

**Bureau of Indian Education's (BIE)  
Reading First Program:  
2006-2007 Summative Evaluation Report**

Submitted on

November 30, 2007

by

Catherine Callow-Heusser, Director

[cheusser@endvision.net](mailto:cheusser@endvision.net)

Heather Chapman, Statistical Analyst

EndVision Research & Evaluation, LLC

41 E. University Blvd. #321

Logan, UT 84321

(435) 881-8811



## Executive Summary

This report includes the overall BIE Reading First Program evaluation findings from the 17 BIE Cohort 1 Reading First schools who have completed three years of Reading First implementation, from 2004-2007, and 13 BIE Cohort 2 Reading First schools who completed their first year of implementation in 2006-2007. A presentation given to the BIE's Reading First Leadership Team has been included in Appendix E. Data for this report were collected over the 2006-2007 year and supplemented with data from previous years for Cohort 1. During the 2006-2007 school year, the external evaluation team visited Cohort 1 schools just at the end of year, and visited Cohort 2 schools at beginning, middle and end of year (only the end of year site visit was conducted at Cohort 2 schools in Maine, Indian Township and Beatrice Rafferty). The external evaluation team administered the DIBELS benchmark assessments to all K-3 students, conducted observations of schools and classrooms, conducted interviews and administered surveys to principals and other school leadership, reading coaches, and teachers, and reviewed other data at the Wireless Generation and University of Oregon DIBELS websites.

On average, a substantial number of K-3 students in Bureau of Indian Affairs (BIE) Reading First Cohorts 1 and 2 schools made progress during the 2006-2007 school year. In fact, over 67% of Cohort 1 K-3 students made progress (e.g., stayed at 'Benchmark' or changed classification in the positive direction, as indicated by Instructional Recommendations on the DIBELS benchmark assessments), with the percentage of K-3 students who made progress in individual schools ranging from 81% to 33%. Percent progress in Cohort 2 schools ranged from 74% to 45%. Additionally, the percentage of students who begin the year at high risk of reading failure has dropped from 40% at baseline to 18% in Cohort 1 schools (3 years of implementation), and from 41% at baseline to 29% in Cohort 2 schools (one year of implementation). The percentage of students reading at grade level has changed from a baseline of 28% to 59% in Cohort 1 schools and from 26% to 44% in Cohort 2 schools based on DIBELS scores.

These positive outcomes were apparent despite the many geographic challenges faced from implementing a Reading First program in remote schools spread throughout the United States:

- Increased engagement of instructional leadership,
- Increased fidelity of implementation,
- Increased progress monitoring with the DIBELS and use of data to adjust intervention,
- Increased knowledge of research-based and effective reading instruction,
- Increased use of effective teaching practices,
- Increased student engagement,
- Increased numbers of students reading at Benchmark or improving based on the DIBELS instructional recommendations and other curriculum-embedded assessments, as well as standardized (e.g., SAT-10) and state-level criterion-referenced tests, and
- Professional development and technical assistance that has changed over time to better meet needs.

Schools in which student outcomes in reading showed the greatest improvement had

- Strong, engaged, and supportive leadership,

- Stable (i.e., high retention), informed and supportive staff with high levels of communication about reading goals, students' progress data, and intervention,
- High expectations for students,
- Well-implemented research-based reading programs,
- Policies that supported implementation of reading programs,
- Established plans for reading intervention for struggling students, and
- Regular progress monitoring and review of data to plan instruction and intervention.

Additionally, the BIE's Reading First schools are better meeting the needs of struggling readers, so that they improve reading skills rather than falling further and further behind. Schools report fewer special education referrals and placements, and greatly reduced referrals for behavior.

Overall, data provide support for positive changes in use of effective teaching practices and fidelity of implementation of reading programs, demonstrating the positive impact of Reading First's professional development and on-going, onsite technical assistance. However, data indicate that additional training for teachers at BIE Reading First schools is needed for

- Fidelity of implementation of reading programs and teaching strategies,
- Systematic and explicit instruction,
- Teaching strategies for students who are at-risk, English Language Learners (ELL), or identified with learning disabilities,
- Efficient and effective time management, and
- Positive management of classroom behavior.

While some teachers and schools made substantial progress in these areas, others struggled as evidenced by student outcomes. The move towards more targeted assistance to schools, based on evidence from student outcomes and school and classroom observations, strengthened implementation of reading programs across schools, and more consistent technical assistance than in previous years increased Reading First school staff satisfaction with technical assistance and likely impacted student outcomes. Continued frustration with communication and mixed messages exists but complaints have decreased from previous years. Data also indicate that some Cohort 1 schools needed continued technical assistance during their third year of implementation, and without this assistance and more onsite feedback and monitoring, five Cohort 1 schools did not meet progress criteria for continued funding, and over half of Cohort 1 schools did not meet their percent progress from the previous year.

While the percentage of students entering school with limited English proficiency certainly plays a role in the BIE's Reading First outcomes, and particularly SAT-10 scores, the positive outcomes achieved within some schools, including those with higher percentages of English Language Learners, demonstrate that strong implementation of research-based reading programs combined with effective and engaged leadership, a knowledgeable reading coach with strong leadership skills, and high quality professional development, contribute to dramatic increases in the number of K-3 students reading at grade level and decreases in the numbers needing intensive intervention. Overall, the BIE's Reading First program has demonstrated high levels of success in changing the reading trajectories of struggling readings and increasing the percentages of students reading at grade level.

## Table of Contents

<b>BIE Annual Reading First Program Report .....</b>	<b>1</b>
The Evaluation Program Model.....	1
Expectations for BIE Reading First Schools .....	1
BIE Reading First Program Evaluation Questions .....	2
Program Effectiveness .....	2
Alignment and Coherence.....	3
Capacity Building .....	3
Methods .....	3
Table 1: School Site Visit Data Collection Methods and Timelines, Cohort I.....	4
Table 2: School Site Visit Data Collection Methods and Timelines, Cohort 2 .....	5
Table 3: Other Data Collected, Cohorts I and II.....	5
DIBELS Scores and Instructional Recommendations .....	6
SAT-10 Reading First Scores .....	6
Demographic Data .....	6
Interviews.....	7
Classroom Observations .....	7
Fidelity of Implementation .....	7
Findings .....	9
Program Effectiveness .....	9
Figure 1: BIE Cohort 1 and Cohort 2 Baseline DIBELS Instructional Recommendations Percentages .....	11
Figure 2: Cohort 1 DIBELS Instructional Recommendations and SAT-10 At Grade Level (AGL).....	13
Figure 3: Cohort 2 DIBELS Instructional Recommendations and SAT-10 .....	15
Figure 4: DIBELS Effectiveness Chart—Cohorts 1 and 2 Combined .....	16
Figure 5: DIBELS Benchmark and SAT-10 AGL by School—Cohort 1 Schools with Higher Percentages of Students Speaking a Language other than English.....	17
Figure 6: DIBELS Benchmark and SAT-10 AGL by School—Cohort 1 Schools with Lower Percentages of Students Speaking a Language other than English.....	18
Figure 7: Percent Progress of Cohorts 1 and 2 Schools.....	19
Figure 8: Graph of Data Used for an RDD Analysis, Cohort 1 2 <sup>nd</sup> Grade .....	24
Table 4: Regression Discontinuity Analyses Results .....	25
Figure 9. Cohorts 1 & 2, Average Trajectories from BOY to EOY by Group.....	27
Table 5: Use of Effective Teaching Practices and Fidelity of Implementation .....	31
Figure 10. Classroom Observation Feedback Form .....	35
Table 6. Special Education Referrals and Placements at Turtle Mountain.....	36
Alignment and Coherence.....	37
Capacity Building .....	40
Additional Findings .....	41
Table 7: Predictive Ability of DIBELS When Compared to SAT-10 .....	42

Figure 11: Average ORF Scores by Group over Time .....	43
Table 8: School-Level Factors Affecting Student Outcomes .....	44
Conclusions.....	45
<b>Appendix A: BIE Reading First Evaluation Questions.....</b>	<b>47</b>
BIE Reading First Subgrant Evaluation Questions.....	47
Administration, Organization, & Communication.....	47
Goals, Objectives, & Priorities .....	48
Assessment.....	48
Instructional Programs & Materials.....	49
Differentiated Instruction, Grouping, Staffing, & Scheduling .....	49
BIE Reading First Program Evaluation Questions .....	50
Program Effectiveness .....	50
Alignment & Coherence .....	50
Capacity Building .....	50
<b>Appendix B: Research-Based Effective Teaching Practices .....</b>	<b>52</b>
<b>Appendix D: Results of RDD Analyses .....</b>	<b>61</b>
<b>Appendix E: Predictive Ability of DIBELS When Compared to SAT-10.....</b>	<b>69</b>
<b>Appendix F: 2006-2006 Reading First Evaluation Report to the BIE’s Leadership Team ....</b>	

## **BIE Annual Reading First Program Report**

This evaluation follows the specifications written in the BIE's Reading First Grant Application, approved by the U.S. Department of Education in 2003. The subgrant evaluation questions used to evaluate BIE Reading First subgrants are listed in that application and adapted in Appendix A in this report. Additionally, this report provides an analysis of data collected through interviews, classroom observations, and DIBELS outcomes for the 2006-2007 school year.

This report in combination with individual school evaluation reports for the 17 BIE Cohort I Reading First schools (whose first year of implementation was 2004-2005) and 13 BIE Cohort II Reading First schools (whose first year of implementation was 2006-2007) provide a comprehensive view of the BIE Reading First Program.

### ***The Evaluation Program Model***

The evaluation program model selected is the one advocated by Posavac and Cary in their 1997 book, *Program Evaluation Methods and Case Studies*. This model requires (1) formative evaluations to progressively and systematically improve practices, and (2) summative evaluations to answer the question, "Is the Reading First program working, for whom, and under what conditions?" This document reports on the methods and findings from the third annual summative evaluation.

The following Reading First expectations for schools, Reading First Program evaluation questions, and the adapted Reading First Subgrant questions included in Appendix A guided the evaluation design, data collection methods, instrument development, analysis of data, and presentation of findings.

### ***Expectations for BIE Reading First Schools***

The BIE's Reading First Grant Application specifies the following components of scientifically based reading research (SBRR) practices that are expected to be replicated in all BIE Reading First schools (p. 49):

- BIE's high expectations for reading achievement,
- Sufficient instructional time (i.e., minimum 90-minute scheduled reading block),
- Careful lesson planning,
- School-wide assessment system,
- School-wide interventions for struggling readers,
- Sound instructional approaches,
- School climate of collaboration, strong leadership, and evidence of commitment, and
- High-quality professional development.

Given these expectations, the grant application states that the following characteristics that should be observable and measurable prominent features in all Reading First classrooms (p. 53-54):

- Reading programs based on SBRR that include instructional content based on the five essential components of reading;
- Coherent instructional design that includes explicit instructional strategies, coordinated instructional sequences, ample practice opportunities, and aligned student materials;
- Ongoing use of assessments that inform instructional decisions;
- Protected, dedicated block of time for reading instruction;
- Clear expectations for student achievement and clear strategies for monitoring progress;
- Small-group instruction, as appropriate, to meet individual student needs, with placement and movement based on ongoing assessments;
- Active student engagement in a variety of reading-related activities which are connected to the components of reading and academic goals;
- Instruction designed to bring all children to grade level, with appropriate, scientifically based intervention strategies aligned with classroom instruction designed for students not making sufficient progress.

### ***BIE Reading First Program Evaluation Questions***

The major questions that guided the evaluation of the BIE Reading First Program include the following areas of focus as adapted from the BIE Reading First grant proposal (BIE/OIEP, 2003, pp. 26-29).

#### **Program Effectiveness**

1. What gains are students making in grades K-3 as a result of Reading First programming?
2. How successful are professional development activities in changing teachers' instructional practices in reading?
3. How do LEA choices regarding assessment measures, embedded in adapted core program, impact schools' and LEAs' use of assessment information when planning instructional approaches?
4. What specific changes are being made at the classroom, school, and LEA levels as a result of Reading First formative and summative evaluations?
5. Are referral rates to Special Education changing in Reading First schools; particularly referrals linked to reading failure in the K-3 grades?
6. Are exit rates for English Language Learners changing in Reading First schools?

## **Alignment and Coherence**

1. Are there specific changes in pedagogical approaches across reading instruction for those identified as needing additional intervention?
2. Are there changes in existing standards and/or benchmarks that improve alignment to Reading First?
3. Are other non-Reading First schools and/or districts adopting Reading First priorities?
4. Are there shifts in LEA instructional practices in other content areas as they relate to reading instruction?
5. Are there shifts in staff development programs that reflect the BIE's Reading First priorities?

## **Capacity Building**

1. Are teachers able to adapt their new knowledge of instructional strategies for teaching reading to a range of situations, such as working with at-risk learners, ELL students, or students with possible learning disabilities?
2. Are Reading First teachers participating in professional activities outside of Reading First programs—providing workshops, attending conferences, or training other teachers, administrators in approaches to reading instruction?
3. Are pre-service and in-service teacher training programs incorporating Reading First approaches instruction in their reading courses?

## **Methods**

The data used in this report and in individual Reading First school reports were collected during site visits to 17 BIE Cohort I Reading First schools that have received continued funding and 13 BIE Cohort II Reading First schools that were funded for the 2006-2007 school year. Additional data were collected during professional development activities and Institutes for Beginning Reading (IBR) attended by evaluators, and during meetings and communication with BIE Reading First key personnel, including the Project Director, Reading Specialists, and school-level principals and reading coaches. Data were collected during the 2006-2007 school years during one end-of-year site visit to Cohort I schools, and three site visits to Cohort II schools. During site visits, data were collected through observations of schools and classrooms, interviews with principals and other school leadership, reading coaches, and teachers and other Reading First staff, and surveys administered to Reading First staff. Data gathering methods and timelines are shown in Tables 1 and 2. Individual instruments are available upon request.



**Table 1: School Site Visit Data Collection Methods and Timelines, Cohort I**

<b>Data Gathering Method</b>	<b>School Site Visit Dates</b>	<b>Number (%) Completed</b>
DIBELS benchmark assessment of K-3 students in 17 BIE Cohort I Reading First (RF) schools (schools completed their own MOY and EOY benchmark assessments)		
End-of-year (EOY)	April-May, 2007	2107 (96%)
Interviews at Cohort I Schools		
Principals	EOY	17 at 15 schools
Reading Coaches	EOY	20 at 15 schools
Teachers and Paraprofessionals	EOY	156 at 17 schools
Surveys at Cohort I Schools		
Principals	EOY	15 at 12 schools
Reading Coaches	EOY	20 at 16 schools
Teachers and Paraprofessionals	EOY	181 at 17 schools
Classroom Observations in Cohort I schools		
	EOY	117 at 17 schools+

+ Includes school staff who provided additional intervention but who were not the primary teacher assigned to a K-3 Reading First classroom (e.g. paraprofessionals, reading coaches, other school staff who provided instruction to small groups of students), or teachers who did not teacher the entire school year (e.g., left after January or were hired in March). Observations were conducted in over 80% of BIE Reading First classrooms.

