional expertise in the form of informal observations to contextualize and enrich the information they receive from the more formal assessments such as IRI, TPRI, DIBELS, Core Phonics Survey, etc.

What is important to underscore is that all of these teachers use data in one form or another from **individual** children to make day-to-day decisions about how they approach that child instructionally and curricularly. Coaches and principals often commented about a teacher's knowledge of and use of assessment data at the individual child level. For example, a coach said the following:

> Her assessment is important. The way that she uses the assessment. She has a deep understanding of the assessment and what she needs to do. She has a deep understanding of the individual needs beyond the assessment.

So, for these teachers assessment data was an important way to look at the individual child, but they also looked beyond the data to the personality of the child and the way the child learns best. Perhaps with the risk of falling into cliché, these teachers had the whole child in mind when they decided how to approach the child to further their literacy learning.

These teachers were education engineers. They recognized strengths and weaknesses in the core program, the supplemental programs and the individual children in their classrooms, and they carefully and strategically processed all of this information along

with assessment data to structure their daily workshop and intervention times to leverage the strengths and ameliorate the weaknesses. They saw these blocks of time during the day as wonderful opportunities to work with individual children and small groups of children on areas of weakness that they had identified through data and teacher observation. These teachers looked forward to these times during the day because it was during workshop and intervention that they could best achieve their goal of individualizing instruction so that each child had the greatest opportunity to learn and improve his or her reading skills. As one teacher said during an interview, "The workshop is key because it is individualized. Also, the concept introduction as done by OC is important and well planned."

This focus on the importance of thoroughly preparing for workshop extended to the use of paraprofessionals in the classrooms. The teachers spent considerable time and energy writing lessons for their paraprofessionals and making sure they did what they were supposed to do with the students. In the following excerpt from an interview with a teacher this attention to detail is underscored. The teacher was asked to what does she attribute her success. The excerpt is a mixture of verbatim teacher response and interviewer paraphrasing of what the teacher said:

> Teacher: I follow the program and expect students to listen attentively.
>
> Interviewer Paraphrase: She also admits that she is very organized, plans thoroughly and she thinks that her enthusiasm is an important element of her teaching style. I agree with this as I saw a tremendous amount of energy and enthusiasm for what she is doing. She showed me her plans and they were detailed, typed and included not only the schedule for the day, but detailed lessons for the aide and the volunteer who come in for the workshop/intervention hour.

This teacher spends considerable time not only preparing the lessons that she will conduct but also lessons for the paraprofessionals and volunteers in her room. In some schools this work was being shared among the teachers at a grade level, but it was just as often the case that the individual teacher did it by herself.

When taken in aggregate, a wealth of supplemental curriculum has been developed by dedicated IRF teachers. IRF should develop an "idea bank" where teachers can go for helpful tips and suggestions about such important topics as fluency, vocabulary enrichment, etc. The "idea bank" should primarily focus on providing support in identified areas of weakness in the core program, helping with building effective workshops and providing tips on how to improve instruction. As much detail as possible should be provided concerning why the idea is important, where it can be applied and how long it takes.

To summarize, the focus these teachers had on getting every one of their students to perform was heartening to see. They didn't give up, but instead diligently worked day after day, constantly making adjustments to curriculum and instruction as new data came in so that all children progressed. Workshop and intervention were key to this progress.

The teachers' emphasis on the importance of workshop to their success and their willingness to spend considerable time developing lessons and materials to make their workshops effective day-in and day-out were hallmarks of the group. Workshop took a variety of forms in these classrooms, all of which appeared to be effective based on observations, interviews and test scores. There were those teachers who developed quite customized small-group activities that were highly focused on skills and concepts with which the students needed more practice and application. Other teachers ran a traditional centers-based workshop where students cycled through the centers while the teacher and paraprofessionals worked with individuals or small groups. These centers were meaningful activities focused on basic skills. The students worked at them diligently, but the level of individualization wasn't as great as in other classrooms.

Focus on vocabulary and vocabulary development, especially for ELL students. As discussed above, vocabulary is being emphasized with approaches and materials beyond those available in the core reading programs. Off-the-shelf vocabulary programs are not being used by these teachers. Instead adjustments are being made to the way the core reading programs approach vocabulary. For example, teachers remarked that OCR chooses to emphasize some words that are not the type of words that children need to know now or even perhaps in the future. The words are sometimes just too obscure, so teachers choose other words from the stories that are more appropriate. Oftentimes their criteria for appropriateness include such things as relevance to students' lives and key words that will aid comprehension. Additionally, teachers adjust the instruction as it is stipulated in the teacher's manual to augment the focus on vocabulary. They spend more time during the 90-minute block discussing words and making connections to other words, the children's experience and previously learned information. Publisher-provided worksheets are modified or replaced with others. During workshop students receive additional activities and/or worksheets focused on the targeted vocabulary words. But it is important to point out that not all of the teachers adjust the treatment of vocabulary during the 90-minute block. There were instances where teachers were not allowed to stray from the teacher's manual. These teachers then focused on vocabulary during workshop and intervention. For instance, while reading decodables during workshop, one teacher explained and made connections for students concerning the words *dawn* and *lawn*. She didn't feel like she could stop and talk about common words during the 90-minute block, but felt like it was within the boundaries to do that during intervention/workshop.

Finding more pictures to put alongside the target words was a common adjustment. The following excerpts from interviews reveal the importance of using pictures to illustrate vocabulary words:

Teachers supplement Open Court vocabulary with teacher-made notebooks with sentences and pictures. They do this during workshop. (Coach comment)

She also uses many visuals (again often homemade) to show a word she knows students will need. (Interviewer paraphrase of teacher response)

She has spent considerable time creating picture cards to match with the vocabulary cards that come with the Open Court program. She gets the pictures from the Internet. The students do a matching game with these. (Interviewer paraphrase of teacher response)

She stops with all students and asks, "What does this word mean?" She says she really takes time with them all to visualize and explain vocabulary, even if Open Court doesn't tell her to explain a word, she knows when to do it. (Interviewer paraphrase of teacher response)

These teachers' emphasis upon using pictures to illustrate vocabulary words should not be oversimplified. Their emphasis was more than just holding up pictures or placing a picture near a word on a worksheet. Six teachers mentioned SIOP training having an impact on their instruction, primarily in the area of teaching vocabulary and the importance of finding visuals to go with vocabulary words. The use of pictures was the overt manifestation of these teachers' knowing that vocabulary was a weakness in both their core program and their students and then their reflecting on how they could change their instruction to address the needs.

The emphasis on vocabulary went well beyond pictures. Teachers were observed acting out words and having the students act them out. The following excerpt, which is a note written by an observer/interviewer, draws parallels between what was observed in the classroom and what the teacher said in the interview about vocabulary instruction:

She emphasizes vocabulary development. The kids acted out the word *exploded* from their pilgrim story before they even read the story. She fully developed all of the vocabulary words before they read the story. She ties the words to their lives. Later on in the interview she says she has them act out words—the whole class does this at the same time. She really stresses vocabulary.

Similarly, the following teacher interview excerpt shows the importance of making vocabulary come alive for the students through both pictures and action:

The kids can read a decodable book sometimes, but not have a clue about what they're reading about. We take a picture walk and talk about the pictures. There just has to be a lot of visuals so we go through and talk about them in the anthology too. I suppose there are times – like when we were talking about the dog today. So today I pretend that I'm walking a dog – going forward. I taught for many years so I'm very cognizant of those kids because of that. (Interviewer paraphrase: She does more previewing and having the kids put it in their own words what's happening in the pictures than what is expected from the series.)

In addition to pictures and acting out words, numerous vocabulary games were observed in classrooms or were mentioned by teachers and coaches in interviews. Teachers and coaches emphasized the importance of this expanded focus on vocabulary development as being a key component of their success and also especially important for ELL

students. The following excerpts illustrate the importance of dynamic and varied vocabulary instruction to ELL student performance:

> Interviewer: What components of your "core" reading series are effective with Hispanic students?
> Teacher: The Green Band—phonics and the sound cards, but it must be accompanied with vocabulary instruction.
> Interviewer: Which are ineffective?
> Teacher: Again, the vocabulary is not adequate.
>
> I think adding in really focusing on vocabulary which is terribly important for any ELL school. (Teacher interview)
>
> I think I use a lot of ISOPP [SIOP] things where I tell a lot of stories and act out things and have kids act out things. And using words in a lot of different directions. And using vocabulary that is 'high' and then have students restate it back to me. I look at their eyes to see if they are engaged in the conversation. The words on the board, I use a lot all week long. (Teacher interview)
>
> Interviewer: What do you attribute your success to?
> Teacher: I have a minor in Spanish, so I use this to help my students understand the vocabulary and the concepts I am teaching. I translate when necessary and I draw a lot of pictures. I do not just teach the words OC has listed as vocabulary, but I teach what I see they need. I do not always have time to draw pictures so I may have to quickly clarify or point to an object.
> Interviewer Paraphrase: She also credited her SIOP training.

These excerpts illustrate the teachers' belief in the importance of additional vocabulary instruction to the success of ELL students. However, enhanced vocabulary instruction was not just for ELL students, but for most of the students in these schools since many of them came from impoverished backgrounds and benefited from an additional focus on word and concept development. Coaches and principals also recognized that the core program was weak in vocabulary, especially for ELL students, and they were well aware of the teachers' efforts to shore up this area of their curriculum and instruction.

This is an important finding for IRF. Hispanic students are not making strong and consistent gains in IRF schools. But in the identified teachers' classrooms Hispanic students did quite well on the ITBS test relative to other IRF classrooms. Not all of this improved performance is due to increased vocabulary instruction, but when listening to the teachers and coaches talk, they believe it to be a very important component of their success with all students, especially their Hispanic students. IRF should explore in greater detail the revised vocabulary instruction that is occurring in the high-performing teachers' classrooms. The words being taught need to be cataloged, and how the words are being taught needs to be captured so other teachers can learn about effective vocabulary instruction in IRF classrooms. An important addition to the idea base

mentioned above could be the inclusion of how teachers are augmenting vocabulary instruction with specific examples of how to accomplish it.

An additional focus on writing. Teachers talked about the blue band in OCR. Some liked it and some didn't. Almost all of them used it, some quite a lot, others less so. In some schools, the most needy learners were pulled out to supplemental services programs such as ELL or special education during blue band instruction and never received this component of the OCR program. But one thing was common across virtually all of these teachers. They did more writing in their classrooms than what the blue band stipulated. Of course, the writing took different forms in different classrooms, but story writing was quite common. These teachers believed in the importance of writing on two levels. First, it was important to learn how to write well. Second, learning to write also helps children learn to read. Some used programs such as Step up to Writing, but others didn't follow a prescribed program. The following excerpt illustrates how teachers talked about their expanded writing programs: "We're doing more writing than what's in the grant. We don't have to do what they say for writing. We turn in a writing score that we give each week. We use Step up to Writing."

Openness and willingness to learn new things and to solve problems. All of these teachers appreciated the professional development they had received as a consequence of being in an IRF school. This makes sense since they all shared a hunger for professional knowledge about their craft. They talked about always looking for new things to try when previously applied techniques, strategies, curricula, etc. fail to work as expected. They were not the type of teacher who wanted some new, quick and unsubstantial thing to do on Monday morning with their class, but instead they looked for substantial curricular and instructional ideas that addressed an identified problem or challenge to make them better educators. One teacher talked about how she enjoys searching Reading First sites on the Internet to learn about what is going on in the schools and to look for promising ideas. Another talked about searching the Internet for specific teaching ideas when she feels the need to look for a different way to approach a problem. In the following excerpts, teachers, their coaches and their principals talk about the drive to learn more and get better all the time:

> I always think I can improve in my instruction of Open Court, I am never complacent in that this is good enough. (Teacher interview)

> I'm constantly looking for research. I suppose the other thing that has put me in a different ball park, for the last 2 years, I've been involved in grant writing. So, again, a lot of research involvement. And I also like the workshops – I attended all of the … [core program] workshops that I could. I'm willing to go to whatever is going to help me in improving how I teach Reading First. (Teacher interview)

> She's determined and conscientious. She finds a lot of things to add to the program. She's a lifetime learner who embraces a lot of different ideas. (Principal comment)

The excerpts exemplify the curiosity and devotion these teachers have for their profession. In a similar vein the following excerpt from an interview with a coach also reveals this attribute of teacher curiosity but in a different and interesting way. What is of additional interest in the excerpt is the teacher's desire to know why a specific procedure in Open Court needed to be followed and her openness to exploring for answers coupled with her willingness to change in the face of evidence that ran counter to her practice:

> As a supporter and a person that she can bounce ideas off. She uses me more as a colleague. I think there's a way you can force people. I always asked why would Open Court recommend this and we would go back and read it. For instance, at first in 1st grade, they had an argument about reading the decodables. Sally (pseudonym for a high-performing teacher) was reading the decodables to her children. Open Court said that decodables needed to be read by the students. So they questioned Open Court and researched it and found out why Open Court said that decodables should be read by the students. So they figured it out together. Understanding of WHY we would do this – for both beginning and experienced teachers.

The spirit of critical inquiry into practice that these teachers manifest is laudable. IRF has been quite successful in a number of schools and classrooms at stimulating teachers to think about their practice. As a previous excerpt stated, Reading First is very good at building a foundation of knowledge about practice in teachers. It is icing on the cake when the teachers critically evaluate this practice so they understand better why they do what they do. They thus reenter their classrooms empowered with the knowledge that their practices are best for their students while also adhering to the tenets of the core program.

This spirit of critical inquiry and curiosity prefaces another important point. Many of the coaches interviewed were astute, knowledgeable educators who provided important insights into best practices in IRF classrooms. In the following excerpt, a coach talks about different types of teachers in IRF schools. She characterizes some as problem solvers and others as non-problem solvers. The identified teachers were problem solvers and that is what stimulated the coach's comment:

> The problem solvers work harder at figuring out the program and how to make it work. Non problem solvers get bogged down in the little things. In Open Court, the believers in efficacy and the rationale of the program do well and excel. Those who don't believe drag their feet and are hard to cultivate. This goes back to the spectrum comment earlier. For teachers to have success, their feelings [about Open Court] make a huge difference. Some teachers do what is asked and no more. Kathy (pseudonym) understands and likes the program.

Kathy in the above excerpt was the identified high-performing teacher. She was a problem solver who was willing to work hard to learn the program and make the changes necessary to make the program work at its full potential. The non-problem solvers, who "get bogged down in the little things," are the teachers who find a weakness in the core

program and instead of digging in and working hard to make the necessary changes, fall into complaining and negativity that undermines the collective work of the IRF teachers at that grade level or perhaps in the entire school. The coach talks about the difficulty of moving these non-problem solvers forward. This quote points out a salient difference between the problem solvers and the non. The problem solvers have an unwavering focus on students and their needs. The non-problem solvers get diverted from this focus and instead dwell on problems in the program, or problems with the students such as home life, etc. Perhaps IRF could use some of the problem solver teachers to work with the others to show them how they can practice fidelity to the core, remain positive about doing so, and maintain a positive outlook on students and their learning.

Strong, well-coordinated supplemental services and collaboration. Superior quality supplemental services that were coordinated well among teachers, coach and supplemental services teachers was a common, but not universal, attribute of these teachers. Common but not universal is an important phrase. Some teachers praised their special education teachers and program, their ELL teachers and program, and their Title I teachers and program as integral components of the overall reading program in their school and also as very important components of their individual success as teachers. In these schools — and most schools were like this — regular classroom teachers, supplemental services teachers and the coach regularly communicated with each other to establish intervention groups, discuss individual children's needs and coordinate delivery of the various curricular programs being used with the children. Where the teachers praised the supplemental services, they often spoke of the importance of one-on-one or small group instruction and that the supplemental services are where the children can get such attention on a regular basis. In the following excerpts from teacher interviews the importance of strong supplemental services and collaboration among all teachers are emphasized:

> The great thing is the working together … that's the most important part of Reading First… Title 1, resource, and challenge. We sit down together and decide what's best for kids. Now we work together on everything.

> Interviewer: What elements of your overall program are important to your success? Teacher: The collaboration and time to seek answers from other people. Help from others: Title 1, resource, challenge, ELL, and the teacher across the hall.

> It's the whole thing working together. Everybody tries to keep in contact with each other, everyone makes an effort. The more you teach something the more you get used to it and the more you know what you have to do to be successful. We're good at sharing things.

These comments are heartening to read since a goal of IRF has been to build cultures of collaborative problem solving in the schools. There is evidence that this has occurred in a number of places and the teachers enjoy it and benefit from it.

But not all schools were highly collaborative environments where identified teachers were part of a team and supplemental programs were important to their success. Several teachers praised their supplemental services teachers but did not regularly meet with them or for that matter have much knowledge about what the supplemental services teachers did with the students when they left the regular teacher's classroom. There were also two or three teachers who will be called "lone rangers." They had almost all of their children for both workshop and intervention and didn't interface with supplemental services teachers, didn't know what the supplemental services teachers did in their classrooms or didn't appear to care all that much, and felt that the supplemental services in their buildings had little influence or perhaps no influence on their success. Some of these same teachers didn't collaborate much with their grade level colleagues either. But coaches and principals provided evidence that the supplemental programs in their buildings functioned well and were effective. So the teachers' not working closely with their supplemental programs is not due to poor programs.

Collaboration among grade level colleagues was another variable in these teachers' success. Some attributed much of their success to such collaboration, whereas others were lone rangers. In the following excerpts both sides of this continuum are represented. The first two excerpts, the first from a teacher interview and the second from a coach interview, reveal large amounts of collaboration and the second two excerpts little collaboration:

> Interviewer: How much of your success is attributable to this collaboration?
> Teacher: Bunches!
>
> Interviewer (Speaking to a coach): How much of her success is attributable to collaboration?
> Coach: The last couple of years quite a bit of collaboration contributes to her success. Because she was so opposed to Open Court, the collaboration helped her a lot.
>
> Interviewer: How much of your success is attributable to this collaboration?
> Teacher: The data and we talk about different ways of doing things. Maybe a third or 25%. Really, you do what you do in your own classroom. We meet, plan, but do our own thing.

This range of attitudes towards collaboration can be seen as a cup half full scenario instead of a cup half empty. What it reveals is that teachers can be successful in IRF schools without a lot of collaboration. For those who don't collaborate much, it may make the journey more difficult but the result can be quite similar to those who do collaborate. This comment is in no way advocating for reducing the pressure in IRF schools to build collaborative cultures. It is simply made to make the point that there are other, albeit perhaps not as ideal, means to the same end.

Special mention needs to be made concerning ELL students with regards to coordination of services and collaboration. In classrooms where teachers are having strong relative

success with ELL students, it's usually not just the teacher by herself who is achieving the success, but a combination of the teacher and the ELL infrastructure that has been built up in the school. These schools have strong ELL teachers or aides and many of them use Language for Learning or other programs focused on ELL students, which by the way was given strong praise by teachers, coaches principals. The ELL students are given a lot of small group and individual attention throughout the day via workshop and intervention times where the classroom teacher and ELL teacher or aides work diligently to help these students achieve. Important attributes of these programs include high expectations for students and quality, coordinated programming delivered by motivated educators. It appears that the infrastructure of small group intervention and appropriate curricular materials such as Language for Learning that has evolved in IRF schools synergizes with the high-performing teachers' style of teaching, devotion to all students and work ethic to produce some excellent learning environments for ELL students.

But just like teachers' use of supplemental services and their levels of collaboration with their grade level colleagues, the degree of ELL student success that depends on the ELL infrastructure built up in a school varied markedly across the teachers. Some teachers praised their ELL programs and said that they wouldn't be as successful without them. Others didn't feel their ELL programs were that important to them, while others didn't even have access to an ELL program, or at least they said they didn't, and attributed their success to what they did in their classrooms.

Connect/make learning relevant to students' lives. All of the teachers seemed intent on letting the students know that the content could be applied to their lives. With vocabulary, teachers asked students to make connections, had students put definitions into their own word and asked for examples from their lives. Some teachers were explicit as to why fluency was important – beyond the score. They talked about fluent readers being readers who could understand, think about and remember what they read. One notable example was a charismatic veteran teacher who continually discussed with her students why literacy was important and why they were learning the things that they were. In the following interview excerpt she discusses this:

> Lots of things like – you're going to grow up and be married and your kids will want to go to Roaring Springs and you might not have enough money. So if they don't 'climb the mountain,' they won't get a good job. I also point out that parents are paying the teacher and they want me to do a good job.

One could quibble about this teacher's focus on learning as strictly a means to a material or vocational end, but these are important pursuits nonetheless. The excerpt does clearly illustrate the immediate relevance this teacher strove for and her constant focus on keeping children engaged and motivating children to achieve.

There were other teachers who talked with their students at considerable depth about the rationale behind learning something that transcended the material or vocational. For example, observers noted that students could tell them why they were working on Read Naturally. The children could talk about fluency, comprehension and higher order

thinking. Sometimes the relevance was to test taking. "When you come to a word that you don't know on the ISAT, chunking the word will help read it." Relevance was an important motivational tool for these teachers and it took many forms.

Dialogue and higher order questioning. Classroom observers noted a number of instances of extended conversation and higher order questioning in these classrooms. The observers who also observed last year were especially aware of this. One noted that she witnessed the most extended discussion and higher order thinking that she had seen in IRF classrooms.

This focus on extended conversation and higher order thinking occurred in both the green band and red band portions of the daily lessons. For example, one teacher modeled mistakes in decoding for the children and had them tell her what she did wrong and why it was wrong. During this there was lively dialogue and debate about how words are put together and what can and cannot be done with different word parts. During story discussions, conversations were deep and questions were asked that made students think about connections between the story and themselves, their worlds outside of school and other stories that had been read. Teachers asked questions that weren't in the teacher's manual.

The use of extended dialogue in the classrooms was corroborated by principals and coaches. A coach remarked when asked what set the teacher apart from her colleagues, "Also engaging the students in a lot more conversation instead of just following the book would probably be a difference." This quote is not included to undermine fidelity but to show another example of how these teachers used the core program as a solid foundation upon which to further improve their instruction and thus student learning.

Coaches' and principals' roles. The role of coaches and principals in these teachers' professional lives was as varied as the teachers themselves. Ideally, the teachers would be close to both their coaches and principals and would draw heavily upon their support for much of their success, but such was not always the case. Some were close to their coaches but not their principals. Others were close to their principals but not their coaches. A few were close to both and a few weren't close to either. Thus the role of leadership in the success of these teachers is not clear. It appears that teachers can be successful even when their perceptions are that their coach and principal do not contribute much to their success. There were a few instances where teachers said the coach was not important to their success and that the coach seldom visited their room, but the coach said that she played an important role in all the IRF classrooms in the school. These diametrically opposed perspectives are in keeping with results from the two previous evaluations.

Closing: The art versus the science of teaching. The art and science of teaching were mentioned at the beginning of this section on findings from the classroom observations and interviews. It will be discussed more thoroughly here since it is a good way to encapsulate the findings.

The identified teachers had mastered both the art and science of teaching. They had strong, positive relationships with their students and they knew the technical strengths and weaknesses of their curriculum and instruction inside and out. Additionally, on the art side, they were creative and dynamic so they could quickly and efficiently adapt and adjust their demeanors, their instruction and their curriculum to meet each individual student's needs. The following excerpt from a teacher interview points out this interplay:

> The step-by-step program, the routine. The hand signals. The phonics are strong. It is reinforced in intervention and workshop. It took some years to get used to it. You have to have your own personality come through so you're not a robot – but you have to become robotic to do it right. It was necessary and I guess it's important then to move into it. If you laboriously go through the steps, you can internalize it first. For young people, if they don't know that they're supposed to add life. Plus the pressure that people are under, I think they get nervous and anal. It always concerns me that young teachers that are trained in this don't realize that you have to add your personality. One young teacher would not look at her audience because she was trying so hard to do what she was supposed to do. They are so intent on following the rules. I see it in older teachers too – they lose touch with the personality of teachers.