**Table 2: School Site Visit Data Collection Methods and Timelines, Cohort 2**

<b>Data Gathering Method</b>	<b>School Site Visit Dates</b>	<b>Number (%) Completed*</b>
DIBELS benchmark assessment of K-3 students in 13 BIE Cohort II Reading First (RF) schools		
Beginning-of-year (BOY)	August-Sept, 2006	1220 (93%)
Middle-of-year (MOY)	January-February, 2007	
EOY	April-May, 2008	1353 (98%)
Interviews at Cohort II Schools		
Principals	BOY, MOY, EOY	14, 13, 13 at 11, 11, 13 schools
Reading Coaches	BOY, MOY, EOY	9, 11, 13 at 11, 11, 13 schools
Teachers	MOY, EOY	81, 89 at 11, 13 schools
Surveys at Cohort II Schools		
Principals	BOY, EOY	8, 0, 14 at 7, 0, 10 schools
Reading Coaches	BOY, EOY	10, 12 at 10, 13 schools
Teachers	BOY, EOY	86, 106 at 10, 13 schools
Classroom Observations in Cohort II Schools	BOY, MOY, EOY	61, not calculated, 94+ at 9, 10, 13 schools

\* We did not visit Beatrice Rafferty or Indian Township in Maine at BOY or MOY. Wireless Generation completed their BOY and MOY benchmark assessments.

+ Includes school staff who provided additional intervention but who were not the primary teacher assigned to a K-3 Reading First classroom (e.g. paraprofessionals, reading coaches, other school staff who provided instruction to small groups of students), or teachers who did not teacher the entire school year (e.g., left after January or were hired in March). Observations were conducted in over 80% of BIE Reading First classrooms.

**Table 3: Other Data Collected, Cohorts I and II**

<b>Method</b>	<b>Dates</b>
Review of Other Data, e.g., Wireless Generation DIBELS and mClass Direct Websites, University of Oregon's DIBELS website	Ongoing for individual school data, graphs, and reports
Participation in Institutes for Beginning Reading (IBRs):	
Summer IBR, Phoenix	August, 2006
Data Summit, Phoenix	April, 2006
Summer IBR, Phoenix	August, 2007
Participation in BIE Reading First Leadership Meetings (fall and summer meetings were cancelled because of low commitment to attend)	February, 2007
Communication with:	
BIE Reading First Project Director, Reading Specialists	Ongoing
BIE Reading First school staff, particularly reading coaches	Ongoing

## **DIBELS Scores and Instructional Recommendations**

Trained assessors who are members of the external evaluation team administered the DIBELS assessments for beginning-, middle- and end-of-year benchmarks for Cohort II schools, and for end-of-year benchmarks for Cohort I schools. Students who were absent on the days of site visits were assessed by reading coaches. School staff at Cohort I schools administered the beginning- and middle-of-year benchmarks. Per requests from Cohort I schools, the external evaluation team will administer the beginning- and end-of-year DIBELS benchmarks after the 2006-2007 school year.

The DIBELS benchmarks were administered on Palm handheld devices and data were uploaded to the Wireless Generation mClass website. Other than occasional glitches using the technology and uploading assessment results, the process of administering the DIBELS on Palm handhelds worked well. Many problems with the technology have been solved over time. However, fall DIBELS when new software has been installed results in more frequent (i.e., approximately 1 or 2 of 100 students) Palm problems requiring that the assessment be re-administered.

Obtaining the DIBELS scores in electronic format continues to be problematic. While the pipeline between the Wireless Generation DIBELS site and the University of Oregon's DIBELS site was operational during the 2006-2007 school year, not all scores or students transferred correctly. Additionally, when a school changed a child's classroom assignment, the University of Oregon's DIBELS website placed that child in an undesignated classroom rather than the classroom based on the new assignment; hence, the student's data were not "assigned" to a specific teacher/classroom.

## **SAT-10 Reading First Scores**

The SAT-10 Reading First data for the 2006-2007 school year was entered by had by EndVision staff. This data were double entered, matched, and cleaned to ensure accuracy in entry. SAT-10 student level data for Atsa' Biya'a'zh and Aneth, and for Cibequé's 1<sup>st</sup> grade students have not yet been received, and are not included in analyses completed for this report. Similarly, the data file uploaded with the report includes group-level data for Atsa' Biya'a'zh, but no data for Aneth or for Cibequé's 1<sup>st</sup> grade students.

## **Demographic Data**

Although several requests were issued to schools asking that students' demographic data (e.g., disabilities in particular) be updated in the Wireless Generation's website so the demographic data would transfer to the University of Oregon's website, and hence be available in electronic form, several schools did not enter demographic data, and many did not update the data for late enrollees. Because several reading coaches admitted the demographic data would not likely be complete or accurate, we opted to not use that data. However, for the purposes of the electronic data submission with this report, and for data analyses conducted by EndVision staff, all students at BIE Reading First schools were considered both American Indian and economically disadvantaged.

## Interviews

Interviews with administrators, reading coaches, and teachers were conducted at the school during site visits. While many informal discussions were held, the table above lists only the formal interviews that were completed. Interviews included questions about (a) school climate and culture, (b) demographic information, (c) successes and challenges in implementing reading programs, (d) changes in response to Reading First, (e) process and implementation of reading programs, and (e) other open ended questions designed to gather responses to assist in providing formative feedback to improve programs, interpreting student outcome data and explain findings.

## Classroom Observations

Classroom observations included data regarding use of effective teaching practices (see Appendix B for research-based effective teaching practices from which many items on the observation instrument were selected, e.g., teacher-student interaction, instructional feedback, pacing, transitions, behavior management, student engagement, classroom environment) and implementation of reading program (e.g., student grouping for instruction, use of materials and centers, activities addressing the five key components of reading instruction, time management). While 20 minute observations of a classroom during a 90 minute or longer reading block may limit inferences about what occurred in that classroom throughout the entire 90 or more minutes, research shows that short and repeated observations of a single classroom can provide accurate snapshots of classroom activities and behaviors, and repeated observations of multiple classrooms can provide reliable composite pictures of trends within a school.

While we have heard concerns about conducting classroom observations during the same site visits in which the DIBELS benchmark assessments are administered, the DIBELS team rarely disrupts more than a few classrooms during the reading block. Observers strive to either observe classrooms in which students are not being pulled for the DIBELS assessments or to conduct the observations with sufficient time provided for students to become re-engaged in reading instruction after students have been assessed. Therefore, as indicated when scheduling site visit dates, we conducted classroom observations during the scheduled reading block with the expectation that typical reading instruction would occur.

## Fidelity of Implementation

Fidelity of implementation was determined using data from both interviews and classroom observations by (a) comparing activities observed during reading time to “A Consumer's Guide to Evaluating a Core Reading Program Grades K-3: A Critical Elements Analysis” (Simmons & Kame'enui, 2003), (b) comparing materials and activities observed during classroom observations to publisher's materials for the core and/or supplemental reading programs and materials, (c) considering pacing and academic learning time (i.e., length of transitions, amount of time students are engaged with reading instruction or activities that support learning to read), both within the scheduled reading block and across time, to insure adequate coverage of the five key components of reading instruction each day and during a single school year, and (d) looking at student grouping, use of whole and small group instruction, and use of centers,

particularly when compared to publisher's guidelines and the BIE's Reading First Leadership's recommendations.

The level of implementation of reading programs was scored for each classroom, from which an overall rating for the school was assigned. The school level rating also included an assessment of additional factors, such as the involvement of school leadership in the Reading First program; the degree and type of communication and meetings that occurred between administration, reading coaches, teachers, other Reading First staff, and the school community; and the extent to which administrators, reading coaches, teachers, and other Reading First staff supported program implementation. See Appendix A for additional information regarding evaluation questions.

### *Classroom Rating of Fidelity of Implementation*

*High* Systematic and explicit reading instruction was consistently observed. Instruction

- Involved exclusive use of core/ supplemental program materials,
- Was aligned with the five key components of reading instruction,
- Used time efficiently to maximize academic learning time,
- Was on pace to cover sufficient content for each grade level, and
- Created a positive learning environment that maximized opportunities for learning, including appropriate modeling and instructional feedback, frequent opportunities for all students in the instructional group to demonstrate mastery of concepts, and high levels of student engagement.

*Fair/Moderate* Instruction occurred using the core/supplemental materials but typically

- Instruction was not consistently both systematic and explicit, or
- The activities were not well-aligned with key components of reading instruction, or
- Pacing was too slow to cover sufficient content for grade level, or
- Not a high rate of opportunities for students to respond so teachers could check student understanding (e.g., group or individual responses during instruction), or
- Academic learning time needs to be increased (e.g., not all students engaged with instruction, pacing too slow, or instructional time reduced for other reasons including interruptions or long transitions).

*Poor* Reading instruction occurred but

- Was not aligned with the five key components of effective reading instruction (e.g., phonemic awareness, phonics, vocabulary, fluency, comprehension), and/or
- Pacing of lessons or pacing across time too slow to ensure sufficient coverage, and/or
- Included use of materials created by teachers or from other reading programs, and/or
- Did not include systematic and explicit instruction, and/or
- Many students were observably off-task and not engaged with instruction.

*None* No reading instruction observed during classroom visit. For example,

- A test was administered to students, or
- Class time was spent reviewing using activities (e.g., games, teacher-created materials not based on publisher’s guidelines, materials from other publishers) that were not in reading program, or
- Students worked independently during the entire observation although reading instruction would be expected given the reading program used, or
- During repeated visits to the classroom, activities were not related to reading (e.g., math, games, recess, bathroom breaks, movies, sustained silent reading, coloring or matching activities not related to the reading lesson).

## **Findings**

Formative evaluation reports for individual Reading First school subgrant evaluations are available upon request. However, this report includes all summative findings for individual schools.

## **Program Effectiveness**

Each of the evaluation questions related to BIE Reading First program effectiveness is addressed in this section. Additional questions related to the first question have been added based on a more thorough analysis of the data using regression discontinuity techniques and other methods, as described subsequently.

Evaluation Question 1.           What gains are students making in grades K-3 as a result of Reading First programming?

First, it is interesting to note that Cohort 2 baseline averages were almost identical to Cohort 1 baseline averages, based on DIBELS beginning-of-year benchmark instructional recommendations, as shown in Figure 1. While 40% of the students in Cohort 1 were initially intensive based on DIBELS instructional recommendations, 41% of Cohort 2 students were intensive. Also, 28% of Cohort 1 students were initially reading at grade level based on DIBELS instructional recommendations, while 26% of Cohort 2 students were at benchmark. Additionally, the distribution of DIBELS instructional recommendations across schools show similar trends in Cohort 1 and Cohort 2, as demonstrated by the distributions of “intensive” and “benchmark” percentages as one looks down the graphs for Cohorts 1 and 2.

This finding could be considered surprising, because evidence suggests cohorts who begin implementation subsequent to the first cohort often behave differently and are different on baseline characteristics. We expected Cohort 2 to have fewer struggling readers, because

- Cohort 2 schools experienced an additional 2 years of the requirements of No Child Left Behind, which includes a focus on preventing academic failure for struggling learners,
- Most Cohort 2 schools applied for Reading First funding during the Cohort 1 competition,

- Cohort 2 schools had an additional 2 years to prepare for funding,
- Some Cohort 2 school staff visited Cohort 1 schools to learn about Reading First implementation,
- Some Cohort 2 schools implemented Reading First models or components prior to Cohort 2 funding, and
- Stories of BIE Cohort 1 Reading First successes and challenges were communicated among staff at BIE schools who were in close proximity or who attended meetings attended by staff at BIE schools.

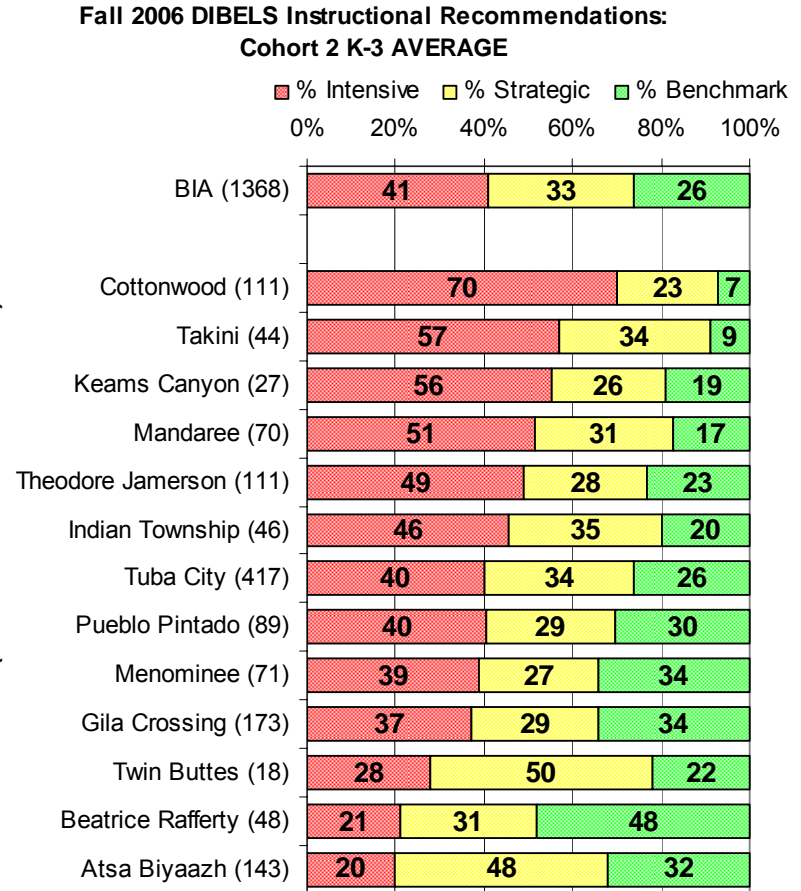
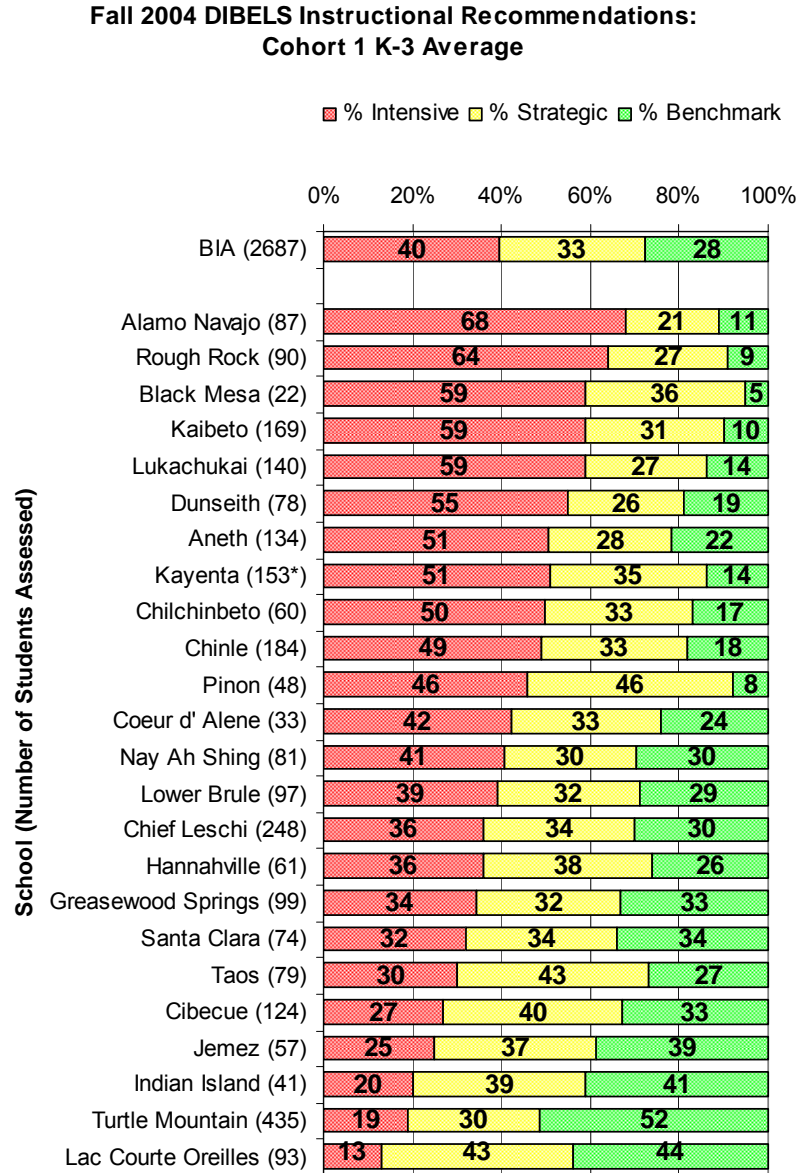
Overall, students at most BIE Reading First schools have made substantial progress in reading skills. As the following figures show, BIE Cohort 1 and Cohort 2 Reading First schools have, on average, increased scores on the DIBELS end-of-year benchmark assessments and across all SAT-10 subtests. In particular, the percentage of students reading on grade level as measured by the end-of-year DIBELS benchmark has increased within BIE Reading First schools overall, and at most schools individually, and the percentages of students at high risk of reading failure has decreased.