What this excerpt points out quite poignantly is that these teachers have rich classroom personalities that interface with the core program and supplemental programs in energizing ways. But this doesn't occur spontaneously or immediately. Teachers have to have fidelity to the program for some time before they become knowledgeable enough about it to be able to put their personality into it. It is a trajectory that these high performing teachers appear to have followed. But importantly, not all teachers complete the trajectory. We have observed classrooms in past years where teachers remained robotic and the instruction lacked energy and personality. Thus it remains a challenge for IRF to try to find the combinations of professional development for teachers and training for coaches and principals so this trajectory is traversed by all teachers and energizing classroom environments becomes a universal attribute of IRF schools. A coach put it quite succinctly when she talked about the strengths and weaknesses of IRF. "So much change so fast. Teachers thought they had to throw that teacher instinct out the window, but that was never the intent." Selective re-introduction of teacher instinct is perhaps in order. These 29 teachers provide a road map for what it looks like. A teacher captured its essential features in the following statement: "I look for where the holes are and where the needs are and I try to fill them in." Identifying holes and filling them in are quite different from moving away from fidelity. What it does instead is allow a teacher to stay close to the core program and implement it with fidelity while also working on making both their curriculum and instruction better. As one teacher said, "There isn't anything wrong with Open Court materials." The best teachers in IRF schools provide quite clear road maps for the interplay of art and science—the interplay of instinct and technical accomplishment.

But this journey is not easy. It hasn't been in the past and won't be in the future. As one of the most dynamic and engaging teachers said, "The first year and a half, I was

miserable learning all those things. Now I see the kids as much stronger readers and I do a lot better job scaffolding which helps them. We're just inundated with new ideas from the principal and reading coach and from classes because of Reading First." This statement needs to be underscored. It reveals that one of the very best IRF teachers says she was miserable for the first year and a half and then she saw results, got to know the program better and things started to look up. This trajectory should become a goal for all IRF teachers, and coaches and principals should continue to strategically work with their teachers to help them travel this path. It is suggested that coaches and principals employ their best teachers to be integral parts of this process. This cadre of 29 identified teachers has hundreds of years of experience, excellent ideas for making the core program more effective while teaching it with fidelity, and binders full of teaching materials that help them do so.

But even keeping the very best teachers moving forward on their individual developmental paths poses challenges. As was discussed above, all of the teachers feel stress about the rigidity in IRF and want more flexibility. Some are at critical stages in their stress. The following quote exemplifies this. The teacher is talking about how Reading First, most notably the core program, is being implemented in her school:

> It's so regimented that some things that worked for teachers for years, they have to throw out. For example, having children read to adults is really important, but Reading First doesn't allow teachers to share successes outside of Open Court. Teachers are afraid to let people know that they are doing something differently.

Whether this statement is true or not, and it is hard to believe that it is true, it graphically represents the sense of constraint that some of these teachers feel. For a teacher to feel so constrained is problematic, and IRF needs to begin to address these feelings.

In closing, based on the findings there is ample capacity in the IRF network to take on most any challenge. The 29 teachers, their coaches and principals were all fine educators willing to work hard to assure the success of all students. The amount of knowledge that has accumulated about the core programs and how to address individual needs is quite astonishing. Of course, IRF schools don't have all the answers and solutions, but a lot has been learned that can now be disseminated across the network. The following quote from a teacher encapsulates the level and quality of criticality and insight that is found in these schools:

> The weakness is that Open Court does not provide adequate challenge to the highest students. What it calls challenge work is really just" busy work." It is also weak with lowest students. It doesn't provide enough re-teaching. I would like to see more phonemic awareness activities and a variety of activities like some games. This is especially important for the intensive group as they often need more motivation.

The additional phonemic awareness activities and the games have been developed. We saw them during our classroom observations and we heard about them during our interviews. Undoubtedly more things will need to be developed and some that have

already been will need revision, but IRF has a strong start on building a powerful instructional and curricular infrastructure to address student needs. What needs to be done now is disseminate this knowledge so others can gain access to it.

Many of the teachers had similar criticisms and suggestions as those articulated in the previous excerpt, and many more had different criticisms and suggestions. In aggregate there is a wealth of constructive criticism and recommendations out there to systematically act upon to ascertain which criticisms are valid and need to be addressed and which recommendations lead to improved student achievement. One recommendation that was made by most of the teachers needs to be started immediately, however, and that is for teachers to watch teachers and share ideas and materials. There remains too much teacher isolation across the IRF network. Teachers are working with one another within their buildings, but little cross network sharing and collaboration is occurring, according to the people interviewed this year.

Finally, watching these 29 teachers reveals that teaching is a complex business and that no single program or teacher can bring all children to criterion. As one coach said, "Open court is not a stand alone program. You need the supplementals to accommodate teacher and student needs." The ongoing challenge for IRF is to find ways to coordinate all of the programs and approaches so that student achievement is maximized.

Principal Turnover Data

A series of school background variables were measured that were hypothesized to influence reading achievement. The turnover rate of school principals was one of those variables and will be the first to be discussed.

During the 2005-2006 evaluation, coaches were asked to fill out a table listing who their building principals and vice principals (where applicable) were during the span of academic years from 2001-2002 to 2006-2007. Data was collected for the two years prior to the start of IRF to establish a baseline. Twenty-eight of 30 IRF schools responded after multiple requests for this data. The two schools that did not respond were Filer and New Plymouth. All those who did respond provided complete data. For the 2006-2007 evaluation this data was updated to reflect the additional year of IRF. All 30 schools responded this year. Table S3-4 shows which schools experienced changes disaggregated by several criteria. If a school is not listed, that means they experienced no change in principals during 2001-2007. The table does not report data for vice principals. The first column after the school's cohort identifies the number of principal changes in the building prior to the beginning of IRF. This provides a measure of the leadership stability in the building leading up to the start of IRF. The column headed "Change During Transition Year to IRF" identifies those schools that experienced a change in principal between the year just prior to starting IRF and then the first year of IRF. This particular transition was identified because IRF can represent quite a change in a building. The leader who worked with the staff preparing the grant application would most likely be the best person to lead the school into the first year of IRF. The final

column identifies schools that have had principal changes during the years they have been in IRF.

Table S3-4: Principal Changes 2001-2007

School	District	Cohort	Change Before IRF Years*	Change During Transition Year to IRF	Change During IRF Years**
Sacajawea	Caldwell	1			3 in 4 years
Adams	Madison	1			1 in 4 years
Roberts	Madison	1		1	1 in 4 years
Archer	Madison	1			2 in 4 years
Union Lyman	Madison	1			2 in 4 years
Acequia	Minidoka	1			1 in 4 years
Heyburn	Minidoka	2	1 in 3 years	1	
East	Mountain Home	1			1 in 4 years
West	Mountain Home	1		1	1 in 4 years
Sherman	Nampa	1		1	2 in 4 years
Snake River	Nampa	1	1 in 2 years		1 in 4 years
Priest River	Priest River	2		1	
Bickel	Twin Falls	1	1 in 2 years		
Oregon Trail	Twin Falls	1			1 in 4 years

* Since there have been two cohorts of schools, some schools have two years of data (i.e., 2001-2002 and 2002-2003) prior to their start of IRF and others have three years (i.e., 2001-2002, 2002-2003, 2003-2004).

**Since there have been two cohorts of schools, some schools have been in IRF three years (i.e., 2004-2005, 2005-2006, 2006-2007) while others have been in IRF for four years (i.e., 2003-2004, 2004-2005, 2005-2006, 2006-2007).

Eleven IRF schools have experienced a principal change during the years they have been in IRF. Three schools, Archer, Union-Lyman and Sherman, have experienced two changes in principals during their four years in IRF. Sacajawea has experienced three principal changes during their four years in the network. Sherman and Sacajawea are two of the lowest-performing schools in the initiative, but on the other hand Archer and Union-Lyman, two rural schools that share a principal, have done well. Thus there is no clear evidence that multiple principal changes strongly correlate with poor school performance. But finding and retaining strong leadership should be a focus in IRF schools since research has shown that leadership is important to school improvement.

Teacher Turnover Data

Idaho Reading First requires a high degree of coordination between teachers within and across grade levels in a school, and the core reading programs are sophisticated curricula that require substantial teacher professional development and practice before they can be mastered. Additionally, prior evaluations revealed that some teachers initially struggle with developing effective workshop/intervention periods but improve with experience. Thus it is important in IRF schools to retain high-quality, experienced teachers. It follows then that high turnover in the teaching staff within a building could negatively impact the success of the initiative. During the 2005-2006 evaluation, reading coaches completed a table asking for the total number of teachers at each grade level and the number of new teachers at each grade level. These data were broken out by year starting with the 2001-2002 academic year and ending with the 2005-2006 academic year. This year, coaches were asked to provide the same information for 2006-2007. Idaho Reading First started at the beginning of the 2003-2004 academic year. For this group, two years of teacher turnover data prior to the beginning of the initiative were collected as a baseline so that the possible impact of IRF on turnover could be measured. A second cohort of 10 schools started IRF at the beginning of the 2004-2005 academic year. Thus, this group has three years of baseline data prior to starting IRF.

Teacher turnover was defined in the following way. If a teacher was new to the building at the beginning of the academic year or had taught in the building in prior years but had been away for a year or more, IRF coaches were instructed to count that teacher as a new teacher. If a teacher left during the year for maternity leave, illness, retirement or employment elsewhere and that teacher was replaced with another teacher or a long term substitute for the remainder of the year, IRF coaches were instructed to count the replacement teacher as a new teacher.

After a number of requests and reminders, all IRF schools returned their teacher turnover data. Therefore, the response rate was 100%. Although all schools reported their data, there was some missing data within schools, which caused group sizes in the final calculations to be smaller than the actual number of schools in the particular cohort. Teacher turnover data were analyzed several different ways. Following is a discussion of those analyses.

Table S3-5 reports teacher turnover by school and grade. Some explanation for how to interpret the table is needed. Under each grade level there are columns labeled "Years" and "Proportions." The "Years" columns report the number of years the school has experienced a change of teachers at that grade level out of the number of years the school has been in IRF. For example Popplewell's kindergarten has experienced a change in one or more teachers each of the three years they have been in the initiative (i.e., 3/3). The "Years" column does not report how many teachers changed. It just reports how many years a change occurred at the particular grade level. The "Proportions" columns report the percent of teacher turnover for each year turnover occurred. For example, Popplewell's kindergarten experienced 33% turnover two years and 50% turnover one year. Please note that the yearly percentages are not in chronological order but instead are in ascending order from lowest to highest. There is no trend in the percentages across the schools showing increasing or decreasing rates of teacher turnover as schools remain longer in IRF, so listing the percentages in ascending order allows readers to estimate a range and average more easily.

Table S3-5: Teacher Turnover by School and Grade

School	District	Kindergarten Years*	Kindergarten Percents**	First Grade Years	First Grade Percents	Second Grade Years	Second Grade Percents	Third Grade Years	Third Grade Percent
Popplewell	Buhl	3/3	33/33/50	1/3	40	2/3	60/60	3/3	25/25/25
Lewis and Clark	Caldwell	2/3	33/33	2/3	50/50	2/3	25/67	2/3	33/67
Sacajawea	Caldwell	1/4	50	2/4	25/80	2/4	25/50	3/4	33/67/75
Wilson	Caldwell	2/4	25/25	2/4	25/43	1/4	14	1/4	33
Butte View	Emmett	0/3	0	0/3	0	1/3	17	1/3	33
Filer	Filer	0/3	0	1/3	25	0/3	0	1/3	25
Gooding	Gooding	4/4	33/33/60/67	2/4	20/25	1/4	20	1/4	50
Homedale	Homedale	2/3	33/67	1/3	25	1/3	20	3/3	40/50/50
Harwood	Jefferson	2/4	33/33	1/4	20	2/4	20/25	1/4	25
Roberts	Jefferson	1/4	100	2/4	50/50	1/4	50	3/4	50/50/50
Adams	Madison	3/4	50/50/50	3/4	50/67/100	0/4	0	2/4	50/50
Archer	Madison	1/4	100	1/4	50	1/4	50	3/4	50/50/50
Union-Lyman	Madison	2/4	33/50	1/4	25	3/4	25/25/100	0/4	0
Acequia	Minidoka	0/4	0	2/4	50/50	2/4	33/50	3/4	50/50/50
Heyburn	Minidoka	0/3	0	3/3	33/33/50	2/3	33/50	2/3	67/100
Paul	Minidoka	1/4	50	1/4	25	1/4	25	3/4	25/33/67
East	Mountain Home	4/4	33/50/50/50	2/4	33/75	1/4	33	1/4	33
West	Mountain Home	2/4	50/100	2/4	25/25	1/4	50	3/4	25/33/75
Sherman	Nampa	4/4	20/20/20/20	2/4	20/50	3/4	25/25/50	3/4	25/50/75
Snake River	Nampa	2/4	33/33	1/4	50	2/4	25/75	2/4	67/100
New Plymouth	New Plymouth	2/4	50/50	1/3	33	3/4	33/33/67	4/4	33/33/33/100
Bickel	Twin Falls	2/4	33/50	3/4	33/33/100	3/4	33/33/33	2/4	33/67
Harrison	Twin Falls	2/3	50/67	0/3	0	1/3	50	1/3	25
Lincoln	Twin Falls	3/3	33/33/50	3/3	25/33/33	2/3	33/67	2/3	33/33
Oregon Trail	Twin Falls	2/4	50/50	4/4	25/25/25/7	1/4	50	2/4	25/25

					5				
Central Canyon	Vallivue	3/3	33/33/33	2/3	20/40	3/3	20/40/67	3/3	25/50/80
East Canyon	Vallivue	3/4	14/25/67	3/4	20/33/43	4/4	25/33/40/43	3/4	17/20/25
West Canyon	Vallivue	3/4	50/50/50	4/4	25/40/50/75	3/4	25/25/50	3/4	50/67/75
Wendell	Wendell	2/4	33/50	4/4	20/20/20/40	2/4	25/25	0/4	0
Priest River	West Bonner	0/3	0	3/3	25/33/33	2/3	33/67	1/3	33

* Number of years teacher turnover occurred / Number of years in IRF

** Proportions of teacher turnover. Not in the order in which the turnover proportions occurred but in ascending order.

There is substantial teacher turnover in all but a few IRF schools. This is not said to point blame at IRF. There is no evidence that IRF is causing the teacher turnover. It could just as easily be due to demographic or economic forces or some other causative agent. But as was discussed above, IRF is a sophisticated, demanding program that most teachers have to experience for a year or more before they can fully implement it. Thus when teachers leave, large holes may be left behind. This challenge is further exacerbated in small IRF schools where there are only one or two teachers per grade level. When a teacher leaves in one of these schools, a large percentage of institutional knowledge leaves with the person. Replacing this institutional knowledge is resource and time intensive, since the new teacher who enters will probably not have the level of expertise and experience with IRF that the previous teacher had accumulated. This holds true even in larger schools, and consequently children may be placed at some disadvantage because of teacher turnover whether it be in a large or small school.

Looking within Table S3-5, Butte View and Filer stand out for their low turnover across all grades. Similarly, Harrison has low turnover in grades 1-3, but their kindergarten has experienced higher rates, having had substantial changes in two of the past three years. Several schools stand out for their consistently high turnover rates at all grade levels. The schools are Sherman, Lincoln, Central Canyon, East Canyon and West Canyon. There are other schools in the table that have experienced high turnover at one or more grade levels and quite low in other grade levels. Gooding, Homedale and Adams are examples of this pattern.

Overall averages were computed for the cohorts to reduce the complexity of Table S3-5. The averages are reported in Table S3-6. The table reports the average teacher turnover across all schools and all years by grade level. It also separates the years into those before a school became an IRF school and those after entry into the initiative.

Table S3-6: Average Teacher Turnover by Grade—Before and During IRF

	Grade K	Grade 1	Grade 2	Grade 3
Prior to IRF*	.24 (.26)	.16 (.13)	.18 (.18)	.21 (.19)
During IRF*	.24 (.15)	.22 (.16)	.21 (.15)	.28 (.18)

* Proportion (Standard Deviation)

An example will help interpret this table. Look at the intersection of the column "Grade K" and row "Prior to IRF." The number in the cell means that the average turnover in kindergarten for all schools in the years prior to their beginning IRF was 24%. The standard deviation of the percentages that make up this average is in parentheses. The averages, however, need to be interpreted with caution. A measure of central tendency for data like this can be misleading. An example will illustrate this assertion. Grade 3 has experienced 28% teacher turnover on average. This average represents schools that have had no turnover at the third grade (e.g., Union-Lyman and Wendell) and schools

that have experienced 100% turnover during a given year (e.g., Heyburn, Snake River, New Plymouth) plus schools that have turnover amounts between these two extremes. The average of 28% doesn't appear to be that much, but when a school loses 100% of its teachers at a grade level in a year that probably has serious consequences for the school. So readers are cautioned to interpret Table S3-6 carefully.

Except for kindergarten, turnover rates appear to have gone up during IRF. This is especially the case at 2nd and 3rd grades where the change has been 6% and 7% respectively. It was stated above but will be emphasized here again that there is no evidence that IRF is causing these increases. Other factors could be the causative variables. But when the data in Tables S3-5 and S3-6 are taken together evidence accrues that substantial and continuing teacher turnover is an issue in IRF schools. It will be shown later in this report that IRF schools are not making strong, consistent gains on test scores. Given the sophistication of the IRF program and the level of teacher expertise and dedication the program requires, the consistent teacher turnover in many of the schools probably exacerbates the seemingly intractable test score problem since experienced staffs who have worked together for all of the IRF years are the exception instead of the norm.

Now, the last line brings up an important caveat. The teacher turnover statistics discussed above should not be interpreted that there is wholesale abandoning of IRF schools. Granted, where there has been 100% turnover at a grade level in a given year there are no teachers left at that grade level with IRF experience in the particular school, but this seldom occurs. Most of the time some teachers come and go, leaving a core of "old timers" to carry current programming forward. This is the case in IRF schools also, but the problem is that there will be classes full of children come the fall in these schools that "get the new teacher." And even though new teachers may be excellent, they may not be as effective as they could be if they had had more experience with IRF programming in that particular building. In short, teacher turnover is not an optimal situation to have in schools where highly experienced, dedicated team players are absolutely essential. Thus, districts where IRF schools are located might consider providing incentives for the best and most experienced teachers to have long and fruitful teaching careers in their IRF schools.

Special Education Referrals

The trend in the number of students referred to special education in IRF schools was explored both last year and this year. It was hypothesized that referrals would drop the longer a school was in IRF as a consequence of more effective reading instruction. If reading instruction becomes more effective, then the number of students who fail to acquire early literacy skills goes down and along with it the number of students referred to special education for evaluation. To explore this hypothesis, last year Reading First coaches were asked to collaborate with their special education teachers and gather the data to report the number of special education referrals by grade level during the years leading up to and then during Idaho Reading First. Coaches were asked to collect data for 2001 through 2006.

51

This data was hard to collect last year. Even after multiple requests during the spring of 2006, only 22 of 30 schools submitted data. Furthermore, some schools did not have complete data for all years and grade levels. Coaches were contacted and asked why the data was so difficult to obtain. They reported a variety of reasons. In a few instances they said their school does not keep track of referrals but only those students who are placed in special education. Others said that their special education teacher was no longer employed at the school and had not left behind information on where these records were kept.

It was thus decided for the 2006-2007 evaluation that the request for the data would be sent to the special education directors in the central administration of each IRF district. It was believed that these individuals would have more ready access to this particular data and the response rate would thus be better. Such, however, was not the case. Only 10 schools submitted data this year. Within a month of the initial request for the data in March it was obvious that the response rate was going to be low since few had responded. To try to increase response rates, additional requests were sent to the special education directors throughout the remaining months of the school year and coaches were also contacted to once again work with their building special education teachers to either provide all of the data if they had not done so last year or update their data to include the 2006-2007 school year. In the end, the response rate was very poor.

Given the worse response rate from the previous year, the lead evaluator contacted reading coaches, building level special education teachers, and central office special education administrators to discuss why this data was so difficult to obtain and what needed to be changed in the request for data so that schools could more readily comply. First, a number of IRF schools don't collect referral data. They only keep records on the students who are formally admitted into special education. Thus, it is likely that some of the data reported in this section of the report represents students who were formally placed in special education, not just referred to special education. Second, some schools do keep this data but it is not available because teachers have left and/or special education directors have left and no one knows where the records are located. Third, some IRF schools have moved to a Response to Intervention (RTI) model and this clouds the picture on how many students are referred to special education. Under RTI numerous students can be monitored and even provided services before a full referral to special education is made. Finally, some schools do keep statistics on referrals and these numbers were submitted for this evaluation. Interestingly, those people contacted who did not currently collect referral statistics all said that they wished they did since it would be an excellent indicator of the substantial changes going on in their buildings that have caused referral rates to drop.

In summary, referral data is for the most part not available from IRF schools. And what has been submitted in the past and will be reported herein is probably a mixture of referrals and students admitted to special education. In other words, those schools that keep track of referrals submitted their data as requested, but those schools that don't keep track of referrals probably submitted the number of students admitted to special education

each year. In future evaluations, it is recommended that instead of referral rates the number of students who are placed in special education be collected by grade level and year. Virtually all IRF schools keep these records so response rates should be high. For now, all special education referral rate data reported herein should be cautiously interpreted. More importantly the 2006-2007 data should be mostly ignored since only 10 schools reported.

But the hypothesis is still of substantial importance to IRF. If special education placements decrease as a direct consequence of IRF programming, then this represents an important outcome. It shows that classroom instruction has improved and is providing better curriculum and instruction for students who have historically been at greatest risk for reading failure. Additionally, the cost savings associated with lower special education populations is also an important benefit. Thus, it is important in future evaluations to continue to explore this variable. But future work should not only focus on special education placements instead of referrals but also whether a school is using an RTI model, since RTI represents nothing short of a revolution in special education programming schools. It can thus significantly impact special education placements in both directions. That is, the number of placements can go up or down under an RTI model independent of other interventions such as IRF. In short, the collection instrument used to gather this data needs to be quite sophisticated if it is to tease out whether or not IRF has been a causative agent in any change in special education placements.

Although the quantitative evidence testing the special education referral hypothesis should be interpreted with caution, additional qualitative evidence in support of the hypothesis was collected during the classroom observations and interviews both last year and this year. Numerous teachers, coaches and special education teachers said that their school's special education referral rate had dropped since the school began IRF. One special education teacher said, "They are doing in their classrooms what we have always done in ours, so not so many kids need to be referred." A regular classroom teacher said, "I don't refer as many because we are handling the problems in the regular classroom with the special educator's help." A coach said, "Our referrals are way down. We have programs in place outside of special education that meet a lot of the kids' needs." Interestingly, a reading coach expressed concern that there was "potential for under-identification." She had seen rates drop in her school to such a degree that she was concerned that children who would need special education support after the primary grades would not be identified early enough and would thus suffer in the later grades. The discussion now turns to the quantitative data analysis and interpretation.

The number of students at a given grade level can vary from year to year. Thus, just looking at raw counts of how many students at a grade level are referred to special education during an academic year may be misleading. For example, the number of students referred at a given grade level within a school could go up because the number of students within that grade level increased because of enrollment fluctuations. The same could be said for the number referred going down. In other words, the proportion of students being referred could remain the same but the raw count could go up or down depending on enrollment trends. To correct for this potential problem proportions were

computed for each grade level in each school. More specifically, the number of students in each grade level referred to special education in a school during an academic year was divided by the total student population in that school at that grade level during that academic year. School and grade level enrollment statistics were taken from the Idaho Department of Education Web site. Table S3-7 provides means and standard deviations for yearly referral proportions by grade level and cohort. It is again important to disaggregate by start date since where a school is in the IRF cycle may influence referral rates.

A quick example will aid in interpreting the large amount of numerical data in Table S3-7. Go to the row labeled "K" and "Cohort" 2003-2004 and move across to the first column labeled 2001-02. This cell has .039(.03) in it. The .039 is the mean proportion of kindergarteners across all IRF schools in 2001-2002 who were referred to special education. The value in parentheses is the standard deviation of the individual school values that went into the computation of the mean proportion. These proportions can be quickly converted into percentages to make interpretation easier. The .039 becomes 3.9% and is interpreted in the following way. The average percentage of kindergarteners referred to special education in IRF schools during 2001-2003 was 3.9%.