Figure 2 shows trends across the 3 years of implementation for BIE averages on the DIBELS benchmark assessments and SAT-10 subtests for Cohort 1. Figure 3 displays the same information across the single year of implementation for Cohort 2. SAT-10 data has not yet been received for Aneth and for Cibequé's first grade students in Cohort 1, and for Atsa' Biya'a'zh in Cohort 2.

Figure 4 shows the numbers of students, by grade, who changed DIBELS instructional recommendations by the end of the 2006-2007 school year, when compared to their status at the beginning of the school year. This figure combines both Cohorts 1 and 2, because low-performing Cohort 1 schools that did not receive continued Reading First funding for the 2007-2008 school year have been removed from our access at the Wireless Generation website. This graphic was downloaded before those schools were removed.

The leftmost column includes students who began the year intensive according to DIBELS instructional recommendations. The middle column includes students who were strategic at the beginning of the year, and the rightmost column includes students who were benchmark at the beginning of year. The red color indicates students who ended the year intensive, and the green color shows students who ended the year at benchmark. Overall, the numbers of students "moving up" in instructional recommendation or staying at benchmark increases substantially. However, not all students who begin the year reading on grade level remain on grade level, as shown in Figure 7. The rightmost column includes students who began the year at benchmark, but the red and yellow numbers in that column indicates students who fell below benchmark cutoffs by the end of the school year. The red in the middle column shows students who fell from strategic to intensive.

**Figure 1: BIE Cohort 1 and Cohort 2 Baseline DIBELS Instructional Recommendations Percentages**





Figures 5 and 6 separate Cohort 1 schools with higher and lower percentages of students who enter school speaking a language other than English. While we don't have a consistent measure for English Language Learners (ELL) across BIE schools to rank schools based on the percentage of ELL students (see Appendix C), we have grouped the schools according to approximate categories of higher and lower percentages of ELL students. In reviewing 2006-2007 SAT-10 results by school, a clear trend becomes apparent: schools with higher percentages of students who enter speaking a language other than English have lower SAT-10 scores when compared to schools that have higher percentages of students who enter school speaking English.

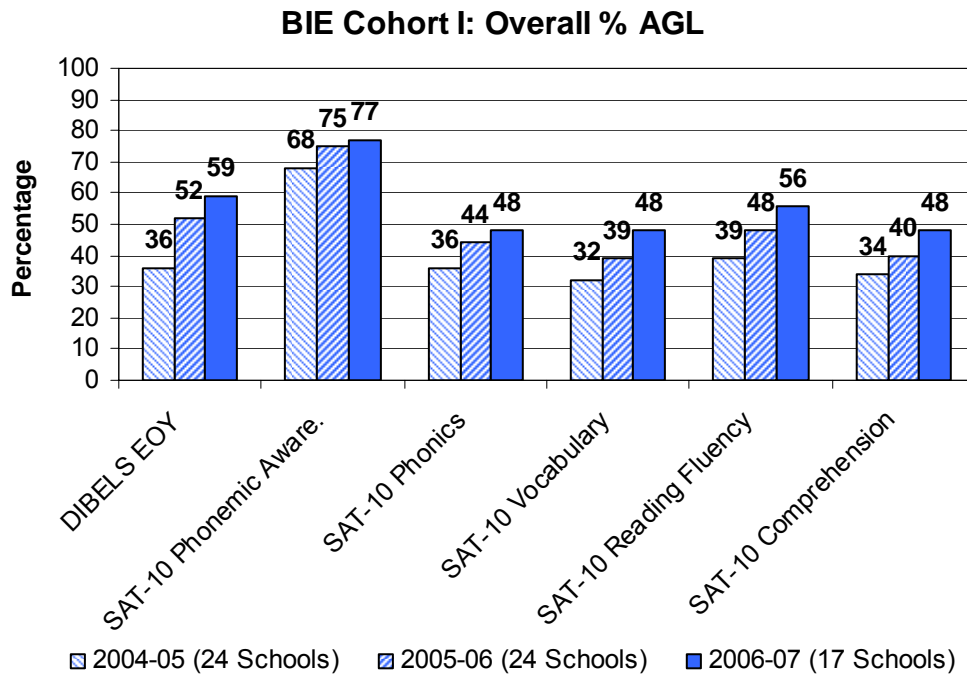
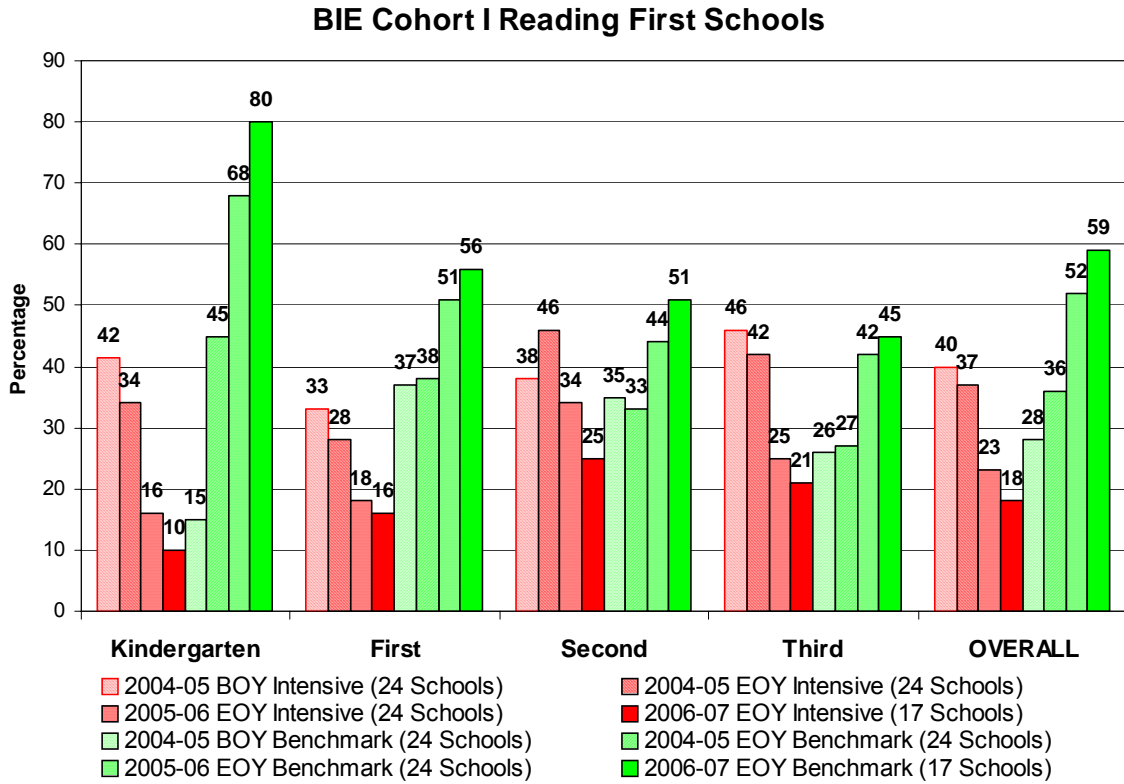
To better measure students' vocabulary skills, schools administered the Peabody Picture Vocabulary Test (PPVT) to all K-3 students starting in fall of 2007, with an end-of-year assessment scheduled as well.

Finally, refunding decisions for Cohort I schools were made based on percent progress, calculated using DIBELS instructional recommendations. Percent progress includes students who changed instructional recommendations in a positive direction or stayed at benchmark from the beginning to end of year. So percent progress is the percent of students who

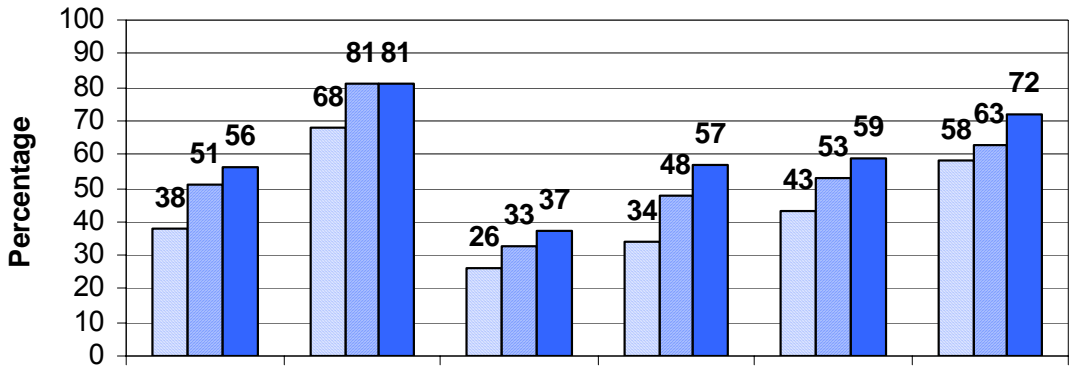
- Started and ended at Benchmark
- Moved from Strategic to Benchmark
- Moved from Intensive to Benchmark
- Moved from Intensive to Strategic

Figure 7 shows the percent progress for BIE Reading First schools. The criteria for continued funding during the 2007-2008 school year—(a) greater than 70% progress based on DIBELS Instructional Recommendations or (b) greater than 60% progress with documented evidence of fidelity of implementation—was established by the BIE's Reading First Leadership Team and communicated to Cohort 1 schools prior to the third year of implementation. All Cohort 2 schools were refunded for a second year of implementation.

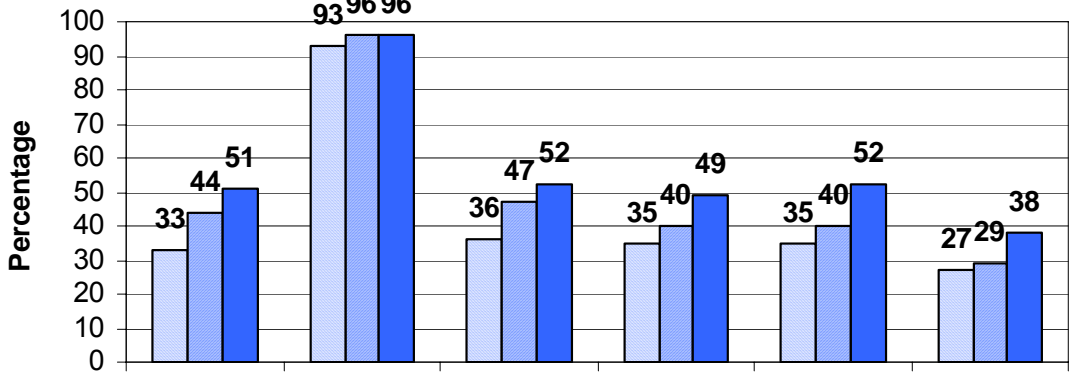
**Figure 2: Cohort 1 DIBELS Instructional Recommendations and SAT-10 At Grade Level (AGL)**



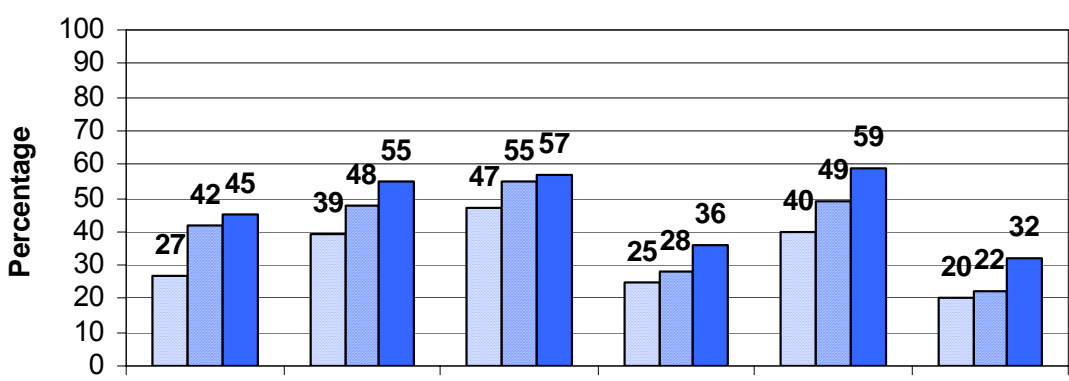
### BIE Cohort I: First Grade %AGL



### BIE Cohort I: Second Grade %AGL



### BIE Cohort I: Third Grade %AGL



2004-05 (24 Schools)
  2005-06 (24 Schools)
  2006-07 (17 Schools)