Table S3-7: Means and Standard Deviations of Yearly Special Education Referral Proportions by Grade and Cohort

Grade	Cohort	2001-02[1]	2002-03[2]	2003-04[3]	2004-05[3]	2005-06[3]	2006-07[4]
K	03-04	.039 (.03)	.033 (.04)	.025 (.03)	.039 (.05)	.017 (.02)	.008(.01)
K	04-05	.014 (.02)	.031 (.03)	.037 (.02)	.038 (.04)	.023 (.02)	.046(.02)
1	03-04	.057 (.05)	.056 (.04)	.043 (.04)	.042 (.03)	.036 (.04)	.024(.02)
1	04-05	.029 (.02)	.035 (.04)	.037 (.01)	.034 (.02)	.032 (.03)	.037(.03)
2	03-04	.046 (.04)	.065 (.04)	.052 (.05)	.051 (.04)	.053 (.03)	.043(.02)
2	04-05	.059 (.06)	.036 (.04)	.046 (.04)	.046 (.04)	.036 (.02)	.024(.02)
3	03-04	.058 (.07)	.055 (.06)	.044 (.05)	.048 (.05)	.045 (.03)	.014(.01)
3	04-05	.030 (.03)	.034 (.03)	.029 (.02)	.026 (.02)	.043 (.01)	.017(.01)

2001-02[1] n=12 for the 2003-2004 cohort and n=7 for the 2004-2005 cohort.
2002-03[2] n=13 for the 2003-2004 cohort and n=8 for the 2004-2005 cohort.
2003-04[3], **2004-05[3]** and **2005-06[3]** n=14 for the 2003-2004 cohort and n=8 for the 2004-2005 cohort.
2006-07[4] n=5 for both cohorts
( ) Standard deviations

Making sense of a large amount of numbers like those in Table S3-7 is difficult. Thus a series of bar charts were created, one for each grade level, to graphically illustrate the tabular data. Each grade level chart will be discussed starting with kindergarten below. In all of the charts Cohort 1 refers to the original group of schools starting IRF in 2003-2004, and Cohort 2 refers to the second group of schools that started in 2004-2005. This means that there are two or three years of baseline data in the charts depending on the cohort. Specifically, Cohort 1 schools have 2001-2002 and 2002-2003 data as baseline (i.e., blue and green bars in all of the charts), and Cohort 2 schools have 2001-2002, 2002-2003, and 2003-2004 data as baseline (i.e., blue, green, and tan bars in all of the charts).

Chart S3-1: Proportion of Kindergarten Students Referred to Special Education by Year and Cohort
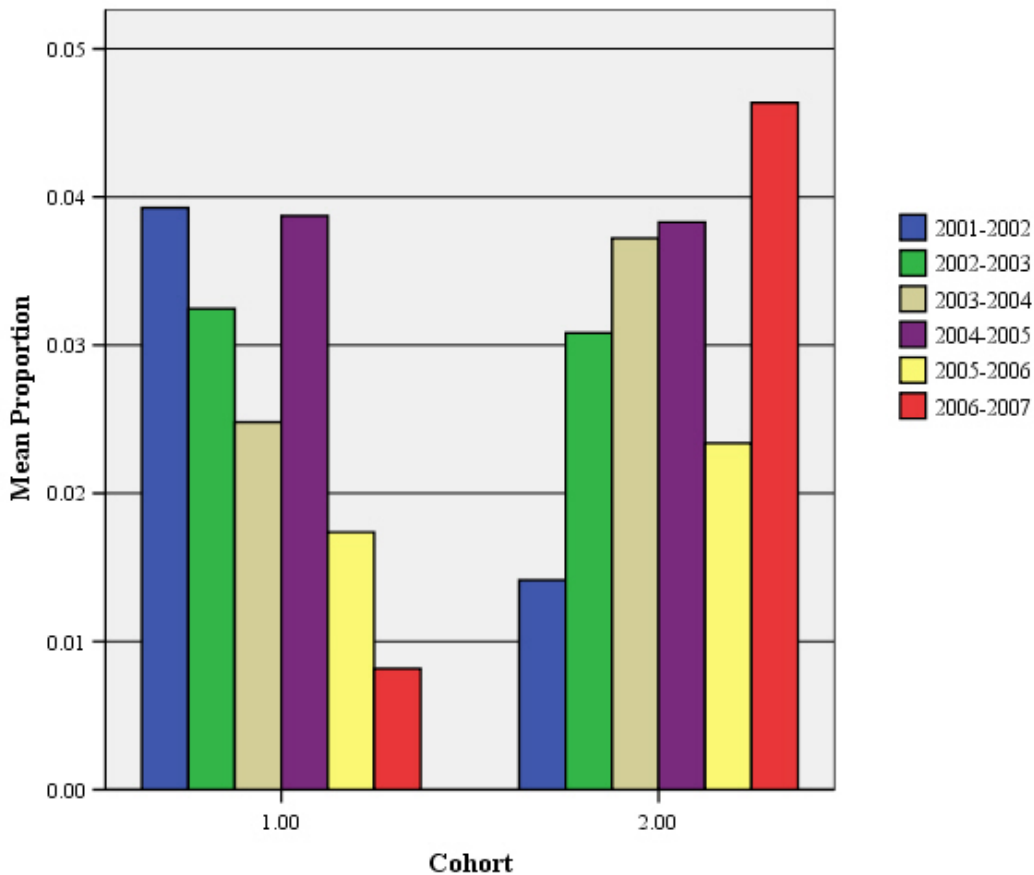
Trends should have no fewer than three data points and five are preferable before a conclusion is drawn. And given that the 2006-2007 data (the red bars in all of the charts) should be interpreted with great caution since it only represents five schools in each cohort, trends cannot be established for either Cohort at this time. However, examining the charts for insights is still possible. In the case of Chart S3-1, cohort 1 may have a trend downward in referrals. Although there was a spike up in 2004-2005, referrals

dropped to an all time low in 2005-2006. At this time, Cohort 2 has no distinguishable trend after the schools started IRF; however, there was an upward trend in referrals prior to and during the first year of IRF. This trend was quite dramatically broken the first year of IRF (i.e., 2005-2006). This could be the beginning of a sustained reduction in referrals but several more years of high quality data are needed to confirm this.

Chart S3-2 shows trends for 1st grade students. It is interpreted the same way as the previous chart for kindergarten students.

Chart S3-2: Proportion of First Grade Students Referred to Special Education by Year and Cohort
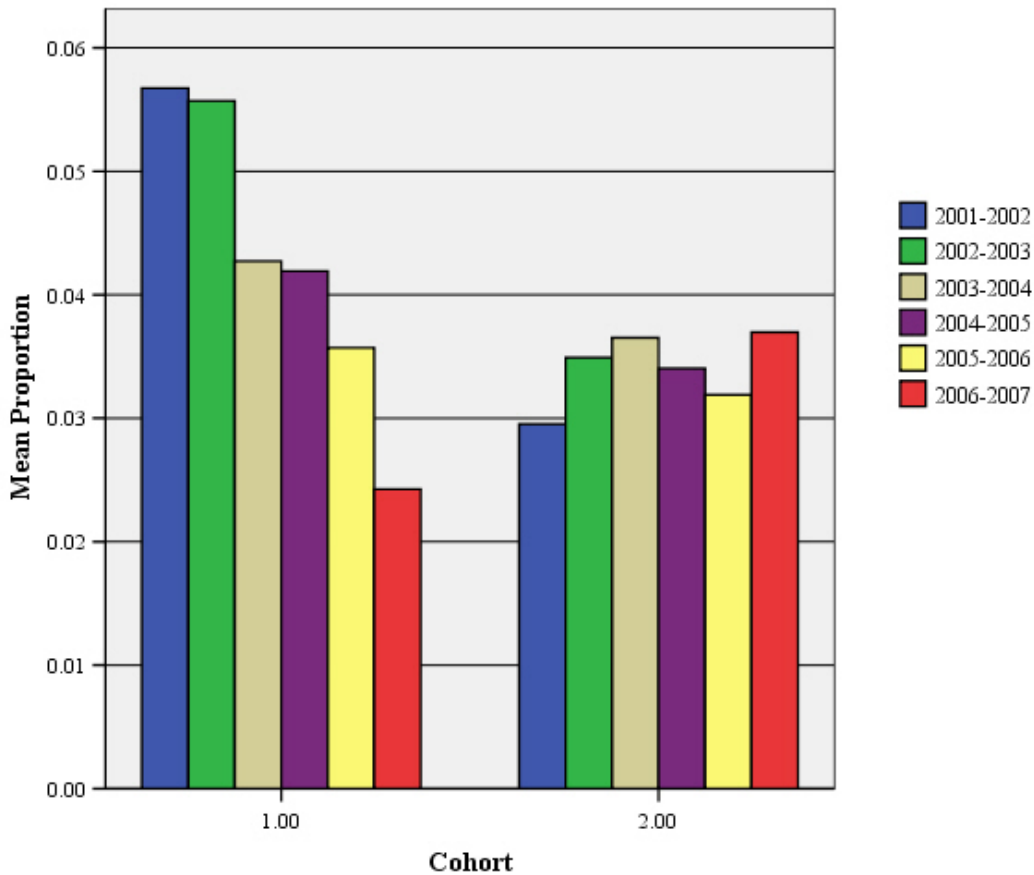
Cohort 1 referrals have dropped each year of IRF. During the two years prior to IRF Cohort 1 schools referred 5-6% of their 1st grade students. By 2005-2006, the rate had dropped to 3.6%. There is a clear downward trend. To put this trend into perspective, in 2001-2002 a total of 41 1st grade students were referred to special education in IRF schools. By 2005-2006 this number had dropped to 29. This represents a 29.3% reduction in the number of students referred. This illustrates how relatively small changes in the proportion of students being referred can result in quite large drops in total number of students being referred across the IRF network of schools. This not only

represents substantial cost savings but also frees up resources for those students who are identified and placed in special education.

First grade Cohort 2 baseline data in Chart 3-2 shows lower referral rates to begin with that are commensurate with the lower rates that have been attained by Cohort 1. These rates appear to have been maintained during the IRF years. Essentially Cohort 2 has been in a sideways trend for a number of years and thus additional years of data are needed to see if the trend breaks out. It could also be that 3-4% is a minimum referral rate that won't decrease over time since it represents the proportion of the population that doesn't respond to the enhanced reading instruction in these schools. This, by the way, is an excellent number and IRF schools should be given high marks for reducing their referrals to this level.

Second grade referral trends are shown in Chart S3-3. Cohort 1 is in a sideways trend with about a 5% referral rate each year. Cohort 2, however, may be experiencing a downward trend, although it is too early to tell.

Chart S3-3: Proportion of Second Grade Students Referred to Special Education by Year and Cohort
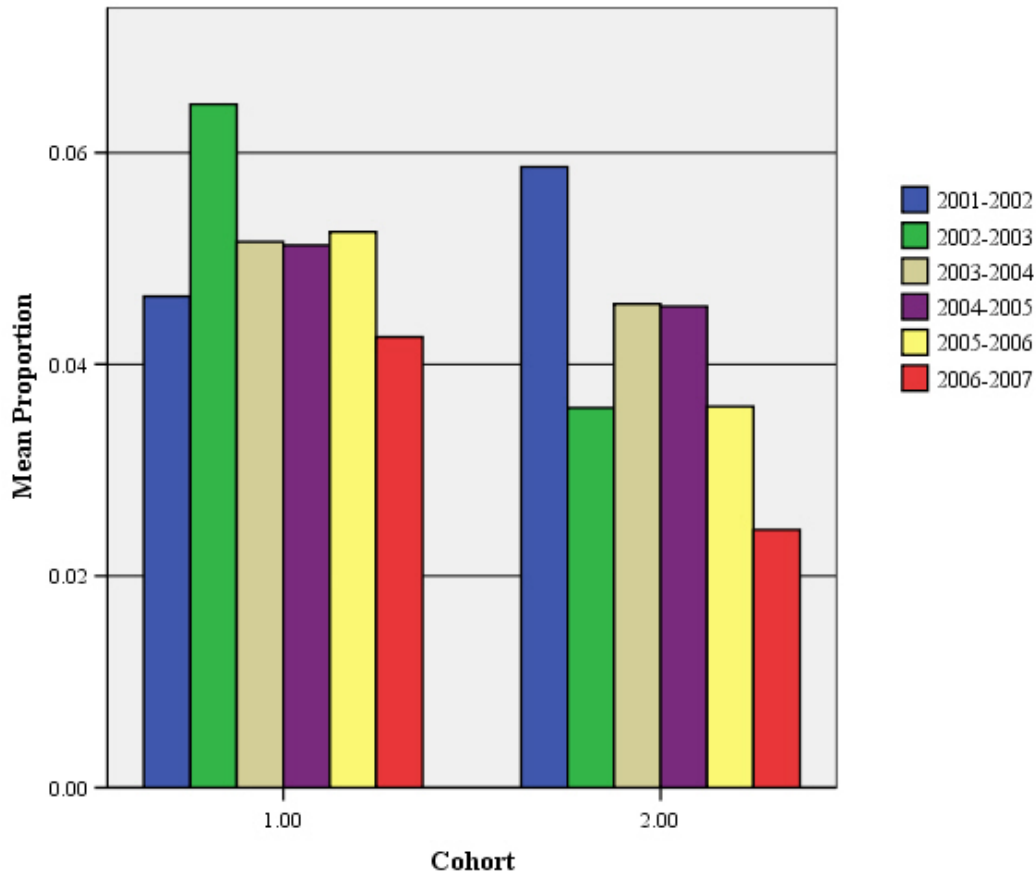
What is interesting about Cohort 2 is that they started with slightly lower referral rates than Cohort 1 and have dropped even lower.

Third grade referral trends are shown in Chart S3-4. It should be noted that the 2006-2007 3rd grade data appears to be quite suspect. The rate of referral is so low in comparison to previous years and the sample size is so small that the data should be disregarded at this time.

Chart S3-4: Proportion of Third Grade Students Referred to Special Education by Year and Cohort
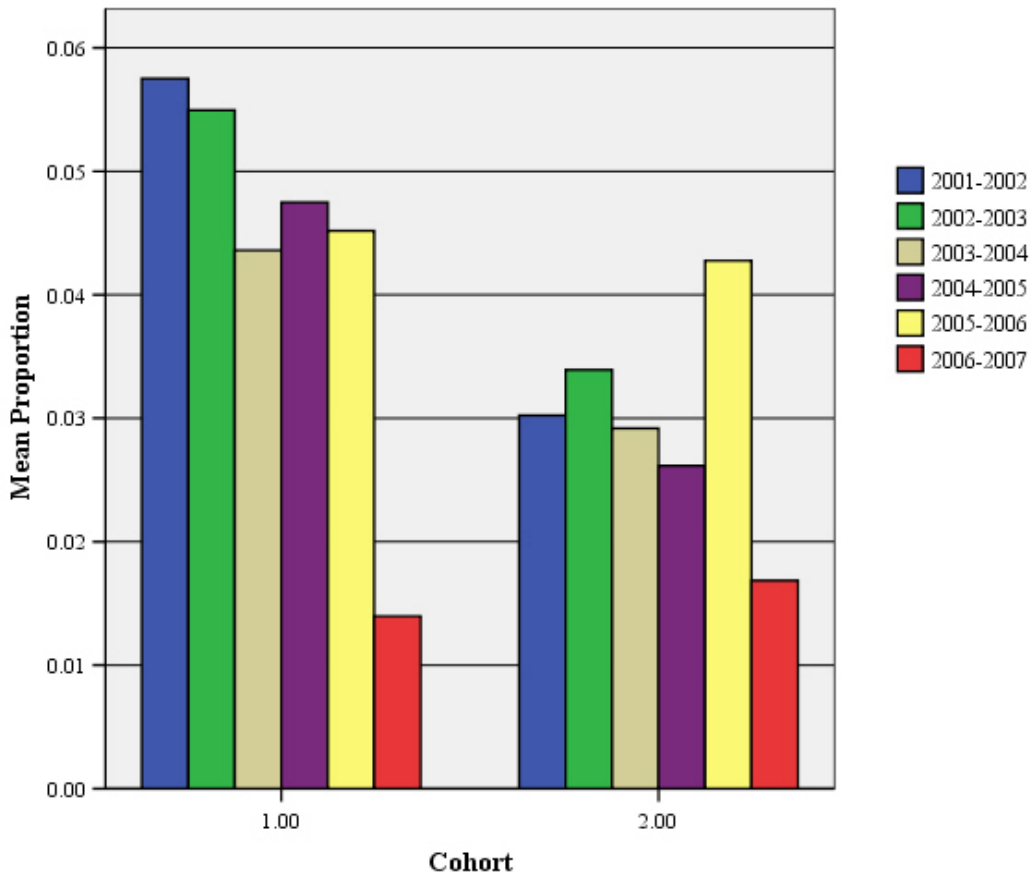
However, looking at the other years reveals that Cohort 1 achieved what appears to be a sustainable reduction in referrals during their first year in IRF (i.e., 2003-2004). The referral rate dropped during this first year and remained roughly at this lower rate for the two following years. Similar to grades 1 and 2, Cohort 2 had lower baseline referral rates and has sustained these lower rates during IRF. Granted rates spiked up in 2005-2006 but only additional years of data will reveal whether this represents an upward trend or just an anomaly. Some insight into this issue can be found in the chart. Referral rates for Cohort 2 schools were trending downward for two years after they hit a high in 2002-2003. It is possible this trend will continue beyond the spike up in 2005-2006.

Keeping in mind the important qualification that the response rate was not good for the special education referral data and the integrity of the data that was submitted is questionable, in aggregate there is accumulating quantitative evidence of some reduction in special education referrals in IRF schools. This represents very positive performance. If trends are confirmed in later years, additional analyses like that done for grade 1 above would be in order to show more specifically what has been accomplished and the cost savings that have resulted.

Student Mobility

In last year's evaluation, student mobility was employed as a predictor of student test scores since it was hypothesized that as student mobility went up test scores would go down. The relationship was not verified by the data, however. Student mobility was not a statistically significant, consistent predictor of student test scores. Although student mobility was not a consistent predictor of student achievement in the 2005-2006 evaluation, the data was collected again this year since knowing how much student inflow and outflow there is in IRF schools can help stakeholders understand the dynamics of these schools.

During the 2005-2006 evaluation schools were contacted and asked to provide the total student enrollment in their buildings during the 2005-2006 academic year. They were also asked to provide the number of new enrollees and the number of students who withdrew. A proportion was computed by summing the number of new enrollees and the number of withdrawals and then dividing it by the total enrollment for the school. This proportion represents the amount of student turnover or mobility in a school. For the 2006-2007 evaluation, the Idaho Department of Education finance office provided the student enrollment data. Student mobility was computed in the same way as described above.

Table S3-8 reports mobility statistics for each school. The mobility average for all schools is .36 (Standard Deviation= .14). This means that on average IRF schools experience a 36% turnover in students each year. The minimum mobility is .12 at Union-Lyman, and the maximum is .66 at West. Some IRF schools experience significant student mobility. Wilson, Roberts, Acequia, Heyburn, East, West, Sherman, Snake River, Bickel, Lincoln and Central Canyon all experienced over 40% student mobility. This is 11 of the 30 IRF schools. Eleven other schools experienced 30-40% student mobility. Four schools have 20-30% and four have under 20% mobility. In short, all but four IRF schools experienced significant student mobility (i.e., greater than 20%).

Table S3-8: 2007 Student Mobility by School

School	District	Mobility	School	District	Mobility
Popplewell	Buhl	.36	Paul	Minidoka	.38
Lewis & Clark	Caldwell	.30	East	Mtn. Home	.51
Sacajawea	Caldwell	.35	West	Mtn. Home	.66
Wilson	Caldwell	.41	Sherman	Nampa	.52
Butte View	Emmett	.28	Snake River	Nampa	.41
Filer	Filer	.20	New Plymouth	New Plymouth	.22
Gooding	Gooding	.34	Bickel	Twin Falls	.49
Homedale	Homedale	.31	Harrison	Twin Falls	.38
Harwood	Jefferson County	.20	Lincoln	Twin Falls	.51
Roberts	Jefferson County	.43	Oregon Trail	Twin Falls	.40
Adams	Madison	.16	Central Canyon	Vallivue	.57
Archer	Madison	.13	East Canyon	Vallivue	.35
Union-Lyman	Madison	.12	West Canyon	Vallivue	.36
Acequia	Minidoka	.45	Wendell	Wendell	.37
Heyburn	Minidoka	.47	Priest River	West Bonner	.16

Test Score Results and Analysis

TEST SCORE RESULTS AND ANALYSIS

A series of standardized tests was administered in Reading First schools, including both state-mandated and additional measures that were not part of the State of Idaho assessment framework. Table S4-1 provides a list of the assessments and the grade levels and times when they were administered. The table is followed by more detail about each instrument, which in turn is then followed by individual sections reporting longitudinal results from the tests since the beginning of Idaho Reading First.

Table S4-1: Assessments Administered by Grade Level and Time

Test	Times of Administration	Grade Levels Administered
Iowa Test of Basic Skills (ITBS)	Spring	K-3
Texas Primary Reading Inventory (TPRI)	K—winter, spring 1-3—fall, winter, spring	K-3
Idaho Reading Indicator (IRI)	Fall, winter, spring	K-3

Iowa Test of Basic Skills (ITBS): The Reading Subtest of the Iowa Test of Basic Skills is administered each spring in IRF schools. The Idaho State Department of Education designated the Iowa Test of Basic Skills (ITBS) as the assessment approved by the U.S. Department of Education for evaluating students' reading performance with respect to grade-level expectations. As described in the National Reading First Assessment Guidelines, a score at the 40th percentile or greater on the National Percentile Rank (NPR) was the criteria used to determine whether a child was reading at his or her grade level.

Texas Primary Reading Inventory (TPRI): The TPRI is an individually administered assessment given in grades K through 3. The following description is taken from the Technical Report of the Texas Primary Reading Inventory, 1998 Edition:

At each grade level, the TPRI consists of a screen and an inventory. The *screen* permits the rapid assessment of individual children. Designations of risk status are yielded, which identify children who most likely *do not* need additional assessment. The *inventory* is a detailed assessment of reading and reading-related skills that allows the teacher to gain more in-depth information that can be used to determine the child's level of risk for reading problems. The inventory is primarily designed to help the teacher set learning objectives for the child. Both the screen and the inventory are individually administered and are designed to be given by a trained teacher (page 11).

Idaho Reading Indicator (IRI): The IRI is a screening measure given three times per year to all K-3 students. The test assesses beginning literacy development in students and is individually administered. The IRI requires about 10 minutes to administer. Although the IRI was originally designed as a screener for teachers to use as a stepping stone for further diagnosis, policy makers in Idaho decided to make it a high-stakes accountability measure by establishing benchmarks that must be met at each grade level each year. The

IRI is scored on a three-category scale (1-3) with students who obtain an overall score of 3 being classified as proficient.

Results from each assessment will be reported in separate sections below.

Iowa Test of Basic Skills: Longitudinal Analysis

Given that this is the fourth year of participation in IRF for some schools and the third year for others, a longitudinal analysis of Iowa Test of Basic Skills (ITBS) scores was undertaken. Granted, for those schools only in their third year of IRF, a clear trend in scores will not be discernible, but trends will be more noticeable for those schools in their fourth year of the initiative. Even a four-year trend, however, is still tentative and in the future five or more years of data will provide much more reliable conclusions.

Although trends in the data at this point are tenuous, it is still important to begin looking at the data longitudinally as soon as possible so the strengths and weaknesses of IRF can be articulated. An important goal of IRF is for schools to steadily improve over time, and trend analysis reveals this. Furthermore, as trends emerge, future staff development and other initiative interventions can be informed by these trends.

The State of Idaho administered the ITBS for the last time in spring of 2001. Idaho 3rd graders scored on average at the 54th percentile, which equates to a grade equivalent score of 3.7. Although the first year of ITBS test data for IRF schools was 2004, a comparison was made between this initial IRF data and the last year of data from the state (i.e., 2001). IRF schools compared quite favorably to the 2001 state average. During spring 2004, IRF 3rd graders averaged a 4.1 grade equivalent score (See Table S4-2 below) and have remained at this level for the past three years. This is substantially above the state average that was last computed in 2001. This represents strong performance for IRF schools, but there remains substantial variability in test score performance across schools and there is little evidence of sustained growth over time.

Table S4-2 provides average grade equivalent scores for each school by grade level and year. The table also includes gain scores by grade and school. Discussion of the findings derived from the table occurs below.

Table S4-2: Average ITBS Grade Equivalent Scores and Gain Scores by Year, Grade and School

School	District	Kindergarten					First Grade					Second Grade					Third Grade					Avg. Gain
		04	05	06	07	G	04	05	06	07	G	04	05	06	07	G	04	05	06	07	G	
Popplewell	Buhl		.7	.8	.8	.1		2.3	1.9	2.0	-.3		3.2	2.9	3.1	-.1		3.9	3.9	3.8	-.1	-.1
Lewis & Clark	Caldwell		.7	.6	.7	0		1.9	2.0	2.0	.1		2.9	2.8	3.1	.2		3.8	3.4	3.7	-.1	.1
Sacajawea	Caldwell	.6	.6	.5	.5	-.1	1.8	2.0	1.8	1.9	.1	3.1	2.7	2.6	2.5	-.6	3.8	3.6	3.7	3.5	-.3	-.2
Wilson	Caldwell	.6	.8	.6	.7	.1	2.0	2.0	1.9	2.0	0	2.9	2.5	3.0	3.6	.7	3.8	4.1	4.0	4.0	.2	.3
Butte View	Emmett		.7	.9	.7	0		2.0	2.0	2.1	.1		3.3	2.7	3.2	-.1		4.2	4.2	4.3	.1	.0
Filer	Filer		1.2	1.3	1.2	0		2.3	2.6	2.4	.1		3.2	3.2	3.5	.3		4.5	4.4	4.2	-.3	.0
Gooding	Gooding	.7	.6	.7	.7	0	2.1	2.1	2.0	2.2	.1	3.0	3.0	3.0	3.1	.1	4.2	4.1	3.9	4.0	-.2	.0
Homedale	Homedale		.7	.7	.6	-.1		2.0	1.9	1.9	-.1		2.9	3.0	3.0	.1		3.7	3.8	3.8	.1	.0
Harwood	Jefferson	1.2	.9	.9	1.0	-.2	2.2	2.1	1.9	2.2	0	3.1	3.1	3.1	3.1	.0	4.3	4.1	4.9	4.2	-.1	-.1
Roberts	Jefferson	.5	.5	.6	.3	-.2	2.4	2.1	2.4	2.4	0	3.2	3.4	2.2	2.6	-.6	3.9	4.2	4.3	3.8	-.1	-.2
Adams	Madison	1.0	1.2	1.4	1.2	.2	2.4	2.4	2.4	2.2	-.2	3.3	3.7	3.7	3.3	.0	4.5	4.8	4.6	4.7	.2	.1
Archer	Madison	.9	1.0	.8	1.0	.1	2.5	2.5	2.9	2.5	0	3.5	3.6	3.3	3.7	.2	4.3	5.0	4.6	4.2	-.1	.1
Union-Lyman	Madison	1.0	1.1	1.1	.9	-.1	2.3	2.3	2.6	2.6	.3	3.3	3.2	3.3	3.6	.3	4.4	4.5	4.4	4.4	.0	.1
Acequia	Minidoka	.7	.8	.9	.7	0	1.8	1.8	2.1	2.4	.6	2.8	3.4	3.4	3.1	.3	3.5	3.7	3.9	4.1	.6	.4
Heyburn	Minidoka		.7	.8	.9	.2		1.9	2.0	1.9	0		2.9	3.0	3.1	.2		3.9	3.7	3.5	-.4	.0
Paul	Minidoka	.6	1.0	.9	.9	.3	2.2	2.0	2.1	2.2	0	2.9	3.4	3.0	2.9	.0	4.1	3.8	4.1	3.5	-.6	-.1
East	Mountain Home	.9	.9	1.0	.8	-.1	2.1	2.4	2.6	2.4	.3	3.1	3.3	3.5	3.4	.3	4.0	4.1	4.3	4.1	.1	.2
West	Mountain Home	.5	.7	.8	.7	.2	2.0	1.9	2.0	2.1	.1	2.9	2.9	3.0	3.1	.2	4.0	4.0	3.9	4.2	.2	.2
Sherman	Nampa		.6	.7	.6	0		1.9	1.9	1.4	-.5	2.6	2.8	2.8	2.8	.2	3.8	3.5	3.5	3.6	-.2	-.1
Snake River	Nampa		.6	.6	.7	.1		1.7	2.0	2.0	.3	2.5	2.6	2.6	2.9	.4	3.8	3.3	3.2	3.3	-.5	.1
New Plymouth	New Plymouth	.6	1.1	1.1	1.0	.4	2.4	2.3	2.5	2.7	.3	3.0	3.4	3.2	3.3	.3	4.6	4.1	4.5	4.4	-.2	.2
Bickel	Twin Falls	1.0	.9	.7	1.4	.4	2.4	2.5	2.1	2.2	-.2	3.4	3.5	3.2	3.4	.0	4.6	4.3	4.1	3.9	-.7	-.1
Harrison	Twin Falls		.7	.8	.7	0		2.2	2.4	2.0	-.2		3.1	3.1	3.3	.2		4.2	4.2	4.2	0	.0
Lincoln	Twin Falls		.7	.7	1.0	.3		2.2	2.3	2.0	-.2		3.2	3.2	3.2	0		4.5	4.2	4.3	-.2	.0
Oregon Trail	Twin Falls	.7	1.0	.6	.7	0	2.1	2.1	2.2	2.0	-.1	3.1	3.0	3.2	2.9	-.2	3.9	3.8	3.9	3.9	.0	-.1
Central Canyon	Vallivue		.8	.8	.6	-.2		2.2	2.0	2.0	-.2		3.1	2.9	2.9	-.2		3.9	3.8	3.6	-.3	-.2
East Canyon	Vallivue	.7	.8	.9	.8	.1	1.9	1.9	2.0	2.1	.2	2.8	2.9	2.8	2.8	0	4.1	3.8	3.8	3.6	-.5	-.1
West Canyon	Vallivue	.8	1.0	.8	.8	0	2.2	2.3	2.3	2.1	-.1	2.7	3.2	3.2	3.1	.4	3.7	4.0	4.0	4.0	.3	.2
Wendell	Wendell	.5	.5	.5	.5	0	2.1	2.1	2.0	1.9	-.2	2.8	3.0	2.9	2.9	.1	3.9	3.9	4.0	3.7	-.2	-.1
Priest River	West Bonner		.9	1.0	1.0	.1		2.3	2.3	2.0	-.3		3.3	3.5	3.2	-.1		4.3	4.3	4.3	0	-.1
Average		.8	.8	.8	.8	.05	2.2	2.1	2.2	2.1	0	3.0	3.1	3.0	3.1	.09	4.1	4.1	4.0	4.0	-.11	

* G = Gain Score   Gain scores are computed in the following manner. For IRF schools who have been in the initiative since 2004: Gain = Average GE 2007 - Avg. GE2004. For schools in the initiative since 2005: Gain = Average GE 2007 – Avg. GE 2005.
** Kindergarten scores are Reading Profile Total scores   *** Grade 1, 2 and 3 scores are Reading Total scores

Table S4-2 provides average grade equivalent scores by school, grade level and year. Composite averages by grade level and year are found in the bottom row of the table. Composite average grade equivalent scores are quite good for all grades except kindergarten. On average students in grades 1-3 are leaving their grade levels with grade equivalent scores equal to or slightly above the next grade level they will enter in the fall. This puts those students who are average or above in excellent positions to begin the next grade. There are, however, half the students who score below the average and are thus not as well-positioned to move into the next grade level in the fall.

Kindergarten composite average scores were mentioned above as not being as high as the other grade levels. Most likely, this lower performance is an artifact stemming from the use of grade equivalent scores. As will be discussed in greater detail below, when normal curve equivalent scores are examined, kindergarteners perform at similar levels to the other grades. Therefore, the lower average grade equivalent score for kindergarteners should not be cause for alarm. Given this, kindergarten scores won't be discussed further until normal curve equivalent scores are reported below.

Gain scores were computed for each school by grade level. There is considerable variability between schools, with some making gains at one or more grade levels, some making few gains and some dropping over time. Average gain scores were also computed for each school and are found in the far right column of Table S4-2. These scores are the arithmetic average of the gain scores across the grade levels. Computing an average for only four scores (i.e., K, 1, 2, 3) is problematic so caution is needed when interpreting these values. They may not be accurate measures of central tendency.

Acequia is the only school that showed sustained positive performance across grades 1-3. Their average gain was .5 for these 3 grades, meaning that across all the grade levels they averaged a five-month increase in achievement between 2004 and 2007. Growth was consistent across grades 1-3 with grades 1 and 3 showing especially consistent positive trends. Take for example Acequia's 3rd grade trend in scores. Between 2004 and 2007 their 3rd graders gained on average six months in reading achievement. Whereas in 2004 third graders averaged 3.5 (i.e., the fifth month of 3rd grade) at the end of the school year, in 2007 they averaged 4.1 (i.e., the first month of 4th grade). Acequia's performance is outstanding and should serve as a model for other IRF schools

No other school stands out for positive, sustained growth in scores. Most schools either gained or lost just a little. A few schools had individual grade levels that showed solid positive performance. For example, Snake River and Wilson at the 2nd grade level made good gains but their 1st and 3rd grade scores either did not go up appreciably or if they did the trend is not clear.

The problem remains that there are too many IRF schools showing no gain or negative gains. Take for example 3rd grade scores. Eighteen IRF schools had negative gains at this grade level and four had gain scores of 0. This means that 22 of 30 IRF schools did not progress at the 3rd grade, and some of the negative gains were substantial. For

example, Bickel and East Canyon both show substantial losses in the 3rd grade, with Bickel's trend line being steadily downward while East Canyon's is only slightly less so. All of the schools experiencing negative growth should be examined to see what might be the causative agents, but those with the most substantial drops should be more intensively scrutinized to see what might be impinging on the schools' ability to make sustained positive growth.

As was mentioned above, students on average in all grades but kindergarten are leaving their grade level at or slightly above the grade level they will be entering the following fall. This is an important and positive finding, but these scores follow mostly normal distributions. This means that half the children fall below the average and thus there remains substantial numbers of children who are below grade level. In order to keep increasing the number of children at the end of each school year who are reading at the next grade level into which they will be moving in the fall, IRF schools need to be experiencing steady growth in their scores from year to year. But this is not occurring. Average grade equivalent scores are holding mostly steady across the years. Future staff development activities should focus on not just maintaining the current solid level of performance but enhancing it.

Another way of examining ITBS scores is through normal curve equivalent scores (NCE scores). NCE scores have a mean of 50 and a standard deviation of 21.06. The benefit of NCE scores is that they can be interpreted in a similar way to percentiles but they don't have some of the measurement weaknesses of percentiles. In Table S4-3, schools that were in the first cohort of IRF have four years of data. Those schools that started in the second IRF cohort have three years. The table is formatted and thus interpreted the same way as the previous table that reported grade equivalent scores.

Table S4-3: Average ITBS Normal Curve Equivalent Scores and Gain Scores by Year, Grade and School*

School	District	Kindergarten**					First Grade***					Second Grade					Third Grade					Avg. Gain
		04	05	06	07	G	04	05	06	07	G	04	05	06	07	G	04	05	06	07	G	
Popplewell	Buhl		47.9	52.0	50.1	2.2		57.5	51.6	53.9	-3.6		54.3	48.1	51.9	-2.4		50.0	49.1	48.4	-1.6	-1.4
Lewis & Clark	Caldwell		47.1	44.3	48.0	.9		50.6	53.1	52.5	1.9		49.8	47.7	52.5	2.7		48.7	41.9	46.0	-2.7	.7
Sacajawea	Caldwell	43.8	43.5	39.9	40.1	-3.7	48.9	54.1	48.2	50.0	1.1	52.8	46.4	44.3	42.0	-10.8	48.3	45.5	46.8	44.2	-4.1	-4.4
Wilson	Caldwell	42.9	49.8	44.2	46.9	4.0	52.5	49.5	51.9	54.7	2.2	49.7	42.6	51.4	60.8	11.1	48.6	51.9	50.5	50.1	1.5	4.7
Butte View	Emmett		49.2	53.4	47.5	-1.7		53.8	53.3	56.2	2.4		56.3	46.7	56.1	-.2		53.5	54.6	56.1	2.6	.8
Filer	Filer		62.7	67.7	64.9	2.2		59.3	67.7	62.7	3.4		54.9	54.8	59.8	4.9		58.0	56.1	54.8	-3.2	1.8
Gooding	Gooding	47.8	44.7	49.0	48.1	.3	55.1	58.1	54.3	59.2	4.1	51.5	52.4	53.0	52.7	1.2	54.6	51.8	49.7	51.2	-3.4	.6
Homedale	Homedale		47.0	48.4	44.8	-2.2		52.3	50.6	52.3	0		48.7	51.7	50.8	2.1		47.0	47.4	48.2	1.2	.3
#Harwood	Jefferson	63.7	56.3	53.3	57.6	-6.1	58.1	54.7	50.3	60.0	1.9	54.0	53.5	52.9	53.0	-1.0	55.5	52.3	64.3	54.4	-1.1	-1.6
#Roberts	Jefferson	41.4	41.6	44.5	34.3	-7.1	64.8	55.3	61.9	66.0	1.2	55.7	58.2	35.4	44.0	-11.7	50.3	54.8	55.7	48.3	-2.0	-4.9
Adams	Madison	58.2	62.8	71.4	65.5	7.3	63.8	62.4	64.6	59.4	-4.4	57.1	62.8	62.3	57.5	.4	58.4	61.3	59.9	61.6	3.2	1.6
Archer	Madison	55.0	57.1	52.6	58.7	3.7	65.8	69.5	76.4	69.1	3.3	61.1	62.7	57.9	64.0	2.9	56.4	64.5	60.1	53.7	-2.7	1.8
Union-Lyman	Madison	57.7	61.4	60.7	56.3	-1.4	60.2	62.6	69.2	70.2	10.0	55.2	55.5	57.6	62.3	7.1	57.7	58.6	56.4	57.5	-.2	3.9
Acequia	Minidoka	45.9	51.6	54.7	46.7	.8	48.2	NS	54.8	63.5	15.3	47.7	58.4	57.8	52.3	4.6	43.2	46.9	49.7	52.8	9.6	7.6
Heyburn	Minidoka		46.7	50.8	56.5	9.8		50.6	53.3	51.5	.9		50.9	52.0	53.9	3.0		50.2	46.5	44.4	-5.8	2.0
Paul	Minidoka	44.8	56.7	56.4	53.4	8.6	56.5	54.0	56.3	58.6	2.1	48.8	58.5	52.5	48.6	-.2	52.8	48.4	52.7	45.0	-7.8	.7
East	Mountain Home	53.3	54.5	57.6	52.1	-1.2	56.9	64.8	67.9	63.1	6.2	54.6	57.3	59.6	58.8	4.2	52.1	53.0	55.2	53.2	1.1	2.6
#West	Mountain Home	41.6	48.2	51.2	46.8	5.2	53.4	51.8	53.5	56.1	2.7	50.2	49.8	52.2	54.0	3.8	50.2	52.0	50.9	54.3	4.1	4.0
Sherman	Nampa		45.7	46.0	43.6	-2.1		50.8	49.5	35.6	-15.2	43.9	47.0	47.5	47.9	4.0	47.5	43.2	43.0	45.2	-2.3	-3.9
#Snake River	Nampa		45.7	43.8	48.2	2.5		44.5	52.4	52.5	8.0	42.2	44.0	44.0	50.8	8.6	47.8	41.1	39.3	40.9	-6.9	3.1
New Plymouth	New Plymouth	44.9	59.3	60.5	57.4	12.5	64.7	62.3	66.2	71.7	7.0	52.2	57.8	55.5	56.8	4.6	59.7	52.6	57.4	56.7	-3.0	5.3
Bickel	Twin Falls	58.0	53.4	47.5	71.1	13.1	62.7	66.5	56.5	57.2	-5.5	58.8	60.2	54.1	58.0	-.8	59.5	55.9	51.3	50.3	-9.2	-.6
Harrison	Twin Falls		49.8	50.5	46.3	-3.5		58.5	63.8	52.5	-6.0		52.1	53.5	57.0	4.9		54.0	53.6	53.3	-.7	-1.3
Lincoln	Twin Falls		50.0	49.0	57.7	7.7		58.2	61.5	55.0	-3.2		55.6	55.2	54.5	-1.1		58.0	53.7	54.8	-3.2	.1
Oregon Trail	Twin Falls	46.4	56.9	44.6	47.6	1.2	56.5	57.6	59.8	52.6	-3.9	52.8	52.0	53.9	50.5	-2.3	49.1	49.1	49.6	49.6	.5	-1.1
Central Canyon	Vallivue		49.9	50.0	44.3	-5.6		59.3	53.9	52.8	-6.5		53.4	50.0	50.7	-2.7		49.7	47.9	46.0	-3.7	-4.6
East Canyon	Vallivue	46.6	52.8	56.3	51.2	4.6	50.5	49.4	53.8	56.5	6.0	46.6	50.6	47.9	46.8	.2	51.7	47.9	48.5	46.1	-5.6	1.3
West Canyon	Vallivue	52.0	57.1	49.8	52.1	.1	56.7	57.1	61.5	55.4	-1.3	45.8	55.3	54.3	52.9	7.1	47.2	50.9	50.5	51.6	4.4	2.6
Wendell	Wendell	42.6	41.2	39.3	39.6	-3.0	53.6	54.2	52.4	50.4	-3.2	47.3	51.1	49.4	51.0	3.7	50.3	49.9	50.9	47.4	-2.9	-1.4
Priest River	West Bonner		55.2	57.1	58.0	2.8		63.5	60.8	52.0	-11.5		57.1	60.8	55.7	-1.4		56.1	55.3	55.5	-.6	-2.7
Averages		49.3	51.5	51.6	51.2	1.7	57.2	56.6	57.7	56.8	.51	51.4	53.6	52.1	53.6	1.6	52.0	51.9	51.6	50.7	-1.5	

68

* G = Gain Score   Gain scores are computed in the following manner. For IRF schools who have been in the initiative for four years: Gain = Average NCE 2007 - Avg. NCE 2004. For schools in the initiative for three years: Gain = Average NCE 2007 – Avg. NCE 2005.

** Kindergarten scores are Reading Profile Total scores

*** Grade 1, 2 and 3 scores are Reading Total scores

\# Harwood 3rd grade 2006 score is perhaps not accurate since n=55 with 35 missing values. Roberts 1st grade 2006 score is perhaps not accurate since n=33 with 13 missing values. West 1st grade 2006 score is perhaps not accurate since n=43 with 33 missing values. West 2nd grade 2006 score is perhaps not accurate since n=67 with 25 missing values. Snake River 1st grade 2006 score is perhaps not accurate since n=55 with 24 missing values.

Information concerning how some of the calculations were made in Table S4-3 is in order. Each column under a grade level that has a year designation contains average NCE scores for each school. To the immediate right of these columns is a column labeled "G." This column contains the gains scores for each school at that particular grade level. This gain score was computed by subtracting either the average for 2004 or 2005 from the average for 2007, depending on whether the school had four years of data or only three. The average gain score in the far right column is an arithmetic average of all of the grade level gain scores for a school. Please note, however, that this composite gain score may not be an accurate measure of central tendency for the following reason. Only four gain scores are averaged and thus a particularly large positive or negative gain score can skew the average appreciably. Readers are cautioned to scan the individual gain scores carefully for a school to see if the overall average gain score for that school is an appropriate measure of central tendency.

The patterns for NCE scores will be quite similar to those discussed above concerning grade equivalent scores since grade equivalents and NCEs are both derived scores. Overall averages are good. When all IRF schools are taken together, average NCE scores for each grade level are at or above the national mean of 50. These numbers are found in the bottom row of Table S4-3. But as was discussed with GE scores, an average around 50 means that half the students are above this average and half are below. This means that substantial numbers of children at all IRF grade levels are functioning below the national average. Additionally, no grade level shows a sustained positive trend in the bottom row of S4-2. Kindergarten and 2nd grades appear to have made quite small gains over the four years and have sustained them, but there is no evidence that growth is continuing to occur. Instead the scores appear to have hovered in the same range the past three years. In the case of 1st and 3rd grades, 1st grade overall averages have hovered around 57 for all four years of the initiative, and 3rd grade scores have trended slightly downward, although again the decrease represents a negligible change.

The best-performing school was Acequia Elementary School. Their overall average gain was 7.6 NCE points. All grades except kindergarten made substantial gains, and there was greater consistency of gain across the grade levels when compared to the other schools. First graders gained 15.3 points, 2nd graders 4.6 points, and 3rd graders 9.6 points. There is one caveat to these findings, however. Acequia had no 1st grade scores for 2005. It appears that the students took the test but did not complete all subscales necessary for a total reading score to be computed. Even taking into consideration this glitch in the data, no other school approached the level of overall gain and consistency that Acequia accomplished. For example, New Plymouth had an overall average gain of 5.3 but their 3rd grade had a negative 3.0 gain score. Granted, their kindergarten gained 12.5, their 1st grade 7.0, and their 2nd grade 4.6, but the negative gain in the challenging 3rd grade attenuates these otherwise excellent results.

Inconsistent performance within schools is a theme in this data. Schools can do quite well at some grade levels but then fail to make progress in the others. For example, Wilson made substantial progress at 2nd grade gaining 11.1 NCE points with emerging evidence of a positive trend. But their performance at kindergarten, 1st grade, and 3rd grade was not nearly as strong. A matter of fact, their 3rd grade scores have hovered around 50 for the four years they have been in the initiative. Another example of

inconsistent performance is Union-Lyman. Their kindergarten and 3rd grade both showed negative gains, although both were negligible. Their 1st grade gained 10 points and their 2nd grade 7.1 points. Perhaps in schools where these inconsistent patterns occur, collaboration between higher-performing grades with lower-performing grades could take place so successful practices could be shared.

Some notable poor performers will also be mentioned. Bickel Elementary averaged a loss of .6 points, but this small amount masks some rather poor performance in all grades but kindergarten. Bickel's kindergarten gained 13.1 points, but all other grades had negative gains. Of greatest concern is the 9.2 loss at the 3rd grade level with the trend quite clearly downward, although a bottom may have been attained. More years of data are needed to establish this, however. Sacajawea was also a poor performer. The school averaged a -4.4 gain across all grade levels. First grade had a negligible 1.1 gain, but all other grades had negative gain scores. Notably, 2nd grade lost 10.8 points and third grade lost 4.1 points. Sherman was similar to Sacajawea. Only 2nd grade showed a positive gain of 4.0. Central Canyon has perhaps the worst gain score profile. Every grade level had a negative gain score: kindergarten at -5.6, 1st grade at -6.5, 2nd grade at -2.7, and 3rd grade at -3.7. Central Canyon has only been in the initiative three years so trends are quite tentative but there is evidence that the scores have gone down and show resistance to moving back up. These schools and others that are similar should be closely examined to see what is occurring that is impacting test score performance so negatively.

Another way of looking at ITBS performance over time is through the 40th percentile criterion established for Reading First schools. A student scoring above the 40th percentile is considered an adequate reader. Table S4-4 reports the percent of students scoring above the 40th percentile by grade level and school on the spring 2007 ITBS. Please note that the first column to the right of the column reporting the number of students in grades 1-3 is a composite of all three grade levels. The individual grade level statistics are located to the right of this column.

Table S4-4: Percent of Grade K-3 Students Over 40th Percentile on ITBS 2007 Reading Subtest by School and Grade

School	District	# of Students in Grade K	% of Grade K Students* > 40th %ile	# of Students in Grades 1-3	% of Grade 1-3 Students** > 40th %ile	% of Grade 1 Students > 40th %ile	% of Grade 2 Students > 40th %ile	% of Grade 3 Students > 40th %ile
Popplewell	Buhl	105	58.7	305	60.1	65.4	57.6	56.7
Lewis & Clark	Caldwell	94	54.3	270	61.7	64.8	61.5	58.5
Sacajawea	Caldwell	92	35.9	259	49.6	60.5	46.4	41.6
Wilson	Caldwell	111	60.4	273	68.8	70.3	75.0	63.6
Butte View	Emmett	128	61.1	400	73.4	67.9	76.9	76.0
Filer	Filer	86	89.5	261	81.2	83.0	78.6	81.8
Gooding	Gooding	105	63.8	279	70.3	76.9	67.0	67.0
Homedale	Homedale	92	51.6	282	60.6	65.2	64.1	55.4
Harwood	Jefferson	106	77.5	268	70.5	85.0	67.3	64.6
Roberts	Jefferson	31	29.0	74	67.1	92.0	54.2	54.2
Adams	Madison	71	94.4	186	83.7	82.6	80.0	87.1
Archer	Madison	28	78.6	58	84.5	93.3	90.9	71.4
Union-Lyman	Madison	51	74.5	131	88.5	92.3	90.9	81.1
Acequia	Minidoka	45	57.8	121	69.4	81.8	69.6	59.5
Heyburn	Minidoka	72	76.1	225	62.4	69.8	67.8	50.0
Paul	Minidoka	77	69.7	220	60.7	78.1	54.7	46.9
East	Mountain Home	74	64.9	247	75.3	81.0	77.9	66.2
West	Mountain Home	97	52.5	257	67.1	63.0	64.6	71.4
Sherman	Nampa	80	47.4	256	49.0	27.7	56.5	53.9
Snake River	Nampa	96	65.3	217	56.7	64.1	63.2	39.3
New Plymouth	New Plymouth	77	74.0	187	81.3	91.2	75.4	78.7
Bickel	Twin Falls	60	94.9	170	74.4	76.8	80.7	65.5
Harrison	Twin Falls	80	57.7	215	69.3	58.7	75.6	71.6
Lincoln	Twin Falls	83	74.7	215	71.5	70.7	72.1	71.9
Oregon Trail	Twin Falls	93	58.1	269	61.7	65.9	62.2	56.8
Central Canyon	Vallivue	153	53.0	413	59.7	66.7	63.6	51.0
East Canyon	Vallivue	166	68.9	449	60.8	72.6	60.2	51.4
West Canyon	Vallivue	106	67.0	328	66.6	71.0	64.8	64.3
Wendell	Wendell	91	39.1	259	64.3	67.4	66.3	58.8
Priest River	West Bonner	84	78.6	232	70.6	64.3	72.4	76.1
Averages			64.3		68.0	72.3	68.6	63.1
Minimum and Maximum			29.0-94.9		49.0-88.5	27.7-93.3	46.4-90.9	39.3-87.1

* Reading Profile Total Scores
** Reading Total Scores

The row second from the bottom of Table S4-4 labeled "Averages" shows the averages for the grade levels collapsed across schools. For example, on average about 68% of grade 1-3 students in IRF schools are reading above the 40th percentile. Seventy-two percent of 2nd graders read above the 40th percentile and 63% of 3rd graders are above this criterion. Conversely, what this means is that roughly 30-37% of students in IRF schools leave their current grade level below this important criterion. This represents excellent overall performance, and this positive outcome needs to be emphasized, but two things significantly detract from this positive finding: (1) the lack of growth over time in scores and (2) the continued performance variability within and between schools.