**Figure 3: Cohort 2 DIBELS Instructional Recommendations and SAT-10**

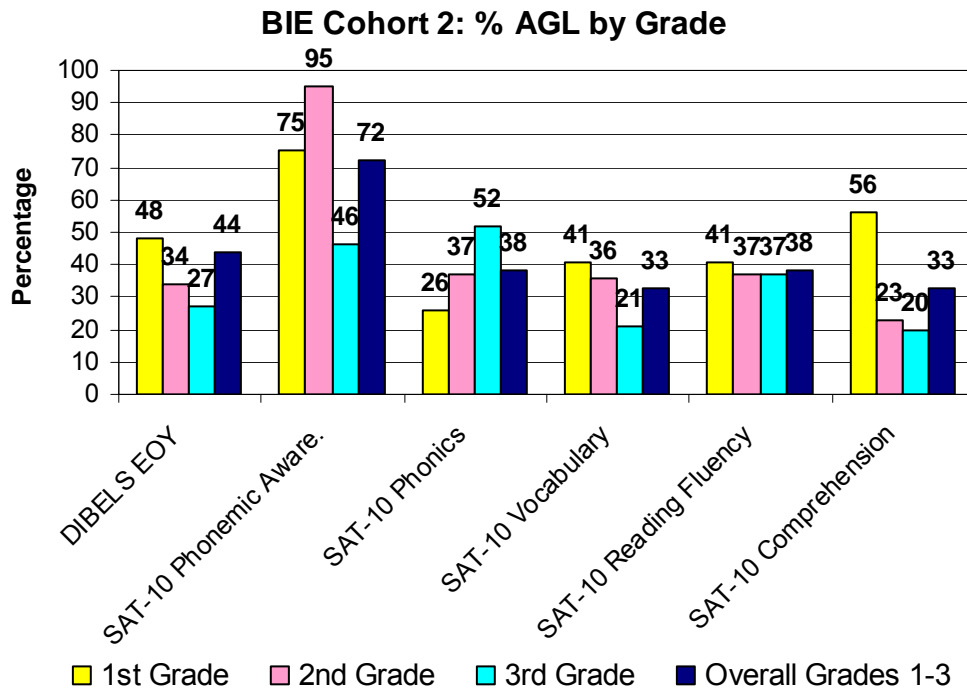
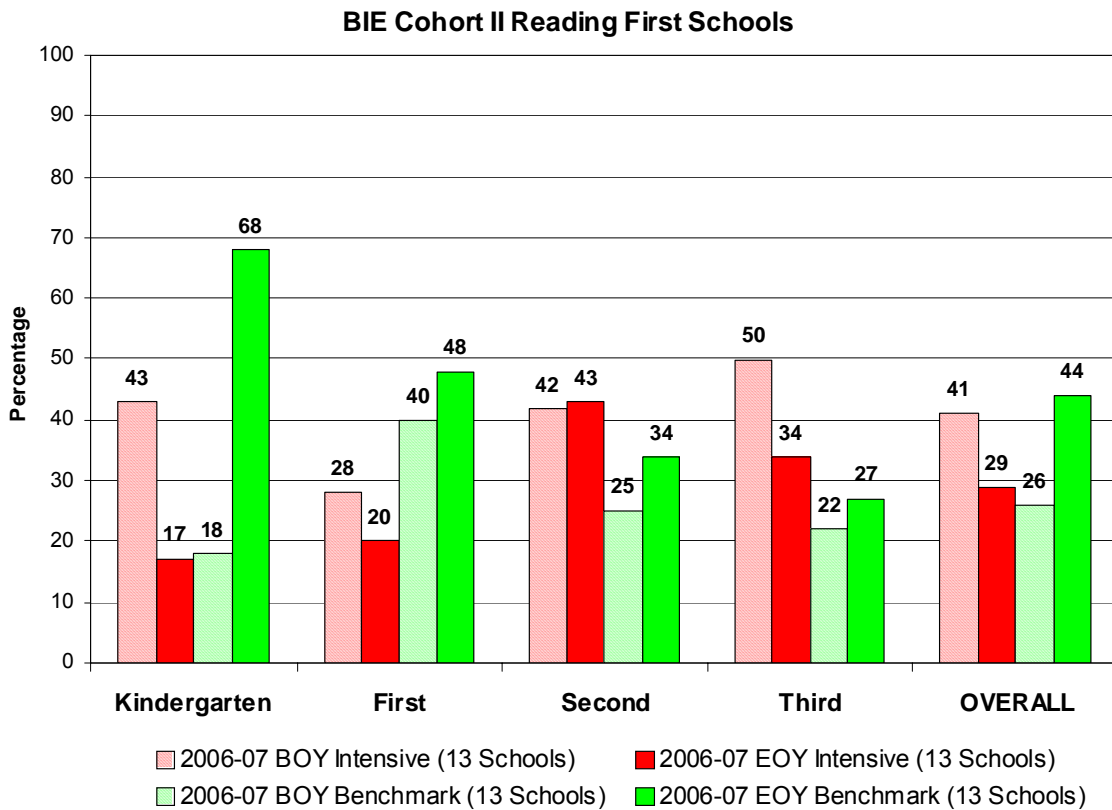
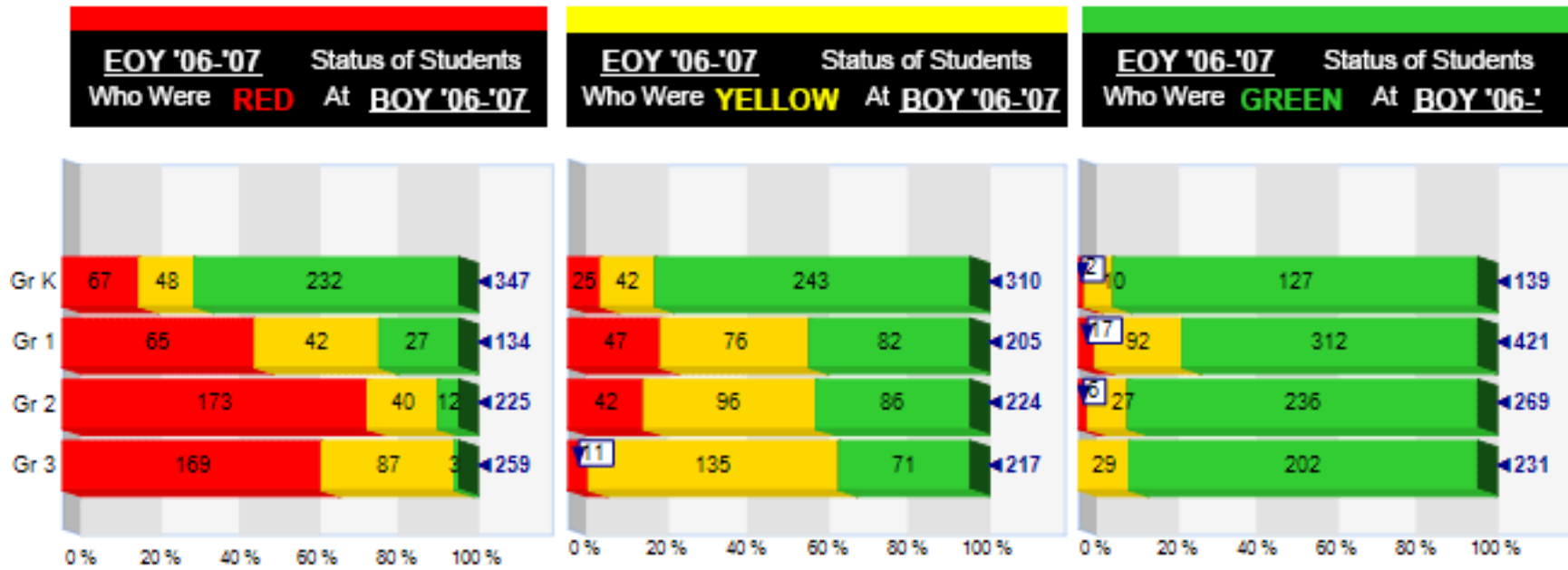
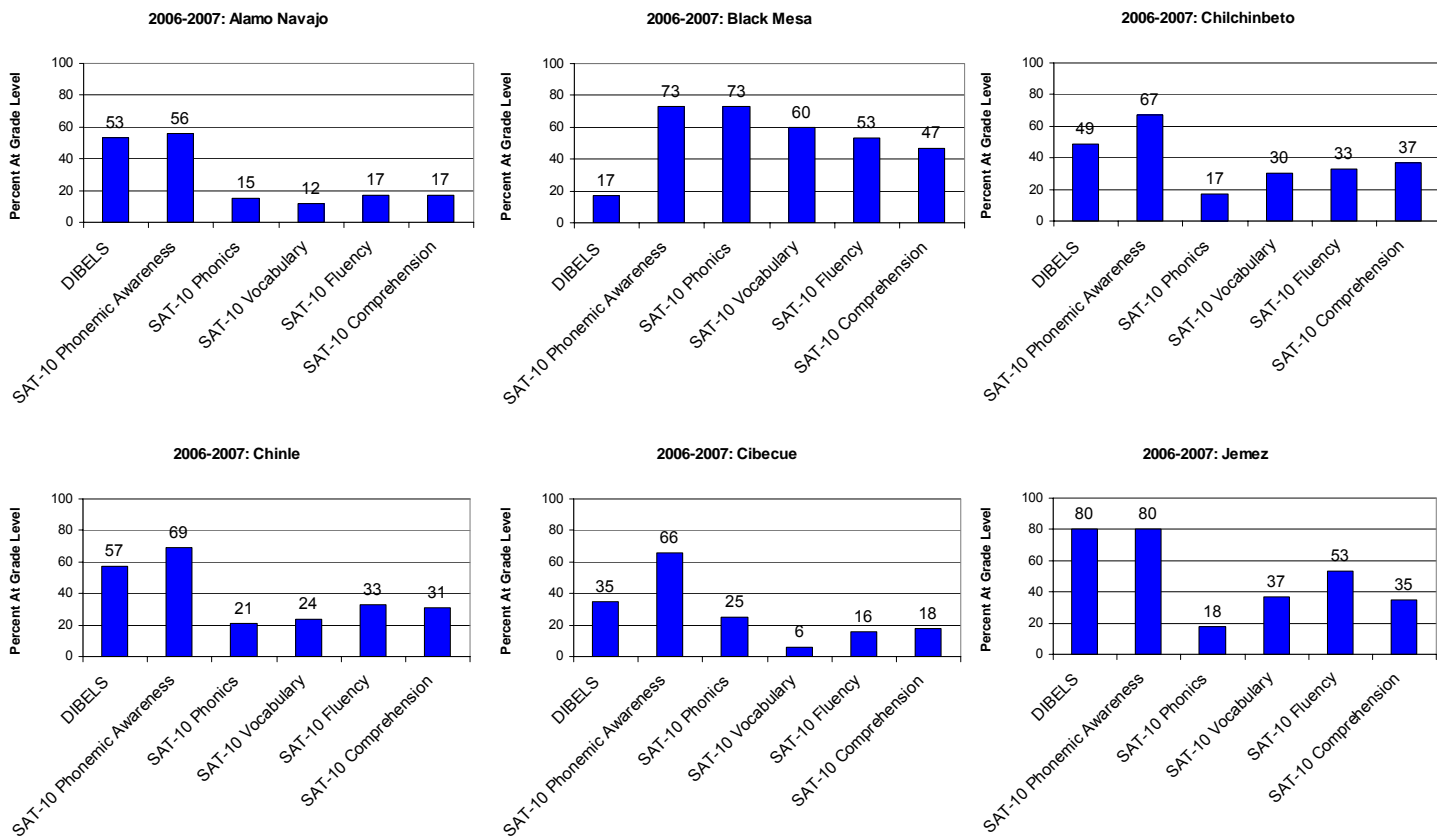


Figure 4: DIBELS Effectiveness Chart—Cohorts 1 and 2 Combined



**Figure 5: DIBELS Benchmark and SAT-10 AGL by School—Cohort 1  
Schools with Higher Percentages of Students Speaking a Language other than English**



**Figure 6: DIBELS Benchmark and SAT-10 AGL by School—Cohort 1  
Schools with Lower Percentages of Students Speaking a Language other than English**

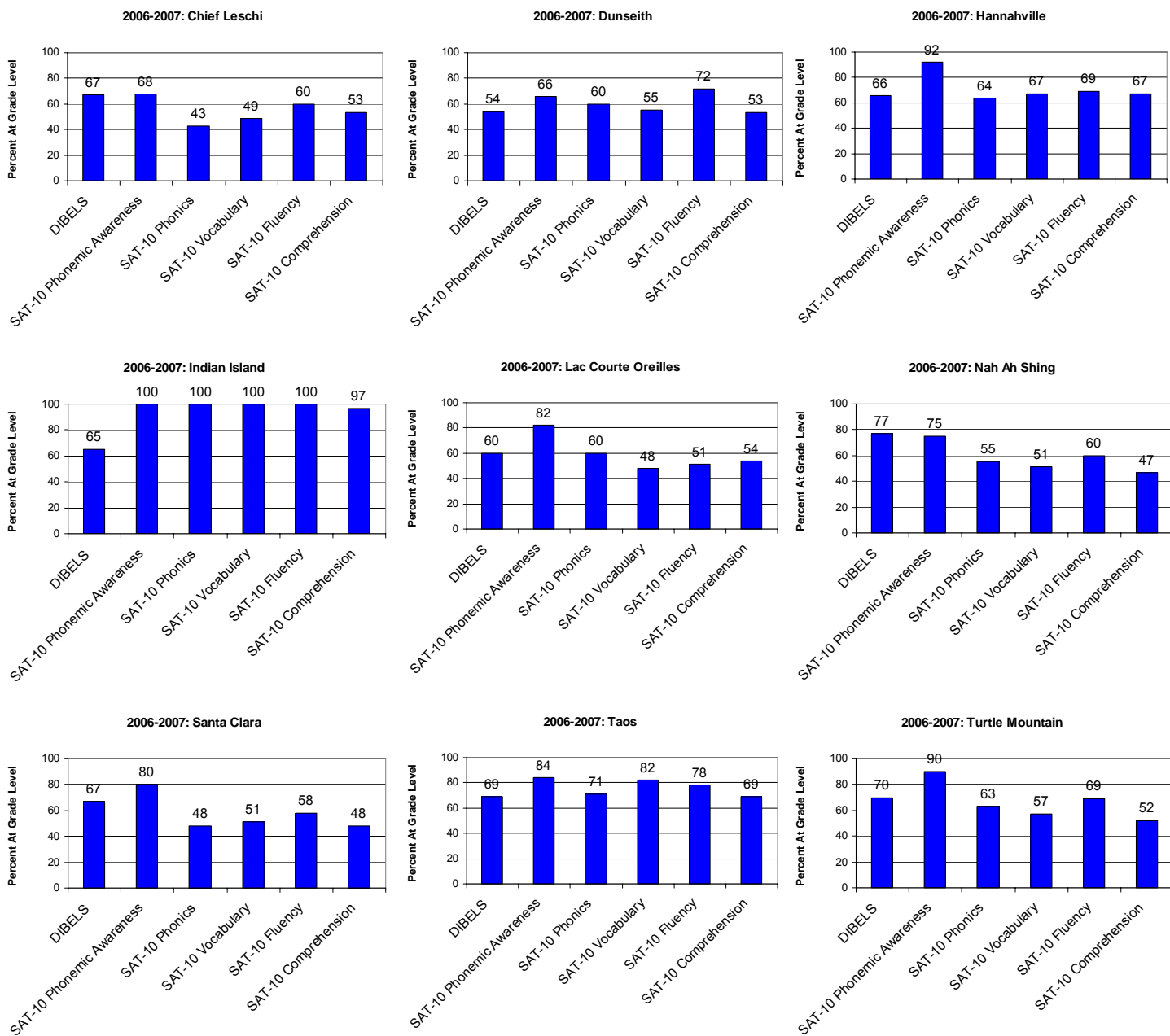


Figure 7: Percent Progress of Cohorts 1 and 2 Schools

<b>Cohort 1</b>	<b>Percent Progress</b>	<b>Cohort 2*</b>	<b>Percent Progress</b>
Nay Ah Shing	81%	Mandaree	74%
Jemez	80%	Theodore Jamerson	65%
Dunseith	75%	Twin Buttes	63%
Turtle Mountain	75%	Pueblo Pintado	63%
Chief Leshi	74%	Atsa' Biya'a'zh	63%
Hannahville	73%	Cottonwood	61%
Taos	73%	Beatrice Rafferty	59%
Alamo Navajo	70%	Gila Crossing	54%
Santa Clara	70%	Indian Township	54%
Indian Island	66%	Menominee	53%
LCO	65%	Takini	46%
Chinle	61%	Keams	46%
Pinon	54%	Tuba	45%
Chilchinbeto	53%		
Aneth	51%		
Black Mesa	40%		
Cibecue	33%		
<b>BIE Cohort I</b>	<b>67%</b>	<b>BIE Cohort II</b>	<b>55%</b>

Refunded  
 Refunded given  
 support  
 documentation  
 submitted by  
 school  
 Not refunded for  
 another year of  
 implementation



\* All BIE Cohort II schools were funded for a second year of implementation



Evaluation Question 2. Do students who begin the school year INTENSIVE on the DIBELS receive intervention that is 'different' than students who are Strategic or Benchmark, and what is the magnitude of that difference?