The first detractor is displayed in Table S4-5. The table compares the 2007 averages to those computed for the 2006 and 2005 ITBS data. The row labeled "Difference" in Table S4-5 supports the assertion made above that IRF schools are not making substantial progress on the ITBS. First and 2nd grade managed a negligible 1-2 point increase over a three year period, but this small gain only came after both grade levels dropped between 2005 and 2006. It is possible that these grade levels will drop again in the future, but it is just as possible that the data reveals the beginning of a positive trend in scores. More years of data are needed to confirm a trend. Third grade performance is weaker with the average percent of 3rd grade students reading above the 40th percentile remaining virtually the same over the three years of testing. In short, what the data reveals is a lack of consistent growth in ITBS scores. The consequence of this lack of positive movement in scores is that until scores begin to move upward each year, 30-37% of IRF students will leave their current grade level below the 40th percentile criterion. This is too many students. Thus it is imperative that test scores begin to consistently move up in all IRF schools.

Table S4-5: Comparison of 2005, 2006, and 2007 ITBS Average Percent of Students by Grade Level Reading Above the 40th Percentile

Testing Year	1st Grade	2nd Grade	3rd Grade
2007	72.3	68.6	63.1
2006	66.5	65.1	63.6
2005	70.5	67.2	63.6
Difference	1.8	1.4	-.5

A second factor that attenuates the positive finding that over 60% of students in IRF schools read above the 40th percentile is the continuing large variability between schools. Schools that have similar demographics employ the same core reading program and have received similar amounts and types of professional development can have quite disparate scores. Table S4-6 shows a subset of IRF schools that have similar demographics but quite disparate achievement. The shaded and unshaded bands contain two schools with similar demographics but dissimilar test score performance. The last school listed in the table is listed by itself and will be discussed below.

Table S4-6: Subset of IRF Schools with Similar Demographics but Disparate Achievement*

School	District	Free or Reduced Lunch	Hispanic	ELL/LEP	% of Grade 3 Students >40th%ile
Wilson	Caldwell	69	50	40	63.6
Snake River	Nampa	73	50	40	39.3
Lewis & Clark	Caldwell	79	55	30	58.5
Sacajawea	Caldwell	77	59	31	41.6
Union-Lyman	Madison	43	16	10	81.1
East	Mountain Home	48	17	7	66.2
Popplewell	Buhl	53	27	26	56.7

The most telling comparison to be made from Table S4-6 is that between Wilson and Snake River. Although Snake River has a slightly higher percentage of students on free or reduced price lunch, the two schools' Hispanic and ELL/LEP populations are identical yet Wilson has 63.6% of their 3rd graders reading above the 40th percentile while Snake River only has 39.3%. Similarly, a comparison between Lewis and Clark and Sacajawea shows quite similar demographics in the schools but quite disparate achievement, and these two schools are in the same district. These disparities are not just found in the high-poverty schools with the largest proportions of Hispanic and ELL/LEP students. The disparities are also found in schools that are less impoverished and have lower proportions of Hispanics and ELL/LEP students. Union-Lyman and East exemplify such schools. They have quite similar demographics but their achievement is quite different. Finally, Popplewell is included since the school's demographics would predict higher test scores yet, Popplewell's performance is slightly lower than Lewis and Clark which has a much more challenging population of students. Popplewell is another case of the variability in IRF school performance. Given that all IRF schools have similar curricula, similar staff training and similar resources provided by their IRF grants, these achievement disparities are troubling. The classroom observations and interviews conducted as part of the annual IRF evaluations provide some specific insights concerning the causative agents behind the achievement disparities. This information should be taken to the schools and discussed to see if teachers, coaches and principals have ideas for how to address the problem.

In summary, with few exceptions, IRF schools are not showing sustained, consistent growth on the ITBS, but averages for all grades represent solid performance that is above the state average for the last year the ITBS was administered in Idaho as a state-wide test. Given that the ITBS is a nationally recognized instrument that is widely used to measure student achievement growth, Idaho Reading First schools' lack of improvement is troubling. Recall that NCE scores have a mean of 50. This mean equates to the 50th percentile. Scanning across the bottom row in Table S4-3 labeled "Averages" shows

most grade levels across the years scoring at or just slightly above the mean. If all students in IRF schools are to make meaningful progress towards grade level performance and if achievement gaps are to be closed, then much better performance on the ITBS in the future is essential.

Closing the achievement gap between Hispanic and White students is an important goal. Table S4-7 reports ITBS results for Hispanic students over the three years of IRF. Please note that in the following table some schools do not have three years of ITBS data reported. If there is no data reported for 2004, then the school started IRF in the second cohort and will thus have only two years of data. If data is missing from other years, then this is due to one of two reasons. Either the school didn't have any Hispanic students take the test in a particular year or there were problems with the test data and no score was reported. This table is interpreted the same way as Table S4-3 above with one exception. There are columns labeled "n" under each grade level. These columns report the number of students who took the test during 2007. This number is included since the number of Hispanic students in IRF schools varies widely and therefore some of the sample sizes are quite small. Statistics from small sample sizes can be highly unreliable so readers should review both the sample sizes and the NCE scores together.

Table S4-7: Average ITBS Normal Curve Equivalent Scores and Gain Scores for Hispanics by Year, Grade and School*

School	District	Kindergarten						First Grade						Second Grade						Third Grade						Avg. Gain
		04	05	06	07	N	G	04	05	06	07	n	G	04	05	06	07	n	G	04	05	06	07	n	G	
Popplewell	Buhl	.	27.8	34.5	31.9	35	4.1	.	42.4	36.1	42.5	37	.1	.	40.6	32.5	36.9	19	-3.7	.	38.9	37.7	31.1	30	-7.8	-1.8
Lewis & Clark	Caldwell	.	39.8	34.8	36.1	48	-3.7	.	43.5	45.4	41.9	53	-1.6	.	44.3	42.1	41.2	47	-3.1	.	41.1	34.9	37.1	46	-4.0	-3.1
Sacajawea	Caldwell	36.2	36.7	33.0	34.9	50	-1.3	46.1	50.0	43.5	45.8	64	-.3	47.5	42.8	40.4	39.5	45	-8.0	41.0	40.0	40.8	39.1	48	-1.9	-2.9
Wilson	Caldwell	30.2	37.3	30.1	39.8	59	9.6	44.6	36.7	42.7	46.4	43	1.8	42.1	40.3	41.3	49.5	33	7.4	41.2	40.3	42.0	39.8	52	-1.4	4.4
Butte View	Emmett	.	35.2	42.2	32.7	22	-2.5	.	43.9	39.7	37.8	21	-6.1	.	49.8	37.3	42.4	23	-7.4	.	48.2	43.8	44.3	16	-3.9	-5.0
Filer	Filer	.	58.1	56.4	48.8	14	-9.3	.	53.3	53.2	53.9	9	.6	.	20.5	52.0	51.3	10	30.8	.	43.4	38.8	49.5	13	6.1	7.1
Gooding	Gooding	28.5	26.6	38.9	35.2	41	6.7	43.3	45.3	44.2	49.9	27	6.6	41.5	45.4	42.8	41.9	30	.4	42.8	41.4	40.3	40.9	19	-1.9	3.0
Homedale	Homedale	.	39.7	39.3	29.2	34	-10.5	.	42.7	45.0	41.9	36	-.8	.	38.5	43.7	41.3	26	2.8	.	38.8	38.8	40.1	49	1.3	-1.8
Harwood	Jefferson	63.7	47.6	33.5	37.7	11	-26.0	36.5	35.7	33.0	34.4	9	-2.1	40.6	43.8	35.7	37.2	14	-3.4	50.3	33.0	62.6	33.3	16	-17.0	-12.1
Roberts	Jefferson	26.8	28.4	36.3	24.2	18	-2.6	55.1	44.0	43.0	53.7	11	-1.4	48.3	50.4	27.4	30.3	8	-18.0	39.3	42.1	49.7	39.1	9	-.2	-5.6
Adams	Madison	27.6	57.5	58.0	51.7	3	24.1	41.8	39.4	48.1	39.8	6	-2.0	47.3	50.3	42.4	54.0	4	6.7	40.0	44.9	46.3	39.5	2	-.5	7.1
Archer	Madison	.	26.0	15.0	20.5	2	-5.5	36.7	.	.	41.0	1	4.3	.	41.0	.	58.0	1	17.0	30.0	43.0	34.3	Nd	0	.	5.3
Union-Lyman	Madison	18.2	48.5	34.8	40.0	9	21.8	38.9	43.3	56.5	49.2	5	10.3	37	39.0	29.0	50.8	11	13.8	43.3	45.2	35.2	37.0	5	-6.3	9.9
Acequia	Minidoka	27.7	37.3	39.9	32.1	17	4.4	36.5	.	32.6	49.9	12	13.4	35.9	47.8	40.4	38.4	15	2.5	30.8	36.0	37.8	35.9	7	5.1	6.4
Heyburn	Minidoka	.	36.8	45.9	51.0	34	14.2	.	46.0	41.4	43.7	33	-2.3	.	48.8	48.9	46.8	35	-2.0	.	43.6	41.4	36.7	39	-6.9	.8
Paul	Minidoka	36.9	45.0	43.6	45.8	28	8.9	42.4	45.8	49.2	49.6	27	7.2	38	46.8	45.4	38.5	32	.5	42.6	40.1	39.4	39.2	31	-3.4	3.3
East	Mountain Home	39.3	39.5	49.3	41.9	11	2.6	46.3	49.3	64.4	50.0	15	3.7	46	53.6	45.8	54.2	13	8.2	46.8	41.6	47.3	41.1	14	-5.7	2.2
West	Mountain Home	31	30.2	36.3	33.3	28	2.3	44.7	41.4	51.0	43.3	22	-1.4	42.3	41.0	41.5	47.0	24	4.7	39.8	45.8	43.6	44.9	25	5.1	2.7
Sherman	Nampa	.	36.1	36.2	34.9	31	-1.2	.	43.0	43.1	31.0	40	-12.0	34.1	42.7	38.3	39.3	34	5.2	38.1	35.1	35.5	39.5	35	1.4	-1.7
Snake River	Nampa	.	36.6	34.6	36.4	46	-.2	.	33.9	41.5	44.4	39	10.5	35.3	43.8	37.0	42.5	33	7.2	43.3	36.6	34.9	36.8	36	-6.5	2.8
New Plymouth	New Plymouth	32.6	37.1	52.0	40.2	10	7.6	55.6	51.6	45.3	56.2	5	.6	38.3	50.2	42.8	36.2	9	-2.1	48.6	41.4	53.5	45.0	8	-3.6	.6
Bickel	Twin Falls	46.6	48.0	36.8	61.4	10	14.8	63.9	59.0	51.0	31.3	3	-32.6	47.1	58.8	53.3	50.0	9	2.9	40.5	55.5	38.1	39.4	8	-1.1	-4.0
Harrison	Twin Falls	.	37.9	43.9	36.9	18	-1.0	.	48.8	56.9	49.5	11	.7	.	44.8	45.9	49.7	15	4.9	.	49.7	47.1	43.5	15	-6.2	-.4
Lincoln	Twin Falls	.	40.7	33.9	47.6	28	6.9	.	57.9	53.3	49.5	19	-8.4	.	59.4	41.2	47.5	11	-11.9	.	58.7	49.5	46.6	13	-12.1	-6.4
Oregon Trail	Twin Falls	34.5	41.3	34.0	35.2	34	.7	38	43.8	47.0	44.3	35	6.3	44.5	36.5	43.9	39.8	29	-4.7	39.7	38.8	34.4	38.2	30	-1.5	.2

School	District	K	Gr1	Gr2	Gr3	N	G	K	Gr1	Gr2	Gr3	N	G	K	Gr1	Gr2	Gr3	N	G	K	Gr1	Gr2	Gr3	N	G	G
Central Canyon	Vallivue	.	42.2	36.8	34.6	72	-7.6	.	49.7	46.7	43.8	62	-5.9	.	45.6	45.5	45.2	69	-.4	.	39.5	39.4	41.6	69	2.1	-3.0
East Canyon	Vallivue	34.7	41.8	45.4	44.0	56	9.3	41.4	40.6	44.4	43.3	38	1.9	34.9	44.2	41.4	39.0	37	4.1	38.7	38.6	42.2	34.1	48	-4.6	2.7
West Canyon	Vallivue	36.7	40.6	40.8	37.9	34	1.2	44.7	43.5	47.6	38.5	34	-6.2	40.8	42.9	42.8	39.9	24	-.9	37.8	42.0	35.8	43.6	22	5.8	.0
Wendell	Wendell	30.2	30.0	31.4	29.3	47	-.9	45.8	45.7	43.6	45.5	52	-.3	41.1	45.3	40.6	44.4	38	3.3	42.1	42.6	43.0	40.0	40	-2.1	.0
Priest River	West Bonner	.	.	56.0	65.5	2	.	.	53.0	59.0	42.5	2	-10.5	.	66.0	47.0	45.0	1	-21.0	.	25.0	56.0	nd	0	.	-15.8
Averages		34.2	38.8	39.5	39.0		2.3	44.6	45.3	46.5	44.5		-.86	41.2	45.4	41.7	43.9		1.2	40.8	41.5	42.1	39.9		-2.6	

* G = Gain Score   Gain scores are computed in the following manner. For IRF schools who have been in the initiative for four years: Gain = Average NCE 2007 - Avg. NCE 2004. For schools in the initiative for three years: Gain = Average NCE 2007 – Avg. NCE 2005.

** Kindergarten scores are Reading Profile Total scores

*** Grade 1, 2 and 3 scores are Reading Total scores

nd = no data

Hispanic performance is inconsistent across the grade levels. The best place to see this is in the bottom row of Table S4-7 which provides overall averages by year and grade level. Kindergarten made steady progress from 2004 to 2006 and then sustained the gains during 2007. First grade made slight gains during 2004-2006 but then fell back to 2004 levels during 2007. Second and 3rd grade scores appear to be bouncing around from year to year with no clear trend.

Looking within the body of Table S4-7 reveals that no school shows consistent growth in scores across the years and grade levels. Most schools show both negative and positive gains across the grade levels. In some cases this high degree of variability is the consequence of some schools having few Hispanic students so scores are going to be quite variable from year to year because of the small sample size. But this only explains a part of the inconsistent performance. There are many IRF schools that have substantial Hispanic populations from year to year and the numbers in the body of the table for each grade level show negative growth rates for some, positive for others, or no change. Two illustrative examples will be provided. Heyburn has 30-40 Hispanic students per grade level. Except for kindergarten where a strong positive gain occurred, all grades show small to large negative gains. Lewis and Clark has 40-50 Hispanics students at each grade level and all of their grade levels show negative gains, although they are not large negative gains so what may be occurring is that scores are staying roughly the same across the years.

One final comparison will be made to put into better perspective how Hispanics are performing in IRF schools. Table S4-8 compares Hispanic and non-Hispanic performance by grade level and year.

Table S4-8: Comparison of Non-Hispanic and Hispanic ITBS NCE Scores by Grade and Year

Group	Kindergarten				First Grade				Second Grade				Third Grade			
	04	05	06	07	04	05	06	07	04	05	06	07	04	05	06	07
Non-Hispanic	55.9	56.4	56.9	57.1	61.9	60.9	62.1	61.1	56.0	57.0	56.4	57.2	56.4	56.0	55.9	55.3
Hispanic	34.2	38.8	39.5	37.4	44.6	45.3	46.5	44.3	41.2	45.4	41.7	42.9	40.8	41.5	42.1	39.3
Diff.	21.7	17.6	17.4	19.7	17.3	15.6	15.6	16.8	14.8	11.6	14.7	14.3	15.6	14.5	13.8	16.0

The achievement gap is not closing but it is not necessarily increasing either. The data shows Hispanic kindergarteners have probably made a 3-5 point NCE gain that appears sustainable. This gain has closed the achievement gap by perhaps a point or two over the four years IRF has been in place. At this rate the gap will take many years to close. However, please keep in mind that these trends and their interpretations are still quite tentative and most anything can happen in future years. For the other three grades there are no clear trends and it appears that scores are trending sideways, going up a little one year and then falling back a little the next. The same is happening for non-Hispanic

students in these grades so the net effect is the achievement gap remains roughly the same from year to year.

In closing, IRF schools on average are doing well. Their average scores fall roughly at the national average and for impoverished schools this is very good performance. But continued variability in scores within and between schools and little evidence of sustained positive growth in scores remain significant concerns going forward.

Texas Primary Reading Inventory (TPRI): Longitudinal Analysis

Two components of the TPRI were selected to report, depending on the grade level of the students: (1) The TPRI Screener for kindergarten and (2) The Word List for Story Placement for grades 1-3. The screeners provide a quick way to rule out students who don't manifest risk characteristics for reading difficulties. Students score either at the "Developed" level or the "Still Developing" level on the screeners. Those scoring at the "Still Developing" level are at risk for reading difficulties. The Word List for Story Placement provides a good metric with which to measure growth in students' passage reading ability. Students read down a list of words that predicts performance on the reading passages contained in the test. The more words the student reads, the higher his or her predicted reading performance and the more difficult the story in which he or she is placed for assessment. More details concerning the measure will be provided below as results are presented and discussed.

Only three years of TPRI data will be considered because there is only three years of data available in the database according to the database administrator. The three academic years are 2004-2005, 2005-2006 and 2006-2007. Three years of data are not enough to establish clear trends, but analyzing potential trends is preferable to just looking at a one year "snapshot" of IRF school performance. Multiple tables are provided because of the large amount of data. The amount of data became quite large as a consequence of the need to add the "N" columns for each grade level and year. These columns provide the number of students who were administered the TPRI in each school at a particular grade level and year. This additional information was necessary because there are inconsistencies in the data sets. The number of students at some grade levels in some schools fluctuates quite dramatically from year to year. This calls into question the validity of some of the comparisons across years. For example, in Table S4-9 Sherman has 22 grade 1 students' scores at the end of 2005 but 92 grade 1 students' scores at the end of 2007. The second table, Table S4-10, manifests similar problems. For example, Wilson has eight grade 2 students' scores at the end of 2005, but the number of grade 2 students jumps to 97 at the end of 2006 and then falls back to 65 at the end of 2007. Thus, as Table S4-9 and S4-10 are interpreted, the number of students across the years at a given grade level needs to be carefully examined and if large discrepancies exist, then comparison should be made cautiously.

Table S4-9 reports the proportion of all kindergarten students who scored at the "Developed" level on a screener at the end of each school year by school. The table also reports by grade level and school the proportion of grade 1 students who were placed in

the highest-level story at the end of each year. Table S4-10 reports the story placement statistics for all students in grades 2 and 3. Both tables include columns for gains or losses made across the three years. An example will aid in interpreting the tables. Examining the first row of school data in Tables S4-9, 78% of 95 Popplewell kindergarteners exited the 2004-2005 school year at the developed level on the screeners. Ninety-one percent of 91 kindergarteners did so in 2006, and 76% of 107 kindergarteners were developed on the spring screeners in 2007. The percentage in the Gain/Loss column is -.02. This means that between 2005 and 2007 the number of kindergarteners in Popplewell who passed the screeners at the end of the school year dropped by 2%. The gain score or loss is computed by subtracting the 2005 percentage from the 2007 value. The rationale behind computing the gain scores in this way follows. An assumption is made that test scores should improve over time in IRF schools. Therefore to arrive at the most accurate gain score, taking the earliest year of data for a school and subtracting it from the most recent should provide the most accurate gain score.

First, 2nd, and 3rd grade scores are all interpreted the same way so a single example should suffice. Continuing with the row of data for Popplewell in Tables S4-9, 59% of 87 1st graders were placed in the most difficult story at the end of 2005. At the close of 2006, 51% of 106 1st graders were placed in the most difficult story, and at the end of 2007 57% of 108 were placed in the highest story. The change from 2005 to 2007 was a loss of 2%. Table S4-10 contains the data for 2nd and 3rd grade and is interpreted the same way that 1st grade was interpreted immediately above. Please note that there is a row in Tables S4-9 and Table S4-10 for unnamed schools. This row is a consequence of some 2005 data in the data set not having school identifiers. This appears to have occurred only during 2004-2005.

Table S4-9: TPRI Results by Grade, School and Year: Grades K and 1 All Students

School	District	Spring K 05	Spring K 06	Spring K 07	Gain/ Loss K	N K 05 n=2303	N K 06 n=2296	N K 07 n=2451	Spring G1 05	Spring G1 06	Spring G1 07	Gain/ Loss G1	N G1 05 n=2363	N G1 06 n=2335	N G1 07 n=2420
Popplewell	Buhl	.78	.91	.76	-.02	95	91	107	.59	.51	.57	-.02	87	106	108
Lewis and Clark	Caldwell	.96	.98	.99	.03	85	83	96	.66	.53	.67	.01	64	91	92
Sacajawea	Caldwell	.93	.91	nd	.	82	32	nd	.57	.45	.49	-.08	92	60	37
Wilson	Caldwell	.93	.95	.99	.06	72	88	111	.56	.69	.58	.02	99	59	93
Butte View	Emmett	.98	.97	.92	-.06	142	141	131	.40	.31	.50	.10	126	139	141
Filer	Filer	1.00	.98	.99	-.01	73	81	87	.63	.65	.57	-.06	87	88	90
Gooding	Gooding	.87	.96	.93	.06	68	84	105	.64	.50	.74	.10	104	103	90
Homedale	Homedale	.94	.98	1.00	.06	51	95	92	.57	.55	.54	-.03	113	82	90
Harwood	Jefferson	.98	.94	.99	.01	97	99	102	.61	.48	.61	.00	100	106	115
Roberts	Jefferson	.97	1.00	.94	-.03	29	24	32	.50	.85	.81	.31	20	27	26
Adams	Madison	1.00	1.00	.97	-.03	62	71	70	.65	.62	.74	.09	57	63	69
Archer	Madison	1.00	.95	1.00	.00	23	21	29	.79	.95	1.00	.21	19	20	15
Union-Lyman	Madison	1.00	1.00	1.00	.00	52	36	50	.74	.89	.82	.08	35	53	39
Acequia	Minidoka	1.00	.91	.89	-.11	47	35	45	.66	.76	.86	.20	35	45	35
Heyburn	Minidoka	.92	1.00	.96	.04	37	59	67	.38	.42	.51	.13	8	72	67
Paul	Minidoka	.94	.97	1.00	.06	68	78	79	.41	.58	.59	.18	68	77	73
East	Mtn. Home	1.00	.95	.97	-.03	64	79	70	.83	.73	.64	-.19	70	70	84
West	Mtn. Home	.82	.91	.99	.17	77	77	95	.71	.71	.53	-.18	84	75	87
Sherman	Nampa	.89	.85	.80	-.09	19	88	78	.32	.37	.37	.05	22	75	92
Snake River	Nampa	.90	.96	.95	.05	61	53	76	.71	.75	.71	.00	45	60	63
New Plymouth	New Plymouth	.98	1.00	1.00	.02	66	56	76	.48	.74	.32	-.16	71	65	60
Bickel	Twin Falls	.92	.84	nd	.	66	56	nd	.63	.55	.61	-.02	56	58	18
Harrison	Twin Falls	.95	.95	1.00	.05	78	79	81	.63	.70	.48	-.15	73	80	86
Lincoln	Twin Falls	.94	.91	.99	.05	69	80	84	.79	.74	.58	-.21	61	78	84
Oregon Trail	Twin Falls	.99	.88	.99	.00	75	80	94	.60	.65	.53	-.07	78	74	92

Central Canyon	Vallivue	.	.98	.93	.	9	110	154	.60	.56	.51	-.09	15	121	127
East Canyon	Vallivue	.	.96	.02	.	9	142	159	.33	.45	.53	.20	15	143	146
West Canyon	Vallivue	.	.85	.94	.	5	91	106	.	.61	.54	.	1	93	116
Wendell	Wendell	.91	.93	.39	-.52	89	100	90	.50	.48	.37	-.13	96	88	99
Priest River	West Bonner	.88	.90	.45	-.43	60	87	84	.59	.49	.38	-.21	69	71	85
Unnamed School		.90	.	.	.	402	.	.	.55	.	.	.	421	.	.
Averages*		.94	.94	.85	-.03				.59	.61	.60	.003			

* Not including unnamed schools

82

When looking at the performance of kindergarteners over the three years, overall school averages did not change. Some schools increased from 2005 to 2007 but others decreased, leaving the overall averages found in the bottom row of Table S4-9 nearly identical. Granted, the overall average for 2007 dropped substantially from the previous years, but this is explained by a few schools (i.e., East Canyon, Wendell and Priest River) dropping dramatically between 2006 and 2007. When these schools are removed from the average, the overall average remains virtually identical to the previous years (i.e., 94%). These precipitous falls are probably not due to a substantial decrease in student performance but to an artifact of who was administered this particular portion of the test. It appears that in these schools only a few students were given the screeners during the spring. The students were given other portions of the TPRI test and thus show up in the N column, but they were not administered the screeners. Only the most at-risk students in these schools appear to have been given the screeners during the spring testing, which would give rise to the large drop in the percentage of students passing a screener

Overall IRF school performance on the TPRI at the kindergarten level is excellent. All but a few schools consistently have more than 90% of their students in the spring pass a screener. Popplewell and Sherman are the only schools that do not have as high nor as consistent performance. Perhaps their kindergarten programs should be examined more closely since there is ample evidence across the IRF network that high percentages are both achievable and sustainable.

Part of the TPRI assessment for grades 1-3 is reading a list of words. For grades 1 and 2 the criterion is simply the number of words read correctly from the list. For grade 3 the criterion shifts from a simple count of the number of words students read correctly from the list to a computed words-read-correctly-per-minute value. No matter the grade level, the students' performance on this list is the criterion used to place them in an appropriately leveled story for them to read. The stories range in difficulty. The first story in the set is considered a beginning-of-year story and the last one in the set is considered a story quite close to the beginning of the next grade level up. For example, in grade 1 there are five stories. The first one is a beginning-of-first grade level story. The fourth one is considered an end-of-first grade benchmark, and the fifth story is considered to be close to a beginning-of-second-grade story. Other grade levels have a different number of stories but their progression can be interpreted similarly. Thus, a good measure of student progress in reading on the TPRI is growth in the level of story placement from the first time the test is administered to the end-of-year administration. In Table S4-9 and Table S4-10 the columns labeled "Spring" represent the proportion of students who were placed in the highest story at the end of the academic year. There is an exception, however. In the case of grade 3, the proportions in the "Spring" columns represent the proportion of students who were placed in either of the last two stories in the sequence. There are six stories in this particular sequence and the last two represent strong reading performance, so both were included.

Looking at the last row in Table S4-9 at the IRF School Averages, 1st grade achieved quite stable performance with about 60% of students being placed in the highest story. The problem is that there is no evidence of a trend of consistently improving performance in these overall averages. Additional concern comes from the continued variability at the individual school level. Some IRF schools remain quite consistent across the years. Examples of these schools are Homedale, Sherman, and Snake River. Other IRF schools made good gains from 2005 to 2007. Examples of these schools are Roberts, Archer, and Acequia. Two of these schools (i.e., Roberts and Archer) appear to have experienced substantial initial jumps and have since sustained their higher performance for the past two years. The third school, Acequia, has achieved a steady 10% increase each year for the past three years. No matter whether the school jumped quickly to a high relative performance and remained there or is steadily growing to high relative performance, these three schools represent solid performance over time. But there are other schools that show quite steady declines in 1st grade percentages over the three years. East and Priest River show the most consistent declines, having dropped about 10% per year over the three years.

This variability in IRF schools remains troublesome since all children should receive the opportunity to improve year-after-year, not just those in the schools where improvement is consistently being made. With three years of data now, trends are more evident but are still somewhat tentative because of the volatility in test scores from year to year. But those schools such as East and Priest River who are consistently trending downward and those that are consistently trending sideways should be examined to see what can be done to improve performance. We now turn to a discussion of grades 2 and 3, the results of which are presented in Table S4-10 below.

Table S4-10: TPRI Results by Grade, School, and Year: Grades 2 and 3 All Students

School	District	Spring G2 05	Spring G2 06	Spring G2 07	Gain/ Loss G2	N G2 05 n=2200	N G2 06 n=2254	N G2 07 n=2328	Spring G3 05	Spring G3 06	Spring G3 07	Gain/L oss G3	N G3 05 n=2257	N G3 06 n=2179	N G3 07 n=2320
Popplewell	Buhl	.96	.81	.85	-.11	67	88	99	.21	.24	.22	.01	76	89	96
Lewis and Clark	Caldwell	.93	.87	.94	.01	67	71	88	.44	.31	.40	-.04	68	74	76
Sacajawea	Caldwell	.93	nd	nd	.	54	nd	Nd	.23	nd	.19	-.04	66	nd	37
Wilson	Caldwell	.88	.88	.94	.06	8	97	65	nd	.37	.46	.	10	81	106
Butte View	Emmett	.90	.95	.96	.06	99	129	133	.25	.33	.44	.19	129	141	125
Filer	Filer	.90	.93	.95	.05	72	86	84	.36	.18	.42	.06	78	84	89
Gooding	Gooding	.90	.95	.97	.07	87	93	97	.37	.21	.39	.02	87	94	91
Homedale	Homedale	.83	.95	.89	.06	47	113	79	.29	.24	.18	-.11	98	84	114
Harwood	Jefferson	.89	.82	.83	-.06	79	100	104	.62	.45	.46	-.16	74	108	104
Roberts	Jefferson	.94	.86	.96	.02	34	21	24	.48	.07	.21	-.27	33	28	24
Adams	Madison	.98	.96	1.00	.02	48	55	54	.36	.24	.25	-.11	56	54	59
Archer	Madison	.96	1.00	1.00	.04	25	23	22	.27	.36	.29	.02	15	22	21
Union-Lyman	Madison	.93	.97	1.00	.07	55	33	54	.50	.41	.51	.01	36	54	37
Acequia	Minidoka	1.00	.97	.83	-.17	8	34	48	.36	.37	.44	.08	45	46	41
Heyburn	Minidoka	nd	.88	.93	.	4	56	86	nd	.13	.21	.	1	40	68
Paul	Minidoka	.95	.97	.88	-.07	65	63	83	.25	.42	.27	.02	75	64	67
East	Mtn. Home	.95	.97	.94	-.01	66	71	86	.26	.11	.38	.12	82	73	79
West	Mtn. Home	.89	.86	.91	.02	90	91	80	.32	.43	.36	.04	96	93	95
Sherman	Nampa	.88	.87	.82	-.06	32	90	91	nd	.22	.27	.	8	91	77
Snake River	Nampa	.74	.81	.91	.17	39	75	78	nd	.33	.26	.	13	69	80
New Plymouth	New Plymouth	.93	.95	.54	-.39	57	66	71	.04	.83	.89	.85	23	23	61
Bickel	Twin Falls	.94	.91	nd	.	51	53	Nd	.44	.31	nd	.	66	54	nd
Harrison	Twin Falls	.89	.90	.96	.07	84	77	81	.48	.47	.49	.01	80	85	71
Lincoln	Twin Falls	.97	.95	.96	-.01	64	63	68	.41	.32	.42	.01	63	62	64

Oregon Trail	Twin Falls	.90	.92	.91	.01	69	86	90	.47	.44	.49	.02	76	70	87
Central Canyon	Vallivue	nd	.85	.92	.	2	127	144	nd	.37	.25	.	9	114	143
East Canyon	Vallivue	nd	.89	.87	.	7	149	151	nd	.24	.24	.	7	125	155
West Canyon	Vallivue	nd	.91	.91	.	2	89	109	nd	.25	.36	.	1	97	100
Wendell	Wendell	.93	1.00	.93	.00	73	83	83	.33	.40	.47	.14	64	83	79
Priest River	West Bonner	.88	.94	.80	-.08	73	71	76	.17	.17	.26	.09	75	77	74
Unnamed School		.89	nd	nd	.	617	nd	Nd	.27	nd	nd	.	541	nd	nd
Averages*		.92	.91	.90	-.01				.34	.32	.36	.04			

* Not including unnamed schools

Grade two is quite similar to kindergarten. All the schools have greater than 80% of their 2nd graders being placed in the highest story at the end of the three academic years. There is one exception to this rule. New Plymouth during 2007 dropped precipitously to 54%. On closer examination it appears that perhaps only the most at-risk readers were given this portion of the test this particular year. The overall IRF School Averages found in the bottom row of Table S4-10 are all slightly above 90% which is stellar. Either Idaho Reading First 2nd grade curriculum and instruction are outstanding, leading to these strong, positive results, or the TPRI for 2nd graders is quite easy. The assessment does include a second word list with accompanying stories that are more difficult, but only 204 students were administered this more difficult word list in 2005, 194 in 2006, and 144 in 2007.

A few schools need to be mentioned regarding their 2nd grade performance. Snake River manifests an emerging trend of solid improvement. Butte View, Filer, Gooding, Adams, Archer, Union-Lyman, East, Lincoln, Oregon Trail and Wendell have over 90% of their 2nd graders placed in the highest story year after year. This represents excellent performance by these schools. Two schools, however, appear to under-perform on a relative basis, although their performance is still quite good. Harwood and Sherman don't achieve the high percentages of the other schools.

Grade 3 performance is quite the opposite of that found in kindergarten and grade 2. On average about one-third of 3rd graders are placed in one of the highest two stories. This percentage has held relatively constant across the two years. Granted, there was a 2% drop from 2005 to 2006, but additional years of data are needed to establish a trend. Most schools struggle to achieve 50% of students being placed in one of the two highest stories, but there are notable exceptions. New Plymouth stands out with 83% of their 3rd graders being placed in one of the two highest stories. They achieved a 78% gain over the two years. Priest River, however, is on the other end of the performance continuum. During both years, only 17% of their third grade students were placed in the one of the highest two stories at the end of the academic year. It is important to also note that the number of students tested each year in Priest River remained quite consistent, lending credibility to these findings, although additional years are needed to establish solid trends.

Closing the achievement gap is an important goal of IRF. The discussion now turns to an examination of TPRI scores for Hispanic students. Table S4-11 presents Hispanic student scores and is interpreted the same way as the previous two TPRI tables.

Table S4-11: Hispanic TPRI Results by Grade, School and Year: Grades K and 1

School	District	Spring K 05	Spring K 06	Spring K 07	Gain/Loss K	N K 05 n=484	N K 06 n=659	N K 07 n=769	Spring G1 05	Spring G1 06	Spring G1 07	Gain/Loss G1	N G1 05 n=615	N G1 06 n=640	N G1 07 n=730
Popplewell	Buhl	.76	.85	.62	-.14	21	27	34	.50	.32	.47	-.03	24	22	34
Lewis and Clark	Caldwell	.93	.95	.98	.05	42	44	49	.51	.41	.60	.09	39	46	53
Sacajawea	Caldwell	.92	.87	nd	.	49	23	Nd	.57	.38	.41	-.16	58	39	22
Wilson	Caldwell	.94	.93	1.00	.06	34	44	61	.42	.57	.49	.07	45	30	43
Butte View	Emmett	.96	1.00	.84	-.12	27	22	19	.38	.19	.36	-.02	16	27	25
Filer	Filer	1.00	1.00	1.00	.00	11	6	16	.67	.64	.43	-.24	15	11	7
Gooding	Gooding	.79	.94	.95	.16	28	31	39	.58	.45	.71	.13	19	31	31
Homedale	Homedale	.94	1.00	1.00	.06	18	40	34	.49	.46	.39	-.10	51	28	38
Harwood	Jefferson	1.00	.88	.90	-.10	10	8	11	.33	.09	.33	.00	15	11	12
Roberts	Jefferson	1.00	1.00	.94	-.06	8	13	18	.27	.80	.73	.46	11	10	11
Adams	Madison	1.00	1.00	1.00	.00	3	5	1	.33	.67	.83	.50	6	6	6
Archer	Madison	1.00	1.00	1.00	.00	1	1	2	nd	nd	1.00	.	nd	nd	1
Union-Lyman	Madison	1.00	1.00	1.00	.00	10	5	8	.50	.82	.40	-.10	4	11	5
Acequia	Minidoka	1.00	1.00	.82	-.18	18	8	17	.50	.59	.80	.30	10	17	10
Heyburn	Minidoka	.94	1.00	.97	.03	16	30	30	.33	.18	.41	.08	6	33	32
Paul	Minidoka	.93	1.00	1.00	.07	27	26	27	.33	.59	.54	.21	39	27	26
East	Mtn. Home	1.00	.89	1.00	.00	8	18	10	.55	.67	.44	-.11	11	9	18
West	Mtn. Home	.73	.86	1.00	.27	26	21	29	.60	.73	.57	-.03	30	26	23
Sherman	Nampa	.83	.81	.76	-.07	6	31	34	.00	.24	.33	.33	6	21	43
Snake River	Nampa	.76	1.00	.92	.16	17	15	26	.59	.61	.77	.18	17	18	22
New Plymouth	New Plymouth	.88	1.00	1.00	.12	8	7	9	.54	.38	.00	-.54	13	8	6
Bickel	Twin Falls	.88	.86	nd	.	8	7	nd	.33	.50	.00	-.33	6	6	2
Harrison	Twin Falls	1.00	1.00	1.00	.00	9	14	18	.64	.67	.50	-.14	14	9	14
Lincoln	Twin Falls	.85	.88	1.00	.15	13	17	30	.78	.50	.55	-.23	18	16	20
Oregon Trail	Twin Falls	1.00	.72	.97	-.03	24	25	36	.41	.56	.38	-.03	29	25	34

Central Canyon	Vallivue	.50	.95	.95	.45	2	44	74	.33	.54	.36	.03	6	56	61
East Canyon	Vallivue	.75	.95	.02	-.73	4	43	53	.17	.29	.32	.15	6	38	41
West Canyon	Vallivue	nd	.80	.97	.	nd	30	33	1.00	.53	.30	-.70	1	19	34
Wendell	Wendell	.85	.91	.40	-.45	41	54	50	.43	.45	.29	-.14	47	40	55
Priest River	West Bonner	nd	nd	.00	.	nd	nd	1	0	nd	.00	.00	1	nd	1
Averages		.82	.93	.86	-.01				.45	.49	.46	-.01			

Grades K and 1 will be discussed here and then results for grades 2 and 3 will be presented in Table S4-12. Again, when interpreting Table S4-11 both the proportions of Hispanic students and the number of Hispanic students needs to be looked at since some schools have few Hispanic students and others have substantial variability in the number of students from year to year. Additionally, there are a few anomalies in the numbers of students tested at a given school that need to be considered. Notably the number of students being tested at Sherman and Snake River appear to be low, considering these schools' substantial Hispanic populations, and the Vallivue schools all lack adequate data for 2005.

Hispanic kindergarteners appear to be doing quite well. The three-year trend shows emerging evidence that about 90% of Hispanic kindergarteners are passing a screener at the end of the school year, but identifying this emerging trend is strictly conjecture at this time. Scores next year could drop further or move back up. If the 90% estimate holds, however, then this percentage is for all intents and purposes the same as the three-year trend for the whole group (about 91%) reported in Table S4-9 above.

Closing the achievement gap between Hispanics and Whites is an important educational goal. About 95% of White kindergarteners pass a screener in the spring. If the 90% estimate holds for Hispanic kindergarteners then they are only about 5% behind. This represents strong performance on the part of IRF schools. It is quite possible that the achievement gap between Hispanics and Whites, as measured by the TPRI screeners, could close in the next few years.

Extra caution must be taken when looking at trends in Hispanic kindergarten scores because of small and variable sample sizes and the finding that some schools don't administer the screeners to all kindergarteners in the spring but instead focus on the most at-risk readers. Keeping this in mind, Gooding and West stand out as schools that have probably made substantial, consistent progress. Identifying sideways trending or downward trending schools is problematic given that there is evidence that not all students are administered screeners in the spring. Finally, on a quite positive note, 11 of the schools had 100% of their Hispanic students pass a screener in the spring of 2007.

Turning now to Hispanic 1st grade performance, the overall averages in the last row of Table S4-11 reveal that 45-49% are placed in the most difficult story at the end of 1st grade. This places Hispanic 1st graders about 10-15% behind their White counterparts and there is currently no evidence that the gap is closing. Some individual schools, however, have shown consistent growth in performance. Roberts and Adams are examples. Although they don't test many students each year, the numbers that they do test are quite consistent and their scores have consistently improved. Paul is another example. Paul tests more students than Roberts and Adams and Paul's relative performance is lower, but the school has shown quite dramatic improvement in their scores, and there is emerging evidence that the improvement is being sustained.

Identifying sideways trending or downward trending schools concerning 1st grade performance is more problematic because of widely varying sample sizes and high variability in scores across the years, but a few will be mentioned. Popplewell, West and Oregon Trail may be trending sideways. Lincoln is perhaps the clearest case for a school that is trending downward. The school dropped about 25% between 2005 and 2006 and the drop was sustained in 2007. If scores do not rebound in 2008, then Lincoln should be looked at more closely to identify the source of the low test scores.

Table S4-12 reports the results for Hispanic 2nd and 3rd graders. The table is interpreted the same way as those above.

Table S4-12: Hispanic TPRI Results by Grade, School and Year: Grades 2 and 3

School	District	Spring G2 05	Spring G2 06	Spring G2 07	Gain/Loss G2	N G2 05 n=425	N G2 06 n=676	N G2 07 n=654	Spring G3 05	Spring G3 06	Spring G3 07	Gain/Loss G3	N G3 05 n=505	N G3 06 n=634	N G3 07 n=724
Popplewell	Buhl	.94	.63	.74	-.20	17	27	19	.23	.23	.10	-.13	13	22	31
Lewis and Clark	Caldwell	.87	.84	.91	.04	39	43	47	.38	.27	.37	-.01	39	48	46
Sacajawea	Caldwell	.91	nd	nd	.	32	nd	Nd	.25	nd	.18	-.07	44	nd	22
Wilson	Caldwell	1.00	.85	.91	-.09	5	46	32	.29	.40	.42	.13	7	43	48
Butte View	Emmett	.79	.89	.96	.17	14	18	23	.25	.35	.37	.12	24	20	20
Filer	Filer	.50	1.00	1.00	.50	2	13	10	.45	.00	.45	.00	11	6	11
Gooding	Gooding	.84	.89	.97	.13	25	18	32	.24	.17	.35	.11	29	24	17
Homedale	Homedale	.77	.93	.78	.01	22	54	27	.30	.22	.10	-.20	43	41	52
Harwood	Jefferson	.50	.65	.82	.32	12	17	11	.71	.41	.33	-.38	7	17	18
Roberts	Jefferson	.93	.82	1.00	.07	15	11	10	.53	.08	.09	-.44	15	12	11
Adams	Madison	1.00	1.00	1.00	.00	4	4	4	.00	.00	.00	.00	7	3	2
Archer	Madison	1.00	nd	1.00	.00	2	nd	1	nd	.00	nd	.	nd	2	nd
Union-Lyman	Madison	.75	.75	1.00	.25	8	4	11	.40	.13	.40	.00	5	8	5
Acequia	Minidoka	1.00	1.00	.71	-.29	3	8	17	.24	.35	.29	.05	21	17	7
Heyburn	Minidoka	1.00	.84	.91	-.09	3	32	34	1.00	.16	.21	-.79	1	19	39
Paul	Minidoka	.86	.96	.82	-.04	22	28	33	.25	.23	.26	.01	28	22	31
East	Mtn. Home	.91	.91	.92	.01	11	11	12	.33	.09	.20	-.13	15	11	15
West	Mtn. Home	.89	.81	.92	.03	28	26	24	.42	.41	.36	-.06	31	27	25
Sherman	Nampa	.75	.79	.85	.10	12	34	35	.00	.21	.29	.29	2	42	31
Snake River	Nampa	.57	.70	.92	.35	14	43	36	.40	.36	.27	-.13	10	36	45
New Plymouth	New Plymouth	.89	.89	.50	-.39	9	9	11	.08	.00	1.00	.92	13	6	8
Bickel	Twin Falls	1.00	1.00	nd	.	6	3	Nd	.29	.14	nd	.	7	7	nd
Harrison	Twin Falls	.77	.86	1.00	.23	13	14	16	.61	.42	.53	-.08	18	12	15
Lincoln	Twin Falls	1.00	1.00	1.00	.00	7	14	11	.67	.54	.38	-.29	9	13	13
Oregon Trail	Twin Falls	.71	.87	.89	.18	24	31	28	.32	.35	.30	-.02	28	23	30

92

Central Canyon	Vallivue	.33	.77	.88	.55	3	56	69	.33	.30	.22	-.11	3	54	69
East Canyon	Vallivue	1.00	.86	.86	-.14	2	49	37	.33	.28	.16	-.17	3	43	50
West Canyon	Vallivue	1.00	.74	.71	-.29	2	23	24	.00	.17	.41	.41	1	18	22
Wendell	Wendell	.93	1.00	.93	.00	30	39	40	.39	.38	.43	.04	28	37	40
Priest River	West Bonner	1.00	1.00	nd	.	1	1	Nd	nd	.00	.00	.	nd	1	1
Averages		.85	.87	.88	.05				.34	.23	.30	-.03			

Hispanic 2nd graders are performing quite well. During spring 2007 Hispanic 2nd graders lagged behind the entire group of students by only 2% and behind White students by about 5%. White students averaged 93% being placed in the highest story. Furthermore, the three-year trend for overall averages is slightly upward. In 2005, 85% of Hispanic 2nd graders were placed in the highest story. That percentage went up to 87% in 2006 and 88% in 2007. Granted these gains are quite small and the variability in these scores can be substantial so the trend is not confirmed yet, but there is evidence of an emerging positive trend.

Individual schools have also shown positive improvements. Harwood, Harrison, and Oregon Trail appear to have made substantial gains in Hispanic 2nd grade scores. Conversely, no school shows a consistent downward trend, but Popplewell experienced a substantial drop in scores that if not reversed could be problematic. Other schools show sideways trends. Some of these schools have room to improve their scores but others have nearly 90% or more of their students scoring at the most difficult story level and probably can't improve their scores all that much because of test ceiling effects. This is truly outstanding performance. East, West, Lincoln and Wendell are examples of this, with Wendell being the star of this group given that they consistently tested large number of Hispanic students and held their percentage at or above 93%. An example of a school that trends sideways and yet has room to grow is probably Paul.

Concerning 2nd grade scores, the test appears to be perhaps too easy. More students in IRF schools should be administered the supplemental word list the publisher provides to address this potential problem.

Hispanic 3rd graders present a more complex case. In 2005 the proportion of Hispanic 3rd graders who were placed in either of the highest two stories was the same as for the overall group, that is 34%, but at the end of 2006 the Hispanic percentage had dropped to 23% while the overall group held up at 32%. At the end of 2007, 30% of Hispanic students were placed in either of the two most difficult stories while 36% of the overall group was. During 2007, 38% of White students were in the two most difficult stories, representing an 8% achievement gap. In short, the trend is not clear, but the achievement gap is substantial and shows little evidence of closing. It is also important to point out that the percentage of students being placed in the top two stories is quite low for all students, including Hispanics.

With few exceptions individual schools are not showing consistent growth. Butte View is the only school that has tested consistent numbers of students across the years and has sustained about a 10% rise in scores for the past two years. Other schools have shown varying degrees of growth but the patterns are not clear and the numbers of students tested has been quite variable. A number of schools show emerging sideways trends, including Lewis and Clark, Paul, West, Oregon Trail and Wendell. Homedale shows a quite consistent drop in scores over the three years even after having consistently tested high numbers of students. Roberts also shows a precipitous decline although they do not test as many students each year, which will give rise to much greater variability in scores.