The Reading First program was designed without opportunities for determining program effectiveness using the “gold standard” of randomized experimental designs, where students, teachers or schools are randomly assigned to treatment and control groups. However, Reading First provides an ideal context for implementing a quasi-experimental regression discontinuity design (RDD), where students below a pre-specified cutoff receive intervention that is different than those above the cutoff. In particular, RDD involves the initial selection of the treatment and comparison groups based on a cutoff score, which minimizes selection bias by basing group membership strictly on performance on a pretest measure.

When appropriately implemented and analyzed, RDD yields an unbiased estimate of treatment effect (Rubin, 1977; Shadish, 2006). The What Works Clearinghouse and current requests for proposals from the Institute of Education Sciences state that the RDD is an acceptable alternative to randomized experiments because RDDs “come close to true experiments in minimizing the effects of selection bias on estimates of effect size.” However, the relationships between the assignment variable and outcome variables must be correctly modeled, including analysis of nonlinear relationships and interactions. Additionally, samples must be of sufficient size to meet power requirements. According to Cappelleri, Darlington, and Trochim (1994), a power estimate of 0.80—the value that is traditionally viewed as “sufficient” power—would require a sample size of approximately 150 to detect a medium effect size (Cohen, 1988) of 0.3, and a sample size of approximately 70 to detect a large effect size of 0.5 or greater. If a cutoff “interval” where subjects between a lower and an upper cutoff are dropped from the design, a smaller sample would be needed to detect similar effects with the same statistical power.

A key goal of Reading First is to prevent reading failure by insuring that children are reading on grade level by the end of third grade. Under the BIE’s Reading First program, children who are “intensive” on the DIBELS benchmark assessments are taught using a replacement core program in almost all BIE Reading First schools, and are double- and triple-dosed with reading instruction, where additional intervention beyond the scheduled reading block is provided to those students at highest risk of reading failure.

If we assume that BIE Reading First students whose beginning-of-year DIBELS benchmark scores place them in the “intensive” category are treated differently than those whose scores are above the “intensive” cutoffs, we can implement a RDD. In this case, the DIBELS beginning-of-year benchmark instructional recommendation of “intensive” defines the treatment group, which creates unbiased assignment to groups as required by RDDs. In most BIE Reading First schools, intensive students are placed in homogeneous reading groups in which the teacher uses a replacement core program that provides more systematic and explicit instruction than the core basal reading program. Additionally, intensive students at all schools receive additional reading instruction outside of the scheduled reading block.

A RDD answers the question, “Do students who begin the school year INTENSIVE on the DIBELS receive intervention that is 'different' than students who are Strategic or Benchmark, and what is the magnitude of that difference?”

An RDD analysis was conducted for Cohorts 1 and 2 separately, by grade, with each of the following outcome measures: DIBELS Oral Reading Fluency, SAT-10 Vocabulary, SAT-10 Reading Fluency, and SAT-10 Reading Comprehension Strategies. For 2<sup>nd</sup> and 3<sup>rd</sup> grades, the groups were determined based on beginning of year DIBELS ORF scores. In 1<sup>st</sup> grade, the groups were based on beginning of year DIBELS Nonsense Word Fluency scores. The sizes of all groups were sufficient to detect a medium effect size. Because the SAT-10 findings were initially non-significant, students who were strategic based on DIBELS instructional recommendations at the beginning of the year were removed from the analysis. Shadish (2006) and Cappelleri, Darlington, and Trochim (1994) recommend a “cutoff interval” when the group within that interval may have been treated similarly to either of the other groups. In BIE Reading First schools, students performing at the low end of strategic are sometimes placed in intervention groups similar to intensive students, and high performing strategic students are often treated similarly to benchmark students. Conducting the analysis in this way still retained sufficient numbers of students to meet statistical power guidelines, and resulted mostly in statistically significant findings with large effects. The RDD analyses using the DIBELS scores as outcome measures were not rerun with students in the cutoff interval removed, because the findings were already statistically significant with large treatment effects.

The measure of effect, or effect size, in an RDD analysis is the standardized estimate of the interaction term, as highlighted in bolder and larger numbers in Table 4. More complete tables from the statistical analyses can be found in Appendix D.

Figure 8 provides a visual example of a RDD analysis, based on Cohort 1 2<sup>nd</sup> grade Oral Reading Fluency (ORF) scores as the outcome measure. In this graph, the vertical line defines the cutoff between the intensive students who receive treatment that we assume is different than students falling to the right of the vertical line—those who are strategic and benchmark based on beginning-of-year DIBELS instructional recommendations. The horizontal line defines the end-of-year benchmark cutoff. Students whose end-of-year ORF scores fall above that line are reading at grade level based on the DIBELS benchmark assessments. Students who are intensive at the beginning of year are indicated in red. Strategic students are shown in yellow, and benchmark students in green. The x-axis is the beginning-of-year DIBELS ORF scores converted to z-scores, and the y-axis represents the end-of-year scores on the DIBELS ORF.

Similar graphs have been constructed for all other analyses but are not included here to conserve space. Most of the graphs show similar patterns, as reflected in the results shown in Table 4 and Appendix D. Additionally, the graphs provide visual support for the linear relationships of the regression lines for each group, per appropriate application of RDD analysis. Further analysis of the linearity of regression lines for each group was conducted using piecewise regression with quadratic terms.

Note the substantially steeper regression line on the left side of the graph when compared to the right side. This steeper slope provides evidence that the intensive students were treated

differently than strategic and benchmark students. The RDD analysis provides further support for those differences, with large effects in a number of cases, as shown in Table 4 and described below.

In RDD, an important indicator of model fit are adjusted  $R^2$  values. Overall, the RDD models used to analyze the data for both cohorts fit the DIBELS data well in 2<sup>nd</sup> and 3<sup>rd</sup> grades, and the SAT-10 data moderately well in 2<sup>nd</sup> and 3<sup>rd</sup> grades, as indicated by the  $R^2$  values which provide an indication of model fit. The  $R^2$  values indicate the amount of variability in outcome scores accounted for by the model.

After controlling for beginning-of-year DIBELS scores (ORF for 2<sup>nd</sup> and 3<sup>rd</sup> grades) and including a treatment and an interaction effect, the fit of the models as indicated by the adjusted  $R^2$  values ranged from 0.722 to 0.776 with end-of-year DIBELS ORF used as the outcome measure in 2<sup>nd</sup> and 3<sup>rd</sup> grades. These values indicate an excellent model fit, as the model accounts for over 72% of the variability in outcome scores. The adjusted  $R^2$  values ranged from 0.308 to 0.513 for SAT-10 outcomes in 2<sup>nd</sup> and 3<sup>rd</sup> grade.

To explain further, between 72% and 78% of the variability in end-of-year DIBELS ORF scores can be attributed to the three covariates included in the model: beginning of year DIBELS scores, group assignment, and the interaction of beginning of year DIBELS scores and group assignment. Based on suggestions for adequacy of model fit made by researchers such as Cohen (1994), this range of  $R^2$  values indicates a high degree of model fit for DIBELS outcomes, and a moderate fit for SAT-10 outcomes.

After controlling for beginning-of-year DIBELS scores in 1<sup>st</sup> grade (Nonsense Word Fluency, NWF, as ORF is not administered at the beginning of year) and including a treatment and an interaction effect, the fit of the models is substantially lower for 1<sup>st</sup> grade. The models account for less of the variability in outcome scores, with  $R^2$  values ranging from 0.13 to 0.434. Further, the treatment and interaction estimates for Cohort 2 in 1<sup>st</sup> grade are not statistically significant, indicating that the relationship between beginning of year NWF scores, group membership, and outcome measures is not strong, and with these data, the differential impact of intervention for struggling learners in Cohort 2 1<sup>st</sup> grade was not confirmed using these models. While the interaction estimates for Cohort 1 vocabulary and comprehension are statistically significant at the 0.10 level, the amount of variability in outcomes accounted for by the model is low, indicating other factors may be more important in predicting outcomes.

According to the regression discontinuity method, there are two indications of success of treatment. The first is a significant positive treatment effect, which indicates that students in the intervention group made greater gains than those in the control group, which consists of strategic and intensive students in our case. These numbers are not included in Table 4 but are included in Appendix D, as the statistical analysis does not continue if the treatment effect is not statistically significant.

The second important indication of increased gain in the intervention group is a significant positive interaction effect. The interaction effect gives an indication of how much progress can be expected for an average member of the intervention group when compared to the control

group after accounting for beginning of year scores. In essence, it is an indication of how quickly average scores for the intervention group approach average scores of the control group.

For an example, as shown in Table 4, the standardized estimate for the interaction effect for Cohort 1 students in 2nd grade based on the DIBELS ORF outcome measure indicates that intensive students gained an average of 1.25 ( $p < 0.001$ ) standard deviation units above what students in the originally higher performing group gained. Thus, although these gains may not be enough to bring the intensive students up to the reading level of their peers in one year, based on these results, the additional intervention provided to these struggling students does positively affect their reading trajectory when compared to the initially higher performing group.

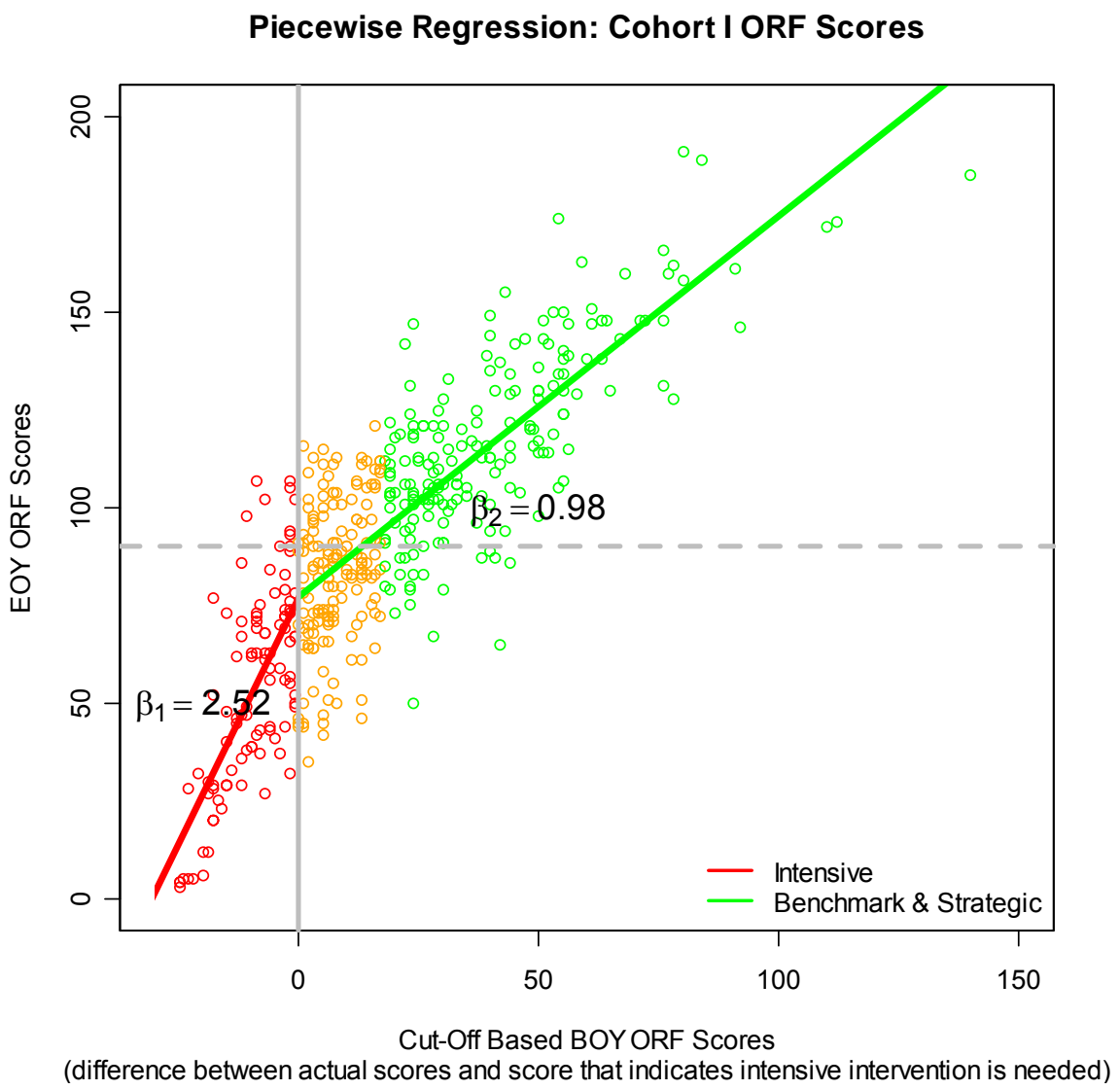
Results for Cohort 1 based on DIBELS outcomes are strong in 2<sup>nd</sup> and 3<sup>rd</sup> grades, with the magnitude of impact greater than 1.25 with a high  $R^2$  values. This indicates that intensive students receive intervention that increases their gains when compared to students who were strategic or benchmark at beginning of year. More importantly, the SAT-10 outcomes for Cohort 1 2<sup>nd</sup> and 3<sup>rd</sup> grade students shows strong effects with a moderate model fit. The interaction estimates ranging from 0.38 to 0.79 show that the intensive students make greater gains on SAT-10 subtests than students in the higher performing group.

Results for Cohort 2 are somewhat mixed, with some outcomes showing strong effects with large or moderate  $R^2$  model fit values. Because Cohort 2 was in its first year of implementation, mixed results are not surprising.

Overall, the RDD results indicate moderate to high  $R^2$  values of model fit, providing support for analyzing the data using these techniques. Additionally, the first step of the data analysis shows large and highly statistically significant treatment effects, indicating differential gains between groups, as can be seen in the graph in Figure 8, for example. Because these treatment effects are large and statistically significant, continuing with the next step of the analysis is warranted, adding the interaction term.

The interaction estimates indicate large effects for most of the outcome measures for 2<sup>nd</sup> and 3<sup>rd</sup> grades, particularly for Cohort 1. These effects indicate that intensive students who receive intervention that we assume is different than the other group (and other data confirm this difference) make greater gains when compared to initially higher performing students. In other words, intervention is positively affecting the reading trajectory for students at high risk of reading failure.

Figure 8: Graph of Data Used for an RDD Analysis, Cohort 1 2<sup>nd</sup> Grade



**Table 4: Regression Discontinuity Analyses Results**

<b>Cohort/Grade</b>	<b>Outcome Measure</b>	<b>Adjusted R<sup>2</sup></b>	<b>Standardized Interaction Estimate<sup>+</sup></b>
Cohort 1, 1 <sup>st</sup> Grade	DIBELS ORF	0.434	0.62 n.s.
	SAT-10 Vocabulary <sup>++</sup>	0.129	<b>1.25+</b>
	SAT-10 Reading Fluency <sup>++</sup>	0.24	0.86 n.s.
	SAT-10 Comprehension <sup>++</sup>	0.215	<b>1.03+</b>
Cohort 1, 2 <sup>nd</sup> Grade	DIBELS ORF	0.736	<b>1.25**</b>
	SAT-10 Vocabulary <sup>++</sup>	0.337	0.08 n.s.
	SAT-10 Reading Fluency <sup>++</sup>	0.418	<b>0.79*</b>
	SAT-10 Comprehension <sup>++</sup>	0.308	<b>0.65*</b>
Cohort 1, 3 <sup>rd</sup> Grade	DIBELS ORF	0.735	<b>1.5**</b>
	SAT-10 Vocabulary <sup>++</sup>	0.513	<b>0.47+</b>
	SAT-10 Reading Fluency <sup>++</sup>	0.419	<b>0.53*</b>
	SAT-10 Comprehension <sup>++</sup>	0.502	<b>0.38+</b>
Cohort 2, 1 <sup>st</sup> Grade	DIBELS ORF	0.296	Negative n.s.
	SAT-10 Vocabulary <sup>++</sup>	0.164	Negative n.s.
	SAT-10 Reading Fluency <sup>++</sup>	0.20	0.2 n.s.
	SAT-10 Comprehension <sup>++</sup>	0.332	0.48 n.s.
Cohort 2, 2 <sup>nd</sup> Grade	DIBELS ORF	0.722	<b>0.72**</b>
	SAT-10 Vocabulary <sup>++</sup>	0.37	<b>0.64*</b>
	SAT-10 Reading Fluency <sup>++</sup>	0.32	<b>0.49+</b>
	SAT-10 Comprehension <sup>++</sup>	0.287	0.42 n.s.
Cohort 2, 3 <sup>rd</sup> Grade	DIBELS ORF	0.776	<b>0.68**</b>
	SAT-10 Vocabulary <sup>++</sup>	0.432	0.26 n.s.
	SAT-10 Reading Fluency <sup>++</sup>	0.471	<b>0.61*</b>
	SAT-10 Comprehension <sup>++</sup>	0.419	<b>0.7*</b>

+ Magnitude of effect, or effect size of difference in intervention for intensive students when compared to other students

++ Students who began the year strategic were removed from this analysis, as they may have been treated more like intensive students if they were low-performing strategic students, or more like benchmark students if they were higher performing

n.s. = not significant

+  $p < 0.10$

\*  $p < 0.05$

\*\*  $p < 0.001$

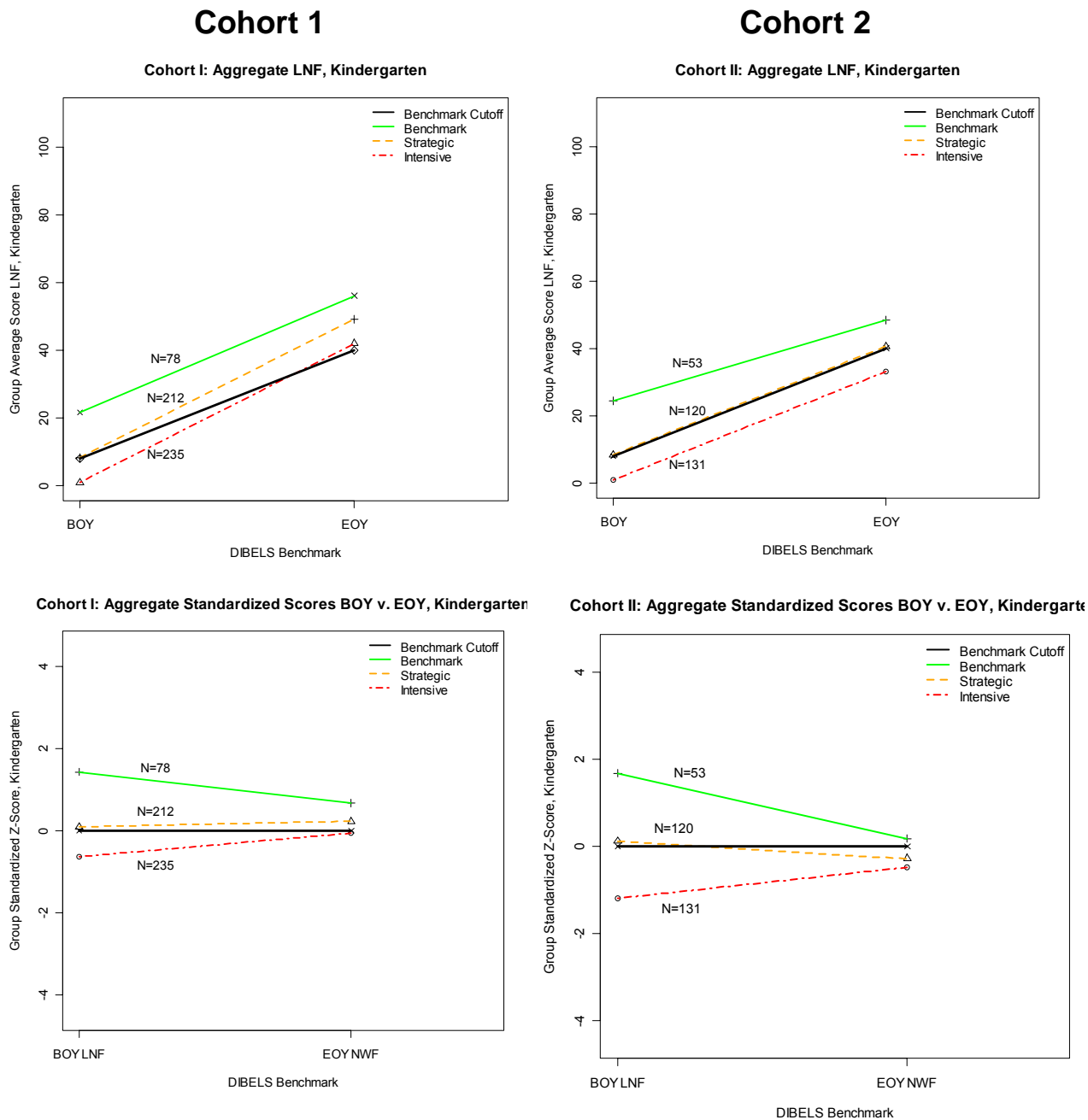
### Evaluation Question 3. Are struggling students “catching up?”

While answers to the previous questions indicate that intervention has been provided to struggling students that has changed their reading trajectories, the question about whether struggling students are “catching up” sufficiently to be able to read on grade level by the end of 3<sup>rd</sup> grade remains. The following graphs show that the BIE’s Reading First program has certainly changed, on average, the trajectories of students who begin the year intensive on the DIBELS, particularly when compared to prior research findings that showed struggling 3<sup>rd</sup> grade students continued to fall behind in subsequent years. However, overall, students in 2<sup>nd</sup> grade do not show a steep enough trajectory, on average, to catch up—based on the increasing cutoffs for Oral Reading Fluency, while students in 3<sup>rd</sup> grade demonstrate, on average, a trajectory that is parallel to the benchmark cutoffs.

In these graphs, the benchmark, strategic and intensive lines represent students whose instructional recommendations on the DIBELS placed them in those categories at the beginning of the year. Those intact groups and the average of their scores at beginning of year and end of year form the lines. The benchmark cutoff line is shown in solid black. Graphs are included for each grade across measures that are administered at both the beginning and end of year. Additionally, using z-scores, trajectories comparing beginning of year Letter Naming Fluency and end of year Nonsense Word Fluency are included for kindergarten, and comparing beginning of year Nonsense Word Fluency to end of year Oral Reading Fluency for 1<sup>st</sup> grade. In these graphs, notice that the average trajectory for intensive students approaches the benchmark cutoff lines, while the average benchmark trajectory parallels the benchmark cutoff or at least stays above it.

Ideally, all three lines in a graph would have a trajectory that placed the end of year averages above the cutoff line. While that happens in several of these graphs, the critical graphs showing Oral Reading Fluency in 2<sup>nd</sup> and 3<sup>rd</sup> grade indicate that one year is not enough for struggling students to catch up, and in 2<sup>nd</sup> grade, struggling students are still at risk of falling even further behind.

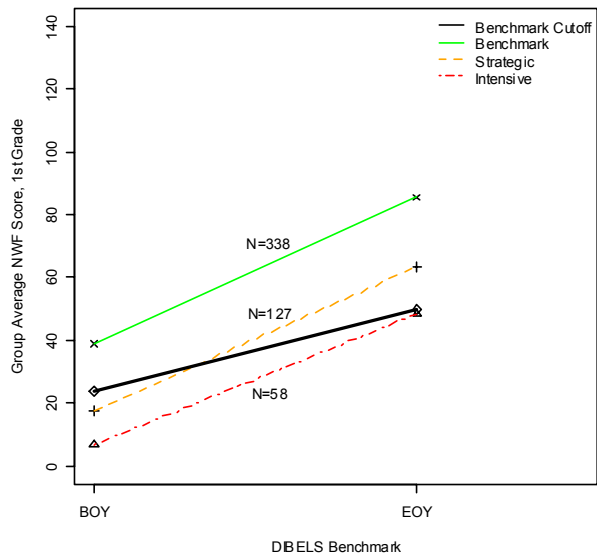
Figure 9. Cohorts 1 & 2, Average Trajectories from BOY to EOY by Group