In closing, all of the statistics for Hispanics across grades K-3 are somewhat tentative because there is only three years of data, which is not enough to establish a trend, and the sample sizes are quite small in some cases and irregular in others. But if these trends hold over another year, or better yet another two years, while sample sizes stabilize, then examination of specific schools showing strong progress and those showing no progress or negative progress will be in order to explore what is working and not working in IRF schools for Hispanic students.

Idaho Reading Indicator (IRI): Longitudinal Analysis

The IRI is administered three times per year in grades K-3. The number and types of activities on the instrument change with each administration to reflect typical literacy development in children. The IRI is an individually administered screener that takes about 10 minutes for each child. A total points score is calculated and then converted into an ordinal value. The ordinal scale ranges from 1-3, with 1 being below grade level reading, 2 being near grade level reading, and 3 being at or above grade level reading. Level 3 is considered proficient and levels 1 and 2 are not proficient. The State of Idaho established a benchmark that at least 85% of all 3rd graders would score a 3 on the spring 2006 IRI. All of the following tables present the percentage of children at level 3. The data contained in the following tables was compiled from the Idaho Reading First database administered by Databases Done Right. A CD ROM containing 2006-2007 IRI scores was provided by Databases Done Right.

The Idaho Reading Indicator (IRI) is given in grades K-3 three times per year: fall, winter and spring. Level 3 on the spring IRI means that the student is prepared to read grade level material appropriate for the beginning of the next grade. Table S4-13 shows three- and four-year trends in spring IRI scores for all students in IRF schools. The "COH" column denotes the IRF cohort to which a school belongs. A "1" in the column means the school started IRF in 2003-2004, and a "2" means the school started IRF in 2004-2005. The bold 2 digit numbers under the grade level designations stand for the years, and the numbers in the cells of the year columns represent the percentage of students at level 3 on the spring IRI. The "G" headings represent gain scores. How these were computed will be discussed below. The far right column, "Avg. Gain," is the average gain score across all the grade levels for a school. The acronym "nd" stands for no data. It means that no data was received from that school for the particular grade level. The acronym "na" stands for not applicable. These are found in the "G" and "Avg. Gain" columns and mean that computing an average percentage was not possible because of missing data.

Table S4-13: Spring Idaho Reading Indicator Scores by School, Grade and Year: Percent of all Students at Level 3

School	District	COH	Kindergarten					First Grade					Second Grade					Third Grade					Avg. Gain
			04	05	06	07	G	04	05	06	07	G	04	05	06	07	G	04	05	06	07	G	
Popplewell	Buhl	2	85	67	84	69	2	74	64	69	64	0	65	66	72	68	2	65	47	58	48	1	1.3
Lewis & Clark	Caldwell	2	.	94	94	93	-1	.	74	83	nd	Na	.	75	73	nd	na	.	66	68	nd	na	na
Sacajawea	Caldwell	1	90	82	76	nd	Na	64	74	63	75	11	68	57	57	56	-12	57	47	45	63	6	1.7
Wilson	Caldwell	1	68	87	85	87	19	59	63	90	61	2	57	62	67	74	17	68	67	81	65	-3	8.8
Butte View	Emmett	2	79	76	82	81	5	64	71	68	84	13	76	72	81	81	9	67	70	69	82	12	9.8
Filer	Filer	2	91	86	89	86	0	74	75	73	73	-2	76	53	74	78	25	88	88	71	87	-1	5.5
Gooding	Gooding	1	65	83	59	77	12	65	78	69	82	17	54	67	83	71	17	64	77	69	81	17	15.8
Homedale	Homedale	2	63	86	89	90	4	57	70	56	80	10	71	62	75	65	3	57	54	52	56	2	4.8
Harwood	Jefferson	1	72	73	72	nd	na	69	66	62	70	1	68	69	63	65	-3	60	60	61	50	-10	-4.0
Roberts	Jefferson	1	76	61	79	60	-16	91	74	78	93	2	74	79	71	38	-36	75	66	84	67	-8	-14.5
Adams	Madison	1	88	94	92	93	5	78	72	77	90	12	76	80	80	85	9	72	74	69	68	-4	5.5
Archer	Madison	1	88	87	95	93	5	83	75	100	87	4	88	76	71	91	3	62	87	68	67	5	4.3
Union-Lyman	Madison	1	78	98	100	96	18	66	83	98	100	34	62	69	68	89	27	59	73	69	68	9	22.0
Acequia	Minidoka	1	79	98	82	80	1	78	61	58	83	5	53	68	59	52	-1	49	45	49	51	2	1.8
Heyburn	Minidoka	2	72	87	73	83	-4	45	53	64	63	10	47	62	58	57	-5	43	41	49	53	12	3.3
Paul	Minidoka	1	79	82	78	75	-4	71	66	69	64	-7	62	75	73	67	5	67	43	58	43	-24	-7.5
East	Mountain Home	1	85	86	85	94	9	75	75	86	81	6	69	79	89	78	9	64	69	66	68	4	7.0
West	Mountain Home	1	72	82	83	84	12	62	75	80	65	3	62	74	70	62	0	56	69	73	73	17	8.0
Sherman	Nampa	1	74	68	78	73	-1	47	60	68	47	0	50	67	59	58	8	45	41	49	46	1	2.0
Snake River	Nampa	1	55	64	85	78	23	46	54	72	69	23	50	47	58	65	15	43	49	60	45	2	15.8
New Plymouth	New Plymouth	1	70	90	95	88	18	86	82	80	92	6	76	75	75	72	-4	69	61	75	76	7	6.8
Bickel	Twin Falls	1	77	86	80	47	-30	87	84	77	56	-31	78	77	79	73	-5	85	76	76	nd	na	-22.0
Harrison	Twin Falls	2	79	90	78	85	-5	68	81	83	66	-15	88	69	73	81	12	86	77	76	72	-5	-3.3
Lincoln	Twin Falls	2	83	81	82	94	13	81	82	69	69	-13	71	77	68	65	-12	79	81	58	64	-17	-7.3
Oregon Trail	Twin Falls	1	93	95	80	91	-2	84	89	86	74	-10	73	83	85	76	3	72	81	80	69	-3	-3.0
Central Canyon	Vallivue	2	83	83	80	63	-20	73	65	66	57	-12	65	66	45	56	-10	43	51	54	42	-9	-12.8
East Canyon	Vallivue	1	77	88	90	84	7	52	57	52	60	8	54	53	56	51	-3	51	48	51	49	-2	2.5
West Canyon	Vallivue	1	74	90	80	77	3	62	72	69	66	4	39	64	65	59	20	48	49	85	58	10	9.3
Wendell	Wendell	1	80	85	92	89	9	78	82	67	58	-20	76	76	78	76	0	77	71	78	80	3	-2.0
Priest River	West Bonner	2	71	80	73	80	0	60	72	68	52	-20	74	59	58	62	3	52	54	57	64	10	-1.8
Averages			77	84	83	79	3.0	69	72	73	69	1.4	66	69	69	69	3.3	63	63	65	63	1.2	1.9

96

An explanation of how the gain scores and averages were computed is in order. If a school was in Cohort 1, then the gain scores were computed by subtracting the percent at level 3 in 2004 from the percent at level 3 in 2007. If the school was in Cohort 2, then gain scores were computed by subtracting the percent at level 3 in 2005 from the percent at level 3 in 2007. The average gain scores in the far right column were computed by summing all of the four gain scores for a school and dividing by four. This average should be cautiously interpreted, however, since computing an average from only four scores is problematic because of the possibility of extreme values.

A note of caution is in order concerning interpreting the gain scores. They can be misleading at times because the scores at a given grade level in a given school oftentimes vary quite a bit from year to year. If there were clear and consistent trends of increasing scores from the time a school entered IRF to the present, then the simple subtraction of the first year score from the current year score would be an excellent indicator for how far the school has grown, but such is not the case. Scores within grade levels within schools can vary markedly from year to year, so simple subtraction like that done to compute the gain scores can be misleading both on the positive side and the negative side. In other words, schools at individual grade levels can show negative gains when the actual trend is either flat or too variable to ascertain a trend. The opposite can occur also. Namely a school can show positive growth at a grade level when the trend is not clear. Readers are urged to study the table carefully before drawing conclusions.

On average, IRF schools did not show appreciable gains in IRI scores over the three or four years they have been in the initiative. Looking across the bottom row in Table S4-13 at the average scores for each year within a grade level reveals that scores are quite consistent across the years within a given grade level. Kindergarten classrooms across the initiative appear to be able to consistently achieve spring IRI percentages at level 3 in the high 70s to low 80s. First grade classrooms hover in the high 60s to low 70s. Second grade classrooms are in the high 60s, and 3rd grade classrooms in the mid 60s. No grade level shows a consistent upward trend in scores over the years. A matter of fact, initiative-wide scores within the individual grade levels have pretty much remained at the point where they were at the start of the initiative.

Looking at individual schools within Table S4-13 reveals substantial variability in performance within and across schools. This is no different from what was found in previous years, but now there is an additional year of data to better confirm or disconfirm the presence of trends. A close look at two schools will illustrate this. Wilson in Caldwell and Butte View in Emmett will be discussed.

Wilson has two grade levels, kindergarten and Grade 2, that appear to be performing quite well on the IRI. After the first year of participation in IRF, Wilson's kindergarten teachers averaged 68% at level 3 on the spring IRI. This percentage jumped to the mid 80s at the end of their second year in IRF and has held at that level for the past three years. This is stellar performance and is a model for what other schools and grade levels should strive for. Wilson's 2nd grade also provides an excellent model, albeit a different

one from their kindergarten. The 2nd grade percentage started at 57% after the first year of the initiative. They have consistently stair-stepped upward five or more percentage points each subsequent year. They currently have 74% of their 2nd graders scoring a 3 on the spring IRI, and after four years there is increasing evidence that this trend upward may hold for two or three more years until the ceiling is reached regarding the number of children who can be brought to level 3. But Wilson's other two grade levels, Grades 1 and 3, do not show such consistently positive trends. Grade 1 went from 59% at the end of the first year in the initiative to 63% at the end of year two, a solid increase that was then dramatically built upon when at the end of the third year of the initiative the percentage jumped to 90%. But the spring 2007 percentage dropped back to 61%. This precipitous rise and fall over the course of three years could be an aberration. If so, then Wilson's 1st grade hasn't improved over the course of their participation in IRF. If the 61% achieved during spring 2007 is an aberration then perhaps their performance has improved. A similar trend is found in the percentages of Wilson's Grade 3. At the end of year three in the initiative, their percentage increased to 81%, a very strong performance on the challenging Grade 3 spring IRI. But during spring 2007, their percentage dropped back to 61%, about where it had been at the end of the first two years of the initiative.

Wilson's inconsistency across the grade levels, that is some grade levels show consistent positive growth while others less so, is quite common in Table S4-13. There are also schools that show no consistent growth at any grade level. Paul in Minidoka is an example of this. Grades K through 2 have basically remained the same throughout their participation in the initiative. Their Grade 3, however, represents a more troublesome phenomenon. The trend is downward, meaning that the average percentage of Grade 3 students at level 3 on the spring IRI has decreased or is decreasing. We now turn to the discussion of Butte View in Emmett.

Butte View is a Cohort 2 school, meaning that they started in the initiative in 2004-2005. A scan across their row of percentages shows that their kindergarten was performing quite well the year before they began and then continued to do so. Grades 1 through 3 also show interesting trends. It appears that after a year or two of base building, all of these grades have increased to having 80-84% of their students at level 3. If these numbers hold for two or three more years, Butte View will have not only achieved, but will have sustained superior performance.

As was mentioned in the introduction to this section on IRI scores, the State of Idaho has set an 85% criterion for the spring Grade 3 IRI. Although no IRF school consistently meets or surpasses this criterion, it is important to note that one school surpassed this level and three others were in the low 80s. The Grade 3 spring IRI is quite challenging and to have 13% of IRF schools meeting or close to meeting this benchmark is a positive outcome.

Finally, a few more schools will be mentioned and briefly discussed to further illustrate how Table S4-13 can be studied for insights. Union Lyman has shown excellent growth across the grade levels. If they can sustain the percentages they achieved in spring 2007 in grades K-2 and both accelerate and then sustain the growth in their Grade 3, they will

be well positioned for the future. Snake River is another school that is showing good growth in grades K-2, although their percentages are not as high as some of the other schools. But like a number of other schools, Snake River appears to be struggling some at Grade 3 since the percentages remain relatively low across the years. East Canyon is a school that appears to be stagnated. Their kindergarten percentages are high and relatively consistent and are therefore cause for celebration, but their other three grades appear to be going nowhere and their percentages are low. West Canyon was highlighted in last year's report as a school that showed consistent improvement, but this year with an additional year of data, the trends are clearer and the school doesn't stand out as much. This underscores the highly variable nature of test scores and also the highly variable nature of schools' performance over time. In the case of West Canyon, their kindergarten appears to be functioning in the high 70s to low 80s, which represents respectable performance. Their Grade 1 functions in the mid to high 60s; their Grade 2 in the low to mid 60s, and their Grade 3 in the 50s or 60s. Grade 3 is a difficult trend to ascertain because of the large increase and then pull back. But all of these estimates could be in error. Any or all grade levels at West Canyon could break out next year and continue consistently moving up or for that matter down. What this underscores is the great degree of variability in IRI scores in IRF schools.

In summary, performance across IRF schools varies widely. If the goal of having all children read at grade level by the end of 3rd grade is to be realized, then all IRF schools are going to have to improve their performance and strive to make their performance more consistent. We now turn to a discussion of Hispanic student scores on the IRI.

Hispanics historically score lower on standardized measures of reading achievement so it is important to look at this subgroup to ascertain the gains being made over time. Table S4-14 reports the results for Hispanic students over the four years IRF has been in operation. The table is interpreted in the same way as Table S4-13 above. Some cells in the table are blank because the schools have no Hispanic students at that grade level.

Table S4-14: Spring Idaho Reading Indicator Scores for Hispanics by School, Grade and Year: Percent of Students at Level 3

School	District	COH	Kindergarten					First Grade					Second Grade					Third Grade					Avg. Gain
			04	05	06	07	G	04	05	06	07	G	04	05	06	07	G	04	05	06	07	G	
Popplewell	Buhl	2	81	50	.	50	0	57	36	.	56	20	50	48	.	47	-1	44	36	.	23	-13	1.5
Lewis & Clark	Caldwell	2	.	92	72	90	-2	.	72	75	nd	.	.	74	56	nd	.	.	61	43	nd	.	.
Sacajawea	Caldwell	1	90	80	75	nd	.	66	76	53	77	11	62	57	53	54	-8	45	45	59	60	15	6.0
Wilson	Caldwell	1	59	82	89	84	25	55	43	80	60	5	43	60	68	63	20	60	56	60	56	-4	11.5
Butte View	Emmett	2	64	77	76	74	-3	53	59	58	75	16	72	68	55	71	3	70	70	44	80	10	6.5
Filer	Filer	2	92	83	80	69	-14	.	93	83	71	-22	.	.	58	78	.	.	.	.	71	.	.
Gooding	Gooding	1	52	74	86	70	18	63	67	37	81	18	63	70	47	64	1	52	65	47	72	20	14.3
Homedale	Homedale	2	47	93	64	85	-8	38	60	65	74	14	60	45	51	52	7	41	48	37	46	-2	2.8
Harwood	Jefferson	1	.	92	74	nd	.	50	47	58	33	-17	.	45	48	42	.	54	.	37	11	-43	.
Roberts	Jefferson	1	67	.	74	72	5	83	70	100	82	-1	80	75	77	30	-50	.	56	60	40	.	.
Adams	Madison	1	.	.	82	.	.	.	.	70	83	.	.	.	69	50	.	.	.	70	0	.	.
Archer	Madison	1	.	.	64	50	.	.	.	59	.	.	.	.	81	.	.	.	.	60	.	.	.
Union-Lyman	Madison	1	.	.	38	.	.	55	.	56	.	.	.	.	84	82	.	.	.	67	20	.	.
Acequia	Minidoka	1	82	100	92	76	-6	65	23	64	70	5	38	53	70	35	-3	37	19	75	14	-23	-6.8
Heyburn	Minidoka	2	75	91	46	94	3	37	48	.	53	5	59	59	40	47	-12	26	52	41	41	-11	-3.8
Paul	Minidoka	1	78	69	77	81	3	58	60	60	58	0	50	67	70	56	6	68	33	79	45	-23	-3.5
East	Mountain Home	1	.	.	.	.	.	50	58	.	61	11	65	.	.	91	26	.	67	.	63	.	.
West	Mountain Home	1	64	74	.	79	15	52	69	.	67	15	60	67	.	52	-8	50	60	.	69	19	10.3
Sherman	Nampa	1	51	52	.	73	22	43	52	.	43	0	32	68	.	53	21	25	27	.	39	14	14.3
Snake River	Nampa	1	39	42	.	61	22	35	37	35	61	26	42	40	.	53	11	40	47	47	40	0	14.8
New Plymouth	New Plymouth	1	60	.	73	50	-10	.	77	55	.	.	77	.	52	45	-32	.	67	45	71	.	.
Bickel	Twin Falls	1	.	.	81	30	.	.	.	64	33	.	.	.	69	67	.	.	.	39	.	.	.
Harrison	Twin Falls	2	75	.	84	72	.	56	86	48	64	-22	89	57	64	75	18	.	89	38	73	-16	-6.7
Lincoln	Twin Falls	2	83	77	.	93	16	.	89	.	60	-29	50	.	70	55	.	78	.	.	62	.	.
Oregon Trail	Twin Falls	1	85	92	.	91	6	52	73	.	73	21	64	52	.	64	0	59	66	.	55	-4	5.8
Central Canyon	Vallivue	2	76	67	83	52	-15	62	58	82	47	-11	54	50	62	54	4	43	40	77	35	-5	-6.8
East Canyon	Vallivue	1	84	78	76	81	-3	45	46	50	40	-5	45	52	64	38	-7	20	41	85	41	21	1.5
West Canyon	Vallivue	1	56	75	69	70	14	50	60	75	47	-3	33	53	73	42	9	48	30	52	50	2	5.5
Wendell	Wendell	1	86	79	68	85	-1	78	86	48	45	-33	65	82	52	80	15	74	77	45	84	10	-2.3
Priest River	West Bonner	2	.	.	.	.	.	.	.	64	.	.	.	.	93	.	.	.	.	.	.	.	.
Averages			70	78	74	76	4.1	55	61	64	63	1.0	57	59	64	57	1.0	49	52	55	48	-2.0	3.6

An explanation of how the gain scores and averages were computed is in order. If a school was in Cohort 1, then the gain scores were computed by subtracting the percent at level 3 in 2004 from the percent at level 3 in 2007. If the school was in Cohort 2, then gain scores were computed by subtracting the percent at level 3 in 2005 from the percent at level 3 in 2007. The average gain scores in the far right column were computed by summing the four gain scores for a school and dividing by four. If the school had only three of four gain scores then the average was computed with the three. If the school had less than three gain scores, then the average was not computed. The same caution that was made above about gains scores holds here also.

Because of the large number of missing values in Table S4-14 and the large amount of variability within some individual schools, the most reliable statistics are probably the grade level by year averages found in the bottom row of the table. This row shows a mixture of trends for the four years of IRF. Kindergarten has consistently remained in the mid to high 70s. Grade 1 appears to hold in the low 60's; Grade 2 in the high 50s; and Grade 3 in the high 40s to low 50s. No grade level shows a consistent continuing trend upward, but kindergarten and Grade 1 have sustained higher scores after the initial year of IRF. Grades 2 and 3 went up from the initial year and sustained the increases for two years but in 2007 fell back to their initial levels.

Discussing individual schools is problematic because of the high degree of variability in the data at the individual school level, but a few will be mentioned to again illustrate how the table can be interpreted for insights into how schools are performing over time. Paul has relatively high scores for Hispanics. The scores at Grades K and 1 are quite consistent, but there is less consistency in Grade 2 and the Grade 3 scores are quite erratic. Wendell also has some high relative scores, but like Paul, the scores exhibit quite a degree of inconsistency. This is especially true at Grade 1 where there was a precipitous drop after year 2 in IRF that has continued for two years. West Canyon was mentioned in last year's report as a school doing well with Hispanic students, but with inclusion of 2007 data the trends are not as positive. Although kindergarten continues to perform consistently quite well, the positive trends in Grades 1 and 2 were not sustained for 2007. Perhaps these pull backs are aberrations and the upward trends will start again next year. West Canyon's Grade 3 appears to be stagnated around the high 40s and low 50s.

A final comparison will be made between Hispanic student performance and White student performance. Table S4-15 presents overall averages by year and grade for White and Hispanic students.

Table S4-15: Spring Idaho Reading Indicator Scores for White and Hispanic Students by Grade and Year

Year	Kindergarten				First Grade				Second Grade				Third Grade			
	04	05	06	07	04	05	06	07	04	05	06	07	04	05	06	07
White	80	87	86	85	74	76	77	74	70	72	71	71	67	66	68	67
Hispanic	70	78	74	75	55	61	64	60	57	59	64	56	49	52	55	50
Difference	10	9	12	10	19	15	13	14	13	13	7	15	18	14	13	17

White kindergarten students appear to have gained about five points from the first year of the initiative to the present. This five-point gain occurred in 2005 and has been sustained ever since. This is very positive performance and perhaps represents scores at or near the performance ceiling, meaning that because of measurement error and children who are not capable of obtaining a 3 on the IRI scores will not go much higher in the future. For Whites at the other grade levels, their performance is very consistent, showing no growth or declines over time. Hispanic student performance paints somewhat of a different picture. For Hispanics both kindergarten and Grade 1 show growth after the first year of IRF implementation that has been sustained for the subsequent three years. Both grades appear to have gained about 5% or perhaps a bit more. Grades 2 and 3 show more variability than their White counterparts, but overall these two grade levels have about the same achievement levels that they had at the outset of the initiative.

An important question that Table S4-15 helps answer is if the achievement gap between Whites and Hispanics is closing. The row labeled "Difference" provides this information. The answer is "no" in grades K, 2 and 3. In these grades the differences between White and Hispanic achievement has stayed relatively constant across the four years. Granted the differences fluctuate up and down from year to year but overall there are no clear trends of decreasing differences at these grade levels. This is not true for Grade 1. Grade 1 experienced a drop from 19 to 15 points difference between 2004 and 2005. The difference score has remained at 13 and 14 for the two subsequent years. Thus, there is emerging evidence that at Grade 1 the achievement gap was initially reduced but has not been reduced further since this initial drop in the difference score.

In closing, much variability in Hispanic IRI scores remains and no school appears to be consistently performing well with their Hispanic students across all grade levels and years. Additionally, the achievement gap is not consistently closing. One thing needs to be underscored, however. Fluctuations in test scores are understandable given the measurement error in the test and changing characteristics of the students tested. Additionally, the smaller number of Hispanic students in schools relative to the number of White students causes statistics for this subgroup to be even more volatile since small sample sizes give rise to greater changes in statistics from year to year.

Conclusions

CONCLUSIONS

Overall, Idaho Reading First (IRF) is having varied effects on participating schools. On the positive side, the average derived scores (i.e., grade equivalent scores and normal curve equivalent scores) are at the national average. This represents excellent performance for impoverished schools. Additionally, IRF schools average about 65% of their students across the grades scoring at or above the 40th percentile on the ITBS. Again, this is excellent performance. Finally, there are individual schools that have made substantial improvements at some grade levels since entering the initiative. On the negative side, however, some IRF schools' test scores may have stalled or regressed. No school has made consistent, sustained progress on all measures at all grade levels since entering the initiative. Overall averages for the entire IRF network have not changed. Achievement gaps between Whites and Hispanics are not closing, and test score performance remains quite variable across IRF schools. Following are additional conclusions:

- High-performing teachers in IRF schools are committed educators who work hard to deliver high-quality curriculum and instruction so that all of their students have the greatest probability of success. These teachers teach the core program with fidelity but also have intimate knowledge of its strengths and weaknesses derived from using the program with intensity and focus for several years. They spend considerable time outside of school working to remediate the weaknesses in the core programs while building on the strengths. They are also constantly searching for new ideas, insights and techniques that will help them become better reading teachers. Inside their classrooms they teach with humor, energy and a strong commitment to the goal that all students will learn. They have excellent classroom management and work hard each day to make the most of every teachable moment so that their students constantly make incremental progress. They use workshop and intervention time to customize instruction for each child and they spend considerable time developing meaningful materials and activities so that workshop and intervention are effective components of the school day;

- Student mobility remains high in some IRF buildings. This is an uncontrollable factor that IRF schools will have to live with while working to develop effective intervention programming;

- Teacher turnover remains a significant but perhaps controllable factor in IRF buildings. Principal turnover is not as widespread as teacher turnover; and

- Special education referrals are trending downward in some grade levels and IRF cohorts. There is no evidence of increases in special education referrals. The quality of this data remains suspect, however.