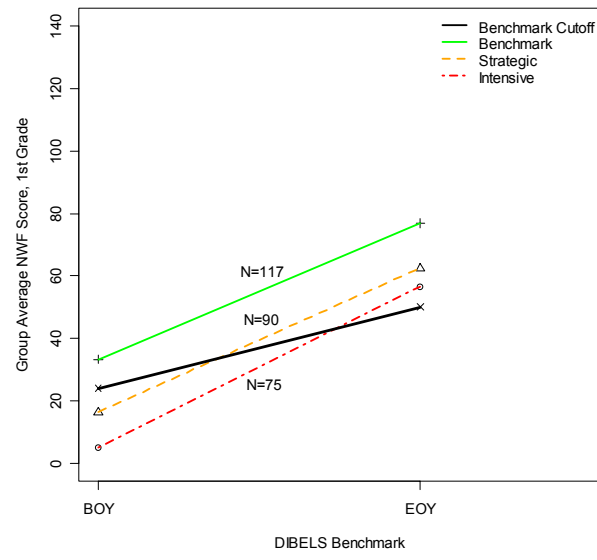
### Cohort 1

Cohort I: Aggregate NWF, 1st Grade

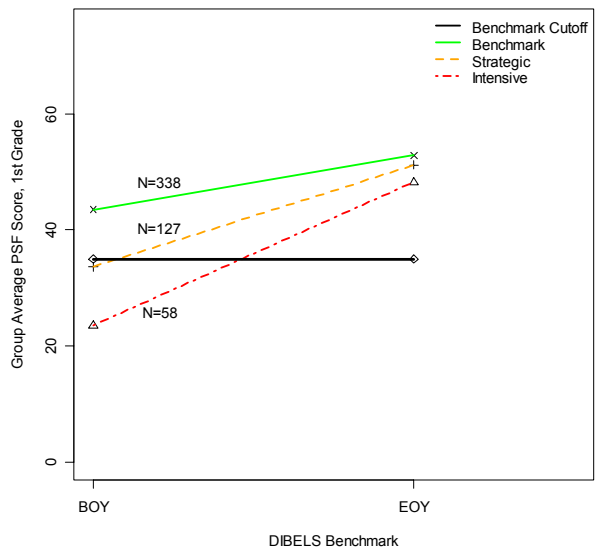


### Cohort 2

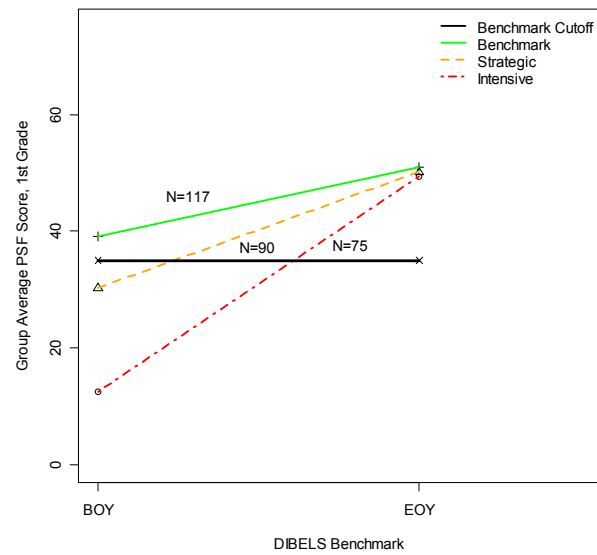
Cohort II: Aggregate NWF, 1st Grade



Cohort I: Aggregate PSF, 1st Grade

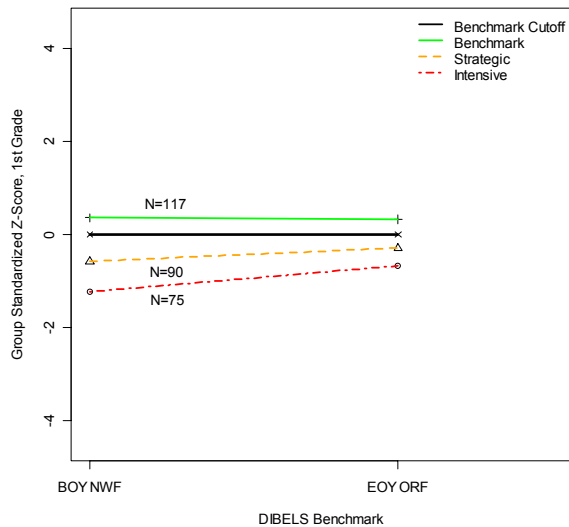


Cohort II: Aggregate PSF, 1st Grade



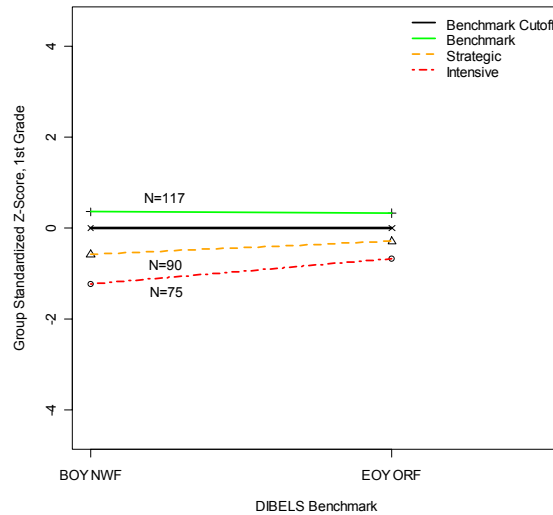
### Cohort 1

Cohort II: Aggregate Standardized Scores BOY v. EOY, 1st Grade

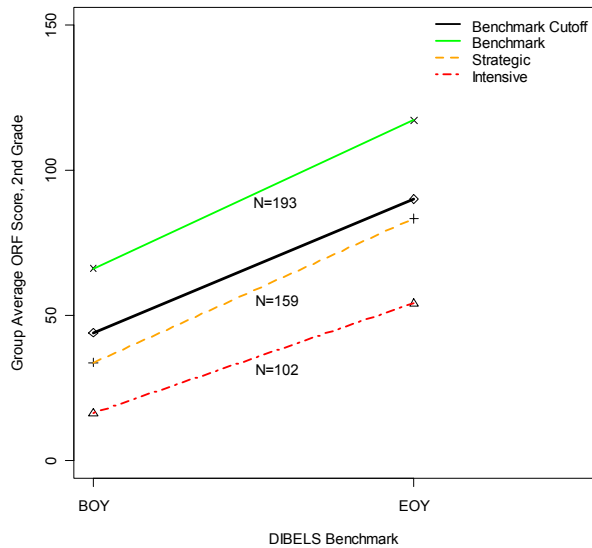


### Cohort 2

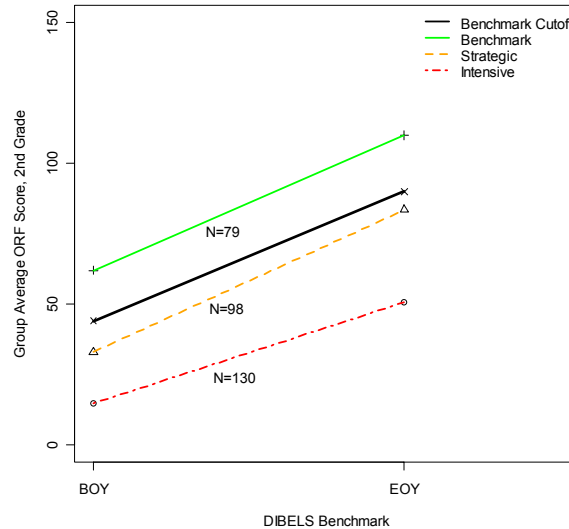
Cohort II: Aggregate Standardized Scores BOY v. EOY, 1st Grade



Cohort I: Aggregate ORF, 2nd Grade

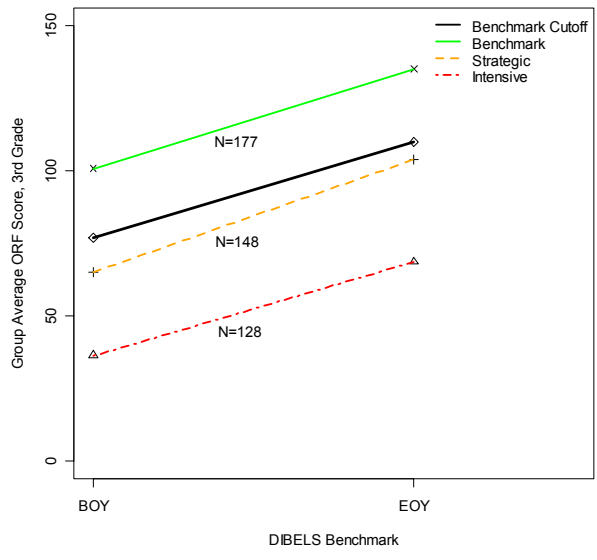


Cohort II: Aggregate ORF, 2nd Grade



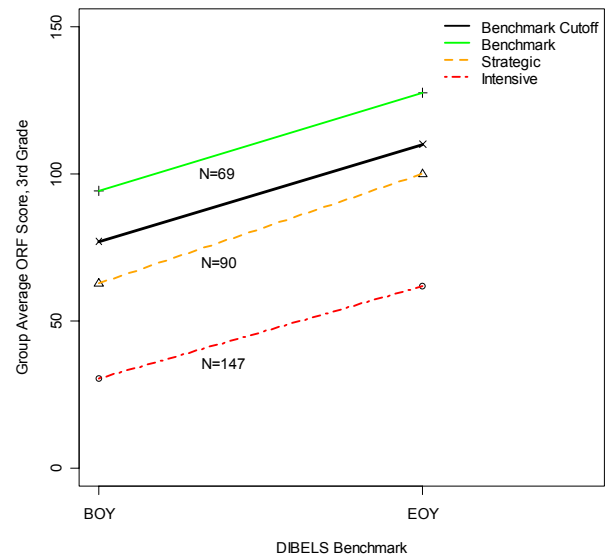
### Cohort 1

Cohort I: Aggregate ORF, 3rd Grade



### Cohort 2

Cohort II: Aggregate ORF, 3rd Grade



Evaluation Question 4. How successful are professional development activities in changing teachers' instructional practices in reading?

As the previous figures indicate, most Cohort 1 schools demonstrated greater gains during the second full year of implementation when compared to the first year. Cohort 2 schools made similar gains in their first year of implementation to Cohort 1's second year. This greater increase in outcomes is largely attributed to better and more targeted professional development and technical assistance provided by the BIE's Reading First Leadership and other technical assistance providers, and improved reading instruction as measured by (a) increased fidelity of implementation in the classrooms (e.g., reading program implementation) and at the school level (e.g., leadership, professional development), (b) increased use of effective teaching practices (i.e., resulting in increased academic learning time and decreased observable off-task behavior), and (c) increased use of assessment data to plan instruction (e.g., DIBELS progress monitoring, curriculum-based measures).

The following table shows the overall averages for BIE Reading First classroom observation measures based on aggregate ratings of use of effective teaching practices and fidelity of implementation. Cohort 1's lack of gain between years 2 and 3 of Reading First implementation is likely due to high staff turnover at many of the BIE's Reading First schools, and to some lingering resistance among some teachers. However, the 3.9 average exceeds the aggregate 3.0 average for the measure was designed. That aggregate 3.0 average holds in other projects which use the same or similar indicators, which means that the BIE's Reading First teacher, overall, demonstrate higher levels of use of effective teaching practices than teachers in many other classrooms in which we conduct observations throughout the country—and fewer of these other classrooms are in such high poverty areas.

**Table 5: Use of Effective Teaching Practices and Fidelity of Implementation**

<b>Cohort</b>	<b>Indicator</b>	<b>End of Year 1 Average</b>	<b>End of Year 2 Average</b>	<b>End of Year 3 Average</b>
Cohort 1	Use of effective teaching practices (5-point scale, 5 high)	3.4	4	3.9
Cohort 1	Classroom level fidelity of implementation (4-point scale, 4 high)	2.3	3.4	3.6
Cohort 2	Use of effective teaching practices (5-point scale, 5 high)	3.7		
Cohort 2	Classroom level fidelity of implementation (4-point scale, 4 high)	3.6		

Observations, interviews and survey responses indicate that changes in these ratings can be attributed to

1. overall implementation of Reading First goals and objectives, which has increased goal-setting, communication, and use of data within and across schools,
2. ongoing professional development and technical assistance activities, and

### 3. increased accountability brought by Reading First.

Observation, survey and interview data continue (similar to the 2005-2006 school year) to indicate that continued challenges to implementing Reading First goals and objectives include

- Reading First staff who are resistant or not committed,
- Poor fidelity of implementation,
- Limited academic learning time,
- Teaching students who have limited English proficiency or who receive special services,
- Staff and student absences,
- Staff turnover,
- “Timely” technical assistance and professional development with reduced time away from schools,
- Understanding and using data appropriately, and
- Communication.

In schools where student outcomes demonstrated the greatest increases in the percentages of students reading on grade level, teachers stated that they were adequately prepared to implement the concepts and they reported a high level of satisfaction with professional development activities coordinated at the BIE and school level. Fewer teachers this year than in previous years reported dissatisfaction with professional development activities (i.e., “It went over my head”) and felt less prepared to teach reading, implement the core reading program, assess student progress, manage student behavior, or implement other concepts addressed in training. However, some lingering resistance to systematic and explicit instruction or to exclusively using the Comprehensive Reading Programs (CRP) selected by a school continues in both Cohort 1 and Cohort 2 schools.

In particular, additional needs for training requested by Reading First staff across schools—including many of those schools with high reading outcomes—included

- Implementing templates and lesson plans developed by the Western Regional Reading First Technical Assistance Center (WRRFTAC),
- Implementing other recently adopted supplemental and intervention reading programs,
- Systematic and explicit (i.e., effective) instruction,
- High levels of interaction with frequent responses from students,
- Efficient and effective use of time during the reading block,
- Differentiated instruction to meet the needs of all students,
- Behavior management,
- Planning, providing and scheduling additional intervention for struggling readers (e.g., double and triple dosing),
- Meeting the needs of students with disabilities and students learning English, and
- Using progress data for instructional planning and student grouping.

Finally, changes to on-site technical assistance during the second year and continuing during the third year of Reading First implementation had a strongly positive impact on schools where the on-site technical assistance was regular and consistent. To the extent that the same provider consistently conducted regularly scheduled school visits, schools reported that this assistance, more than any other professional development or assistance provided under Reading First, impacted implementation of reading programs and teaching. Some schools continue to report “mixed messages” or have received information that is inaccurate from some technical assistance, and this miscommunication has resulted in frustration and inappropriate implementation of WRRFTAC templates and lesson maps (i.e., using them only with intensive students, while they were designed for strategic and lower performing benchmark students).

Evaluation Question 5.           How do LEA choices regarding assessment measures, embedded in core programs, impact schools’ and LEAs’ use of assessment information when planning instructional approaches?

Few Cohort I schools implemented regular progress monitoring using the DIBELS measures during the 2004-2005 school year, but most schools implemented regular progress monitoring during the 2005-2006 school year, and all Cohort 1 schools that received continued funding for the 2006-2007 school year conducted regular progress monitoring on recommended schedules. Additionally, it is clear that at these schools, progress monitoring data based on the DIBELS and curriculum embedded assessments were regularly discussed in grade level meetings with reading coaches (and administrators in most schools), and instructional planning decisions were based on this data. Most Cohort 2 schools conducted regular progress monitoring during their first year of implementation, but not all regularly used the data to plan instruction. Again, those schools in which students demonstrated the greatest progress were mostly likely to administer and review data from DIBELS progress monitoring and curriculum-based measures regularly, particularly with regard to struggling students.

In most schools, regular progress monitoring was conducted by teachers. In these schools, teachers were more likely to be involved in instructional planning and grouping decisions to maximize opportunities for students to learn and to provide needed intervention. These schools clearly showed the greatest student progress. In some schools, reading coaches and other specialists administered most progress monitoring assessments. To the extent that they involved teachers in reviewing the data and planning instruction, student progress was greater compared to schools where reading coaches or other staff administered the DIBELS but did not systematically and regularly review data with teachers.

Most schools regularly administered curriculum-embedded reading assessments, particularly fluency and comprehension measures. Yet, some teachers in schools where student progress was poor admitted that they rarely assessed students because it took too much time. Additionally, many teachers reported that they struggled to use the student progress data to plan instruction and group students.

Based on observations and data from school staff, the top performing schools based on DIBELS percent progress were most effective in reviewing and using data for instructional planning. Evaluators were able to observe grade band meetings at Chief Leschi, which were very

effectively managed by the reading coach with full participation from teachers. Teachers at Jemez showed evaluators examples of student data linked to instructional planning and intervention. At Nay Ah Shing, substantial evidence of using progress monitoring data to plan intervention was available, with DIBELS progress monitoring conducted regularly and frequently. At Mandaree, a student data wall was regularly discussed by staff, with a focus on students who were not making sufficient progress. These are only a few examples of the many schools regularly and appropriately using data for instructional planning and intervention.

Evaluation Question 6.           What specific changes are being made at the classroom, school, and LEA levels as a result of Reading First formative and summative evaluations?

The BIE's Reading First Leadership (e.g., Lynann Barbero, Verla LaPlante, and Casey Sovo) conducted site visits to BIE Reading First schools to provide formative feedback. However, most Cohort 1 schools did not receive these site visits during the 2006-2007 school year. Some of the Cohort 1 schools struggled without the additional technical assistance, as evidenced by five of them not meeting the criteria for continued funding. Data from end-of-year surveys and interviews at most Cohort 1 schools indicated a desire and perceived need for continued assistance from these BIE Reading First leaders.

Overall, schools reported both satisfaction with the assistance and feedback provided by the leadership team, and frustration at "mixed messages" they felt occurred. This concern continues to be expressed by a few Cohort 1 staff, and by staff at approximately 50% of Cohort 2 schools. Additionally, staff at some schools experienced difficulties in agreeing on the messages contained in feedback and steps that should be taken to address issues. Clear written feedback helped prevent varying interpretations of recommendations, and helped to provide clear documentation of expectations.

The external evaluation team also provided formative feedback to schools during and after site visits. In the first year of Reading First implementation for both cohorts, staff at many schools seemed to feel threatened by the external evaluation team and were less cooperative during evaluation site visits and classroom observations. By at this point, all schools were cooperative and valued formative feedback. In particular, reading coaches were receptive to recommendations concerning teachers' need for assistance with classroom management, use of center-based activities, increased efficiency during instructional time to maximize learning opportunities, and the need for greater levels of systematic and explicit instruction. Teachers and paraprofessionals appreciated the immediate feedback contained in the classroom observation feedback forms given to teachers after observations, such as the one shown below (copies provided during site visits to teachers and principals). In a number of instances, teachers tracked down evaluation team members to make sure they received their feedback form.

One additional point deserves mention here. In our first 2 years of evaluating the BIE's Reading First schools, many teachers would either be absent during our site visits, or they had scheduled an activity not aligned with their reading program or instruction (which we request to observe), such as testing students, playing teacher-developed vocabulary building games, or working 1-on-1 with a student while other students worked independently on coloring or silent reading

activities. During the 2006-2007 school year, we were able to observe instruction during the scheduled reading block in a much higher percentage of classrooms, with approximately 5% of observations resulting in no data about instruction (down from a high of nearly 20% during our initial observations in the first year of implementation).