In summary, Idaho Reading First continues to actively work to solve the reading challenges facing high-poverty schools. The percentage of children not reading at grade level is still high in some schools and test scores are not moving consistently upward across the years, but there are grade levels in individual schools that have made significant progress. Thus, what remains a challenge for IRF is to reduce the variability

both within and between schools. As a consequence of this evaluation and earlier evaluations, variability is better understood than it has been in the past. Based on the findings, the key to reducing variability is improving the teaching work force so that all teachers in IRF buildings share the characteristics of the highest-performing teachers.

RECOMMENDATIONS

RECOMMENDATIONS

Following are recommendations distilled from all sections of the report. They are organized under the specific headings of areas needing immediate and continued attention and areas needing attention in future evaluations.

Areas Needing Immediate and Continued Attention

- Longitudinal test score analyses reveal that Hispanic students are not making gains, and large numbers of Hispanic students continue to not meet Idaho proficiency criteria. Regression analyses conducted during the 2005-2006 evaluation revealed strong predictive power of the proportion of Hispanic students in a school, especially for ITBS scores. The consistent improvement in this subpopulation's performance is thus critically important to any school improvement effort. Consequently, an important criteria used to identify high-performing teachers who were intensively studied this year was strong relative success with Hispanic students in their classrooms. Studying the high-performing teachers revealed common characteristics that are important to their success with Hispanic students: (a) high expectations for all students and a work ethic to diligently pursue this goal day in and day out, (b) an enhanced focus on vocabulary throughout the lesson cycle each day, (c) high-quality supplemental programming such as ELL/ESL programs, and (d) effective workshop/intervention periods where individual needs of the students could be consistently addressed;

- The persistent lack of consistent test score growth within and across schools and the persistent large degree of variability in test score performance within and across schools should be immediately addressed. The combined results of the previous evaluations with this one point to some quite salient variables important to student success. Most importantly, the quality of the teaching work force is of paramount importance. Continued staff development informed by the results of these evaluations and the insights from other states that have addressed similar problems in their IRF networks is extremely important if gains are to be realized by the conclusion of the program in Idaho;

- Teacher and principal mobility in IRF schools needs to be reduced to an absolute minimum so staff experience and training levels can be as consistent as possible. It is difficult to implement a program as sophisticated and far reaching as IRF when teacher turnover is high;

- High student mobility is something that some IRF schools will probably have to live with for some time to come, but steps should be taken to ameliorate the potential negative impact this can have on school performance. Active and dynamic new student induction programs that are currently operating in some IRF schools should be the norm for all IRF schools;

- Some high-performing teachers are under much job-related stress. IRF schools losing their best exacerbates the teacher turnover problem. These high-performing teachers need to be recognized within their buildings for what they are: outstanding educators who not only have a lot to offer their students but also other teachers who can benefit from the sharing of best practices and enthusiasm for all children's learning;

- A Web-based clearinghouse for best practices needs to be developed. The highest- performing teachers across the network have extensive knowledge about how to leverage the strengths and remediate the weaknesses of their core programs and the IRF literacy framework. This clearinghouse could provide ideas for lessons, teacher insights focused on curriculum and instruction, specific interventions that worked for particular types of children, ways to read and use data, etc. Areas that need to be showcased in the clearinghouse are vocabulary and fluency;

- Teachers should be included in calibration visits. The high-performing teachers expressed a strong desire to observe other teachers and share best practices and insights. The amount of institutional knowledge and expertise that has accumulated as a consequence of IRF is quite striking. Mechanisms need to put in place whereby this body of knowledge can be shared. The clearinghouse mentioned above is one way, but face-to-face meetings and classroom visits between teachers were also emphasized by high-performing teachers;

- Based on the results of this and previous evaluations, much information about what are best practices in IRF classroom has accumulated. It is time for the very best IRF teachers to be given permission to experiment, while maintaining fidelity to the core program, to see how they can impact student achievement. Their experimentation might result in increased test scores. If this were to occur, then these practices should become part of the clearinghouse and other mechanisms for disseminating information. What this recommendation boils down to is customized oversight at the individual teacher level within IRF buildings. Those teachers doing well can be given more freedom to experiment to improve student achievement while others are provided continued support to maintain fidelity to the program and improve their overall teaching;

- There are outstanding teachers all across the IRF network. Developing a cadre of these teachers to travel from school-to-school to do demonstration lessons and to discuss curricular and instructional issues with their IRF colleagues could serve as highly effective staff development.

- If cross-school sharing and collaboration are not possible, then facilitating sharing across grade levels in individual buildings could be an alternate possibility. As the test score data analyses consistently showed, some grade levels within individual IRF buildings have done quite well. The best practices within these grade levels should be systematically shared with the other grade levels in the building. Where this has already occurred and the best practices have not been incorporated at all the other grade levels or the best practices have not resulted in

higher student achievement, then leadership needs to step in to ascertain the root cause of these outcomes.

Areas Needing Attention in Future Evaluations

- IRF leaders should consider an expanded evaluation design for the remaining years of this initiative. This design might include continued, extensive classroom observations with individual teacher interviews, but it could also include focus groups of teachers, parents and students. Additionally, looking beyond the IRF network to schools outside of the network and perhaps even outside of Idaho which have had success might provide insights as to how to proceed and accelerate school growth;

- A possible bright spot of significant importance is the impact on special education in IRF schools. There are hints in the current qualitative and quantitative data that referral rates and perhaps even the number of students who are placed in special education are down. Special educators in IRF schools and special education directors in IRF districts need to be systematically contacted to ascertain the most appropriate data to be collected that is widely available in IRF districts. Once this data is identified, then it needs to be collected from all IRF schools, including at least three years of baseline data. Cost/benefit analyses should be conducted to show the impact of IRF, and special educators should be surveyed and interviewed for their insights about the impact IRF programming has had on special education; and

- Explore more thoroughly the role of building and sustaining leadership in creating effective, high-performing IRF schools.

LIMITATIONS

LIMITATIONS

All evaluation designs have limitations that need to be mentioned so readers can make appropriate interpretations concerning the quality of the data, findings and conclusions. Limitations of the 2006-2007 evaluation are listed below:

- The number of schools reporting special education referral data was low last year (22 of 30) and even lower this year (10 of 30). This undoubtedly impacts the reliability and validity of this data and the conclusions derived from it;

- The reliability and validity of instruments is always less than perfect. In the case of the IRI, high-quality reliability statistics are not available for the instrument. And in the case of the TPRI, reliability coefficients are available but since it is an individually administered test that is administered by different people at each IRF school, reliability may be compromised. The same holds true for the IRI since it, too, is an individually administered instrument;

- Teachers provided only their perceptions of important variables to school success when responding to survey questions or discussing with an observer/interviewer. Thus, the reality in the buildings could be quite different;

- Classroom observers could have been biased regarding Idaho Reading First, thus influencing their interpretations and descriptions of classroom events; and

- Classroom observations and interviews were conducted in 15 of 30 IRF schools. Despite efforts to draw a representative sample of teachers and schools, it is possible that sample bias occurred, and thus the findings and conclusions will not generalize to all IRF schools.

Appendix

APPENDIX A

Idaho Reading First Best Practices Synthesis
Teacher Interview Questions

Introduction:

You have been identified as a highly successful Reading First teacher. We would like to find out what you attribute your success to, so we have asked to talk with you today. We will also be interviewing your reading coach and principal to get their perspectives on why you are successful. Thank you for your time.

You will not be identified by name or school in the final evaluation report submitted to the state next fall. No one will know you have been interviewed except your principal and coach, and they will not know what you said since we do not share the contents of these interviews with anyone. We have always conducted these evaluations in a confidential manner since we want teachers to feel free to respond in any way they choose.

You will be paid $75.00 for your participation in the interview. I have some paper work for you to complete in order for you to be paid and we will do that at the conclusion of the 60-90 minutes we will spend talking.

You are participating in what is called a **best practices synthesis.** We are interviewing 25 Reading First teachers—about 6 per grade level – and their coaches and principals. The teachers are in about 15 of the 30 Idaho Reading First schools. The 25 teachers are some of the most successful based on test scores out of the roughly 400 K-3 teachers in Idaho Reading First schools. What we want to do is find out what the most successful teachers are doing and disseminate that information to the rest of the Reading First teachers.

The motivation behind this project came from some of the results from last year's evaluation. The results showed that test scores in most Idaho Reading First schools are not consistently going up across all grade levels. There are pockets of sustained improvement at certain grade levels in individual schools, but no school consistently grew at every grade level over the years they have been in Reading First. There was also no initiative-wide sustained improvement in test scores. But, some individual teachers are doing quite well. Thus, we thought it important to talk with these teachers to find out what they attribute their success to and then share that with the rest of the Reading First teachers. We also felt it important to talk with the coaches and principals in the buildings where the teachers are for additional insights and perspectives on what makes the particular teacher so successful.

Do you have any questions before we begin? **GET CONSENT FORM SIGNED!!!**
I would like to discuss some terms before we get started.

I will ask you questions about your "core" reading program. What I mean by "core" reading program is the basal reading series your school adopted as part of Reading First.

I will ask you questions about the supplemental materials and programs that you have purchased to augment the core reading program. These are the materials or programs you have purchased to address the specific needs of subgroups of children you have identified needing something more or different from what the core program provides.

I will also ask you questions about your overall reading program. When I use the term overall reading program, I mean all the elements that go into your K-3 reading program including all of the materials you use (i.e., core reading series and supplemental materials and programs); the daily schedule; the supplemental services provided by Title I, ELL, and special education; your role as principal; and the coach's role.

I'll now begin with some general background questions concerning your experience as a teacher.

1. Name of Interviewee: _____

Circle One: Teacher   Coach   Principal

2. School: _____

3. Grade Level Taught: _____

4. How many years have you been a teacher? _____

5. How many years have you taught at your current grade level? _____

6. What other grade levels have you taught? _____

7. How many years have you taught in a Reading First school? _____

Now I would like to turn to our discussion of what you attribute your success to:

8. Let's begin by talking about your "core" reading program and the supplemental materials and programs that you use. A big part of Reading First is the adoption of a

"core" reading program that all teachers then use. Also many of the Reading First schools have purchased additional supplemental materials and programs. I would like to talk about the core program and any supplemental materials and programs you have.

a. What core program do you use?

b. What supplemental materials and programs are part of your overall reading program?

c. Please tell me how your core program and supplemental materials and programs influence your success with your students.
    Probes:
- What are the strengths and weaknesses of the core program?
- What are the strengths and weaknesses of the supplemental materials and programs?
- Which elements of these programs are the most important to your success?
- Which elements are not as helpful?
- How important are the supplemental materials and programs you have purchased? With which students?

9. Now that we have discussed the materials that you use to teach reading, let's broaden our discussion to include your overall reading program.

a. Please describe your overall reading program?

b. What elements of your overall program are important to your success?

NOTE: Make sure the probes are answered.

Probes:
- What does a typical day in your classroom when you are teaching reading look like?
- How are students grouped for their literacy instruction?
- Which elements of your overall reading program are the most important to your success?

- What about supplemental services such as your Title I program, ELL program, or special education program? Are they important elements of your success? Please explain.
- How do you use data to inform your literacy instruction?

10. You have been quite successful with your Hispanic students. What do you attribute your success to?

   Probes:
- What components of your "core" reading series are effective with Hispanic students? Which are ineffective?
- What components of your supplemental materials and programs are effective with Hispanic students? Which are ineffective?
- What elements of your overall reading program are effective with Hispanic students? Which are ineffective?
- **If in the above, ELL programming and Title I programming are not discussed, ask about their effectiveness with Hispanics.**

11. How closely do you collaborate with your fellow grade level teachers?

   Probe: Do you have grade level meetings? How often? What do you do at these meetings?

12. How much of your success is attributable to this collaboration? Please explain.

13. <u>IMPORTANT QUESTION:</u> How do you differ from your colleagues?

14. How much do parents contribute to your success? Please explain.

15. What role does your reading coach play in your success?

16. How does your principal contribute to your success?

17. Summary Questions:
   a. Is there anything else that you've thought of as we have been talking that would answer the question "Why do you think you are doing so well with your students?"

b. What are the strengths and weaknesses of Reading First?

c. What does Reading First need to do so that all teachers attain the level of success that you have?

Thank you for your time.

Don't Forget to Complete the Payment Paperwork!!

Idaho Reading First Best Practices Synthesis
Coach Interview Questions

Introduction:

_____ (Teacher's Name) has been identified as a highly successful Reading First teacher. We would like to find out to what you attribute her success, so we have asked to talk with you today. I will also be interviewing _____ (Teacher's Name) to ask her about her program and to what she attributes her success, and _____ (Principal's Name), your principal. Thank you for your time.

You will not be identified by name or school in the final evaluation report submitted to the state next fall. No one will know you have been interviewed except your principal and _____ (Teacher's Name). The contents of this interview will not be released to anyone. We have always conducted these evaluations in a confidential manner since we want teachers and coaches to feel free to respond in any way they choose. The interview will take 30-45 minutes.

You are participating in what is called a **best practices synthesis.** We are interviewing 25 Reading First teachers—about 6 per grade level – and their coaches and principals. The teachers are in about 15 of the 30 Idaho Reading First schools. The 25 teachers are some of the most successful based on test scores out of the roughly 400 K-3 teachers in Idaho Reading First schools. What we want to do is find out what the most successful teachers are doing and disseminate that information to the rest of the Reading First teachers.

The motivation behind this project came from some of the results from last year's evaluation. The results showed that test scores in most Idaho Reading First schools are not consistently going up across all grade levels. There are pockets of sustained improvement at certain grade levels in individual schools, but no school consistently grew at every grade level over the years they have been in Reading First. There was also no initiative-wide sustained improvement in test scores. But, some individual teachers are doing quite well. Thus, we thought it important to talk with these teachers to find out

what they attribute their success to and then share that with the rest of the Reading First teachers. We also felt it important to talk with the coaches and principals in the buildings where the teachers are for additional insights and perspectives on what makes the particular teacher so successful.

Do you have any questions before we begin? **GET CONSENT FORM SIGNED!!!**

I would like to discuss some terms before we get started.

I will ask you questions about your "core" reading program. What I mean by "core" reading program is the basal reading series your school adopted as part of Reading First.

I will ask you questions about the supplemental materials and programs that you have purchased to augment the core reading program. These are the materials or programs you have purchased to address the specific needs of subgroups of children you have identified needing something more or different from what the core program provides.

I will also ask you questions about your overall reading program. When I use the term overall reading program, I mean all the elements that go into your K-3 reading program including all of the materials you use (i.e., core reading series and supplemental materials and programs); the daily schedule; the supplemental services provided by Title I, ELL, and special education; your role as principal; and the coach's role.

I'll now begin with some general background questions concerning your experience as a coach and teacher.

1. Name of Interviewee: _____

Circle One: Teacher   Coach   Principal

2. School: _____

3. How many years have you been a coach?   _____

4. How many years have you been a teacher? _____

5. What grade levels have you taught? _____

6. Did you teach in a Reading First school before becoming a coach? _____

Now I would like to turn to our discussion about what you attribute _____ (Teacher's Name) success to:

7. Let's begin by talking about your "core" reading program and the supplemental materials and programs that you use. A big part of Reading First is the adoption of a "core" reading program that all teachers then use. Also many of the Reading First schools have purchased additional supplemental materials and programs. I would like to talk about the core program and any supplemental materials and programs you have.

    a. What core program do you use?

    Note: "Your" in the following questions refers to the coach, not the targeted teacher. We want to hear from the coach's perspective about the materials being used.

    b. What supplemental materials and programs are part of your overall reading program?

    c. Please tell me how your core program and supplemental materials and programs influence your success with your students.
       Probes:
- What are the strengths and weaknesses of the core program?
- What are the strengths and weaknesses of the supplemental materials and programs?
- Which elements of these programs are the most important to _____ (Teacher's Name) success?
- Which elements are not as helpful to _____ (Teacher's Name)?
- How important are the supplemental materials and programs you have purchased to _____ (Teacher's Name)? With which of her students?

8. Now that we have discussed the materials that you use to teach reading, let's broaden our discussion to include your overall reading program.

    a. Please describe your overall reading program?

    b. What elements of your overall program are important to _____ (Teacher's Name) success?

NOTE: Make sure the probes are answered.
    Probes:
- What does a typical day in _____ (Teacher's Name) classroom look like when she is teaching reading?
- How are students grouped for their literacy instruction in _____(Teacher's Name) classroom?
- Which elements of her overall reading program are the most important to her success?
- What about supplemental services such as your Title I program, ELL program, or special education program? Are they important elements of _____ (Teacher's Name) success? Please explain.
- How does _____ (Teacher's Name) use data to inform her literacy instruction?

9. _____ (Teacher's Name) has been quite successful with her Hispanic students. To what do you attribute her success?

    Probes:
- What components of your "core" reading series are effective with Hispanic students? Which are ineffective? Which components do you see _____ (Teacher's Name) using?
- What components of your supplemental materials and programs are effective with Hispanic students? Which are ineffective? Which components do you see _____ (Teacher's Name) using?
- What elements of your overall reading program are effective with Hispanic students? Which are ineffective? Which elements do you see _____ (Teacher's Name) using?
- **If in the above, ELL programming and Title I programming are not discussed, ask about their effectiveness with Hispanics.**

10. How closely does _____ (Teacher's Name) collaborate with her fellow grade level teachers?

    Probe: Does she participate in grade level meetings? How often? What occurs at these meetings?

117

11. How much of her success is attributable to this collaboration? Please explain.

12. **<u>IMPORTANT QUESTION</u>: How does _____ (Teacher's Name) differ from her colleagues?**

13. How much do parents contribute to _____ (Teacher's Name) success? Please explain.

14. What role do you play in _____'s (Teacher's Name) success?

15. How does your principal contribute to _____ (Teacher's Name) success?

16. Summary Questions:
     a. Is there anything else that you've thought of as we have been talking that would answer the question "Why do you think _____ (Teacher's Name) is doing so well with her students?"

     b. What are the strengths and weaknesses of Reading First? How does _____ (Teacher's Name) manifest these?

     c. What does Reading First need to do so that all teachers attain the level of success that _____ (Teacher's Name) has?

Thank you for your time.

Idaho Reading First Best Practices Synthesis
Principal Interview Questions

Introduction:

_____ (Teacher's Name) has been identified as a highly successful Reading First teacher. We would like to find out to what you attribute her success, so we have asked to talk with you today. I will also be interviewing _____ (Teacher's Name) to ask her about her program and to what she attributes her success, and _____ (Coach's Name), your reading coach. Thank-you for your time.

You will not be identified by name or school in the final evaluation report submitted to the state next fall. No one will know you have been interviewed except your reading coach and _____ (Teacher's Name). The contents of this interview will not be released to anyone. We have always conducted these evaluations in a confidential manner since we want teachers, coaches, and principals to feel free to respond in any way they choose. The interview will take 30-45 minutes.

You are participating in what is called a **best practices synthesis.** We are interviewing 25 Reading First teachers—about 6 per grade level – and their coaches and principals. The teachers are in about 15 of the 30 Idaho Reading First schools. The 25 teachers are some of the most successful based on test scores out of the roughly 400 K-3 teachers in Idaho Reading First schools. What we want to do is find out what the most successful teachers are doing and disseminate that information to the rest of the Reading First teachers.

The motivation behind this project came from some of the results from last year's evaluation. The results showed that test scores in most Idaho Reading First schools are not consistently going up across all grade levels. There are pockets of sustained improvement at certain grade levels in individual schools, but no school consistently grew at every grade level over the years they have been in Reading First. There was also no initiative-wide sustained improvement in test scores. But, some individual teachers are doing quite well. Thus, we thought it important to talk with these teachers to find out what they attribute their success to and then share that with the rest of the Reading First teachers. We also felt it important to talk with the coaches and principals in the buildings where the teachers are for additional insights and perspectives on what makes the particular teacher so successful.

Do you have any questions before we begin? **GET CONSENT FORM SIGNED!!!**

I would like to discuss some terms before we get started.

I will ask you questions about your "core" reading program. What I mean by "core" reading program is the basal reading series your school adopted as part of Reading First.

I will ask you questions about the supplemental materials and programs that you have purchased to augment the core reading program. These are the materials or programs you have purchased to address the specific needs of subgroups of children you have identified needing something more or different from what the core program provides.

I will also ask you questions about your overall reading program. When I use the term overall reading program, I mean all the elements that go into your K-3 reading program including all of the materials you use (i.e., core reading series and supplemental materials and programs); the daily schedule; the supplemental services provided by Title I, ELL, and special education; your role as principal; and the coach's role.

I'll now begin with some general background questions concerning your experience as a principal.

1. Name of Interviewee: _____

Circle One: Teacher   Coach   Principal

2. School: _____

3. How many years have you been a principal?   _____

4. How many years were you a teacher? _____

5. What grade levels did you teach? _____

6. Did you teach or coach in a Reading First school before becoming a principal of a Reading First school? _____

Now I would like to turn to our discussion of what you attribute _____ (Teacher's Name) success to:

7. Let's begin by talking about your "core" reading program and the supplemental materials and programs that you use. A big part of Reading First is the adoption of a "core" reading program that all teachers then use. Also many of the Reading First schools have purchased additional supplemental materials and programs. I would like to talk about the core program and any supplemental materials and programs you have.

a. What core program do you use?

Note: "Your" in the following questions refers to the principal, not the targeted teacher. We want to hear from the principal's perspective about the materials being used.

b. What supplemental materials and programs are part of your overall reading program?

c. Please tell me how your core program and supplemental materials and programs influence your success with your students.
   Probes:
   - What are the strengths and weaknesses of the core program?
   - What are the strengths and weaknesses of the supplemental materials and programs?
   - Which elements of these programs are the most important to _____ (Teacher's Name) success?
   - Which elements are not as helpful to _____ (Teacher's Name)?
   - How important are the supplemental materials and programs you have purchased to _____ (Teacher's Name)? With which of her students?

8. Now that we have discussed the materials that you use to teach reading, let's broaden our discussion to include your overall reading program.

a. Please describe your overall reading program?

b. What elements of your overall program are important to _____ (Teacher's Name) success?
NOTE: Make sure the probes are answered.

   Probes:
   - What does a typical day in _____ (Teacher's Name) classroom look like when she is teaching reading?
   - How are students grouped for their literacy instruction in _____(Teacher's Name) classroom?
   - Which elements of her overall reading program are the most important to her success?

- What about supplemental services such as your Title I program, ELL program, or special education program? Are they important elements of _____ (Teacher's Name) success? Please explain.
- How does _____ (Teacher's Name) use data to inform her literacy instruction?

9. _____ (Teacher's Name) has been quite successful with her Hispanic students. What do you attribute her success to?

    Probes:
- What components of your "core" reading series are effective with Hispanic students? Which are ineffective? Which components do you see _____ (Teacher's Name) using?
- What components of your supplemental materials and programs are effective with Hispanic students? Which are ineffective? Which components do you see _____ (Teacher's Name) using?
- What elements of your overall reading program are effective with Hispanic students? Which are ineffective? Which elements do you see _____ (Teacher's Name) using?
- **If in the above, ELL programming and Title I programming are not discussed, ask about their effectiveness with Hispanics.**

10. How closely does _____ (Teacher's Name) collaborate with her fellow grade level teachers?

    Probe: Does she participate in grade level meetings? How often? What occurs at these meetings?

11. How much of her success is attributable to this collaboration? Please explain.

12. **IMPORTANT QUESTION: How does _____ (Teacher's Name) differ from her colleagues?**

13. How much do parents contribute to _____ (Teacher's Name) success? Please explain.

14. What role do you play in _____'s (Teacher's Name) success?

15. How does your coach contribute to _____ (Teacher's Name) success?

16. Summary Questions:
   a. Is there anything else that you've thought of as we have been talking that would answer the question "Why do you think _____ (Teacher's Name) is doing so well with her students?"

   b. What are the strengths and weaknesses of Reading First? How does _____ (Teacher's Name) manifest these?

   c. What does Reading First need to do so that all teachers attain the level of success that _____ (Teacher's Name) has?

Thank-you for your time.