**Figure 10. Classroom Observation Feedback Form**

	EndVision Research & Evaluation dibels@endvision.net
When visiting your classroom today, I saw...	
☺ <i>Excellent use of modeling, e.g., "My turn, your turn"</i>	
☺ <i>ALL students responding during group responses, with appropriate pacing and management of student behavior</i>	
☺ <i>Well prepared lesson with quick pacing, many short transitions, and fluent use of word templates during lesson</i>	
Opportunities for students to learn can be increased by...	
<i>Praising appropriate behavior to set expectation for all: "I like the way Dakota is sitting quietly with her eyes on me and listening to my instructions."</i>	
<u><i>Catherine Heusser</i></u> Signature	<u><i>08-24-06</i></u> Date

Evaluation Question 7. Are referral rates to Special Education changing in Reading First schools; particularly referrals linked to reading failure in the K-3 grades?

Although the quantitative evidence available to answer this question was incomplete and not available for many schools, substantial anecdotal information was collected. For example, staff



at Chief Leschi reported striving to increase children's opportunities for reading success and limiting referral to Special Education or decreasing the special services required previously. The Reading Coaches at Chief Leschi and Hannahville reported that only one student was referred this past school year in Reading First grades. Almost all schools report that classroom behavior problems have decreased, accompanied in most schools by a decrease in Special Education referrals.

Turtle Mountain Elementary School, the largest school funded under the BIE's Reading First, has provided the following results regarding Special Education referrals and placements. Notice both the dramatic drop in assessments and placements, and particularly the reduction in the discrepancy between assessments and placements. Turtle Mountain has approximately 650 students in grades K through 5.

**Table 6. Special Education Referrals and Placements at Turtle Mountain**

<b>School Year</b>	<b>Students Assessed</b>	<b>Students Placed</b>
2003-2004	52	31
2004-2005	31	13
2005-2006	13	4
2006-2007	2	2

Evaluation Question 8. Are exit rates for English Language Learner (ELL) students changing in Reading First schools?

The answer to this question cannot be determined given the quantitative data provided or available to external evaluators. However, many of the BIE Reading First schools have high percentages of English Language Learners, and most schools demonstrated substantive progress in reading outcomes. For example, Chinle and Jemez report high percentages of ELLs, and for many of the teachers and students at these schools, Navajo or Towa is the primary language spoken at home. Yet these schools and others in the 4-corners area made substantive gains in reading outcomes, indicating that ELL students are more likely to be promoted and/or on grade level for reading. However, as mentioned previously, overall gains on SAT-10 measures are lower for schools in which a large percentage of students enter kindergarten speaking a language other than English, or speaking little at all.

## Alignment and Coherence

Evaluation Question 1. Are there specific changes in pedagogical approaches across reading instruction by those identified by the BIE's Reading First program?

Teachers are clearly changing pedagogical approaches in the BIE's Reading First schools across both cohorts. The additional professional development activities provided under Reading First, as well as the levels of accountability required under Reading First and the presence of the external evaluation team, the Reading First leadership team, and other external technical assistance providers created pressure for change. Change was observed and measured in a large percentage of classrooms (50-90% visited per site visit per school), as previously shown, particularly in management of student behavior and implementing activities to keep students engaged. As shown previously, the increased use of effective teaching practices and fidelity of implementation of reading programs were noteworthy (see Table 5). These changes were also reflected in improved student outcomes.

Additionally, many schools report spillover effects from Reading First to other content areas and other grades. In particular, systematic and explicit teaching strategies, such as those included in the WRRFTAC templates, are used by teachers to teach math and science concepts. Classroom management strategies, higher rates of student-teacher interaction to check understanding, modeling, and increased use of praise are also specific examples of pedagogical changes reported in surveys during both reading instruction and instruction in other content areas.

Evaluation Question 2. Are there changes in existing standards and/or benchmarks that improve alignment to the BIE's Reading First approach?

Staff at all BIE Reading First schools reported that state (i.e., often both the state in which the school was located as well as BIE) standards were addressed or met. Review of Comprehensive School Reform Plans and Reading First proposals during the 2005-2006 school year indicates an increased awareness of and ability to articulate reading goals and activities across time in school that provided these documents across years. Increased use of DIBELS benchmark assessments and on-going progress monitoring using the DIBELS and program-embedded measures, as well as regular meetings and data-based decision-making regarding student grouping and instructional planning, demonstrate improved alignment to Reading First goals and objectives.

Two other stories are worth telling here. In a Cohort 1 Reading First school, the principal regularly visits classrooms to observe instruction. He admits he "does not know much" about reading instruction in the elementary grades. Yet, during one observation, he watched a teacher inappropriately implementing a WRRFTAC template. When he was sufficiently frustrated with her lack of fluency in implementing the template, he asked if he could finish teaching that part of the lesson, which he did!

In another Cohort 1 Reading First school, teachers report that the principal conducts classroom walkthroughs almost every day. In fact, nearly 25% of the teachers made statements parallel to

the following: “He’s in my class every day. Well, I think he observes my classroom nearly every day. He visits so often, I don’t even notice anymore when he’s here.”

Evaluation Question 3. Are other non-Reading First schools and/or districts adapting Reading First priorities?

According to the BIE’s Reading First Director, other BIE schools were invited to most Reading First professional development meetings. Additionally, Reading First staff at most BIE Reading First schools reported attending state (e.g., AZ, WI, ND) Reading First professional development activities. In fact, many of the BIE’s Cohort 1 reading coaches are regularly invited to present at local and state meetings, indicating a strong external interest in the Reading First activities at the BIE’s Reading First schools.

Evaluation Question 4. Are there shifts in LEA instructional practices in other content areas as they relate to reading instruction?

In many schools, substantial changes were made because of Reading First goals and priorities. Establishing and enforcing a 90-minute or more reading block was a substantial change for most schools. The focus on systematic and explicit instruction, selection of activities aligned with the five key components of reading instruction, and providing differentiated large-group instruction for longer periods of time (i.e., as recommended by the Houghton Mifflin reading program) was new for many teachers. Yet, many teachers reported feeling successful using these instructional practices and stated that they were implementing the same approaches in teaching other content, particularly math.

Additionally, BIE Cohort 1 K-5, K-6, K-8, or K-12 school systems have made changes to better align reading and math programs throughout the grades to Reading First goals and objectives. For example, implementing a single reading program across more grades, establishing lengthier blocks of time dedicated to reading instruction across grades, and providing on-going targeted intervention outside of established reading time for struggling readers are just a few of the changes reported. Implementing math programs using a Reading First-like model was occurring or planned at almost all Cohort 1 schools. Reading coaches for grades beyond K-3 and math coaches have been appointed at most successful BIE Reading First schools with more than one classroom per grade. Overall, the focus on No Child Left Behind and the success of Reading First in many of these schools has changed instructional approaches, including use of student data to drive instruction.

Some schools were discussing or reported even greater changes. For example, at Chief Leschi where the dropout rates were high and many upper-grade students were reading well below grade level, a high school reading program with intensive intervention was established rather than teaching American and British Literature classes to students who were unable to read on grade level. Chief Leschi also hired reading coaches for grades 4-6 and 6-12, and established a K-8 aligned math program with a math coach. Turtle Mountain, Alamo Navajo, Hannahville, Nay Ah Shing, Indian Island, Lac Courte Orielles, and other schools reported establishing math programs across contiguous grades and encouraging teachers to work together to implement the

math programs to create cohesive and integrated instruction across multiple grades with instructional planning based on student progress.

Clearly, the successes of Reading First have driven decisions to establish Reading First-like models in upper grades and in other content areas in most BIE Reading First Cohort 1 schools. Cohort 2 schools are also planning implementation of similar changes. The most frequently mentioned “spill-over” successes are the following:

- Spread of reading programs and DIBELS assessments to 4th-6th or 4th-12th grades
- Increased use of intervention programs and intervention models in upper grades
- Spread of “Reading First”-like goals, programs, instructional strategies, and assessments to other content areas, e.g., writing/language arts, math, science
- Increased instructional leadership and use of coaches in upper grades

Evaluation Question 5. Are there shifts in staff development programs that reflect the BIE’s Reading First priorities?

The BIE’s Project Director and Reading Specialists have monitored the effectiveness of Institutes of Beginning Reading (IBR) and other staff development opportunities provided by the BIE Reading First program. While the IBRs have been aligned with Reading First priorities, school and teacher readiness played a large role in overall satisfaction with IBR training and transfer into practice. High levels of satisfaction with IBRs and implementation of strategies for which training was provided (e.g., activities aligned with the five key components of reading instruction, effective teaching strategies, evidence from on-going assessment used for instructional planning, action plans, etc.) were reported or observed in most BIE Reading First schools, and particularly schools with (a) an engaged leadership and staff who were involved in implementing Reading First goals and (b) classrooms where students were engaged, positive behavior management was used, teacher-student interaction was high, and routines were established to maximize time for reading instruction and increase learning opportunities.

Based on feedback from school staff, the external evaluation team, and their own observations, the BIE’s Leadership team has continually addressed school recommendations and concerns in planning IBRs and in providing targeted technical assistance. Additionally, during the second and third years of implementation, on-site technical assistance and staff development provided additional assistance to schools categorized as ‘intensive’ or ‘strategic.’ For example, schools where student outcomes were high, regardless of fidelity of implementation, were allowed to select or provide staff development opportunities for the 2005-2006 and 2006-2007 school years. Schools where student outcomes were low, regardless of fidelity of implementation, the BIE Leadership team played a much larger role in providing direction for staff development activities and ensured more regular onsite technical assistance through the BIE’s project leadership and external technical assistance providers. This model of tiered-assistance positively impacted schools that needed assistance most. As mentioned previously though, some Cohort 1 schools that needed ongoing support to continue improving did not receive enough support during the 2006-2007 school year. For the most part, these schools did not receive continued funding for the 2007-2008 school year as they did not meet the criteria for continued funding.

## Capacity Building

Evaluation Question 1. Are teachers able to adapt their new knowledge of instructional strategies for teaching reading to a range of situations, such as working with at-risk learners, ELL students, or students with possible learning disabilities?

One of teachers' most frequent requests for additional training involved how to best manage and teach students who were at-risk or struggling learners, ELL, or identified with disabilities. Clearly, teachers throughout all BIE Reading First schools felt they needed additional assistance to be more successful in teaching these students—and these needs continue to be mentioned in surveys and interviews.

However, there were observable changes across time in teachers' management of struggling readers. In particular, in those classrooms we observed across time, we saw increased engagement of struggling students and use of classroom and instructional strategies to promote opportunities to learn, and decreased levels of severely off-task and disruptive behavior in some classrooms. Additionally, in some observations conducted at the end of the school years, we selected a struggling student who needed intensive intervention in reading to specifically observe. In many classrooms, these students were engaged in instruction at higher levels than we expected from the research literature and from our extensive experience observing classrooms. In some, teachers provided differentiated instruction for these students through more frequent questioning, scaffolding, and instructional feedback, and providing additional opportunities to practice. On the other hand, our observations and interview data clearly support the need for additional training for teachers in positive classroom management and use of instructional strategies for teaching reading, particularly with struggling students who are at-risk learners, ELL, or may have learning disabilities.

Evaluation Question 2. Are Reading First teachers participating in professional activities outside of Reading First programs—providing workshops, attending conferences, or training other teachers and administrators in approaches to reading instruction?

In some schools, Reading First reading coaches and teachers were involved in professional activities to increase use of research-based reading instruction among other Reading First and non-Reading first schools. Again, the staff who reported being most involved in these activities were those who were better implementing reading programs and activities aligned with Reading First goals and priorities. In particular, reading coaches who were more effective and who were able to provide modeling and instructional feedback to teachers also more actively participated in professional activities to promote use of Reading First approaches to reading instruction. Among the successful teachers and reading coaches, the level of enthusiasm and support for Reading First was high, while struggling reading coaches and teachers clearly focused more on their own efforts to improve, particularly given the increased accountability for positive student outcomes associated with Reading First.

Because of the geographical isolation of many of the BIE's Reading First schools, and the number of days they are required to miss school to attend BIE Reading First professional development, many teachers report attending few other professional activities outside Reading First. On the other hand, those teachers and paraprofessionals who provide intervention to students receiving special services, and those who are working towards advanced degrees more frequently report involvement in other professional activities.

Evaluation Question 3. Are pre-service and in-service teacher training programs incorporating Reading First approaches instruction in their reading courses?

With the BIE's unique status as a "state" with Reading First schools spread across 11 states, many pre-service and in-service training programs are available to teachers with the BIE's Reading First schools. As external evaluators, we have not reviewed program materials from these hundreds of programs to determine the extent to which they are increasingly incorporating Reading First approaches to their reading courses. However, we regularly read about Arizona's, New Mexico's and Washington's efforts to promote Reading First and all children reading on grade level. Given the nationwide focus of No Child Left Behind, data we have reviewed from local institutions, and published journal articles and internet communication, there is no question that reading instruction is under scrutiny, which potentially forces changes within institutions of higher education. Yet, we do not have the resources to determine the extent of this change, particularly across the many states and regions in which BIE Reading First schools operate.

## **Additional Findings**

Evaluation Question 1. To what extent can end of year performance on the DIBELS predict SAT-10 performance?

Given the controversy over the appropriateness of the DIBELS as a progress monitoring or outcome measure, we conducted an analysis to determine the predictive validity of the DIBELS when compared to SAT-10. The following table shows the findings. In particular, note that on average, an end of year DIBELS instructional recommendation of benchmark predicts proficiency on the SAT-10 with 79% accuracy. Even more noteworthy is the finding that on average, 92% of students who are intensive based on end of year DIBELS instructional recommendations will be NOT proficient on the SAT-10. A document that includes more complete analyses and findings is included in Appendix D.

**Table 7: Predictive Ability of DIBELS When Compared to SAT-10**

DIBELS End of Year Instructional Recommendation: BENCHMARK	Proficient on SAT-10 (40 <sup>th</sup> Percentile)	DIBELS End of Year Instructional Recommendation: Intensive	NOT Proficient on SAT-10 (40 <sup>th</sup> Percentile)
Grades 1-3 Average	79%	Grades 1-3 Average	92%
1 <sup>st</sup> Grade	86%	1 <sup>st</sup> Grade	89%
2 <sup>nd</sup> Grade	80%	2 <sup>nd</sup> Grade	92%
3 <sup>rd</sup> Grade	67%	3 <sup>rd</sup> Grade	95%

Evaluation Question 2. To what extent do students lose ground over the summer, and when do they catch up again?

Figure 11 shows data for a cohort of students that started Reading First in 2<sup>nd</sup> grade and continued in 3<sup>rd</sup> grade—a matched data set including only students who had scores across all six time points. The students are grouped according to their beginning of 2<sup>nd</sup> grade instructional recommendation. The graph shows the drop in reading fluency that occurs across all three groups over the summer. Note that students do not regain their end of 2<sup>nd</sup> grade average until the following middle-of-year benchmark assessment—7 months later.

Evaluation Question 3. What school-level factors affect student outcomes?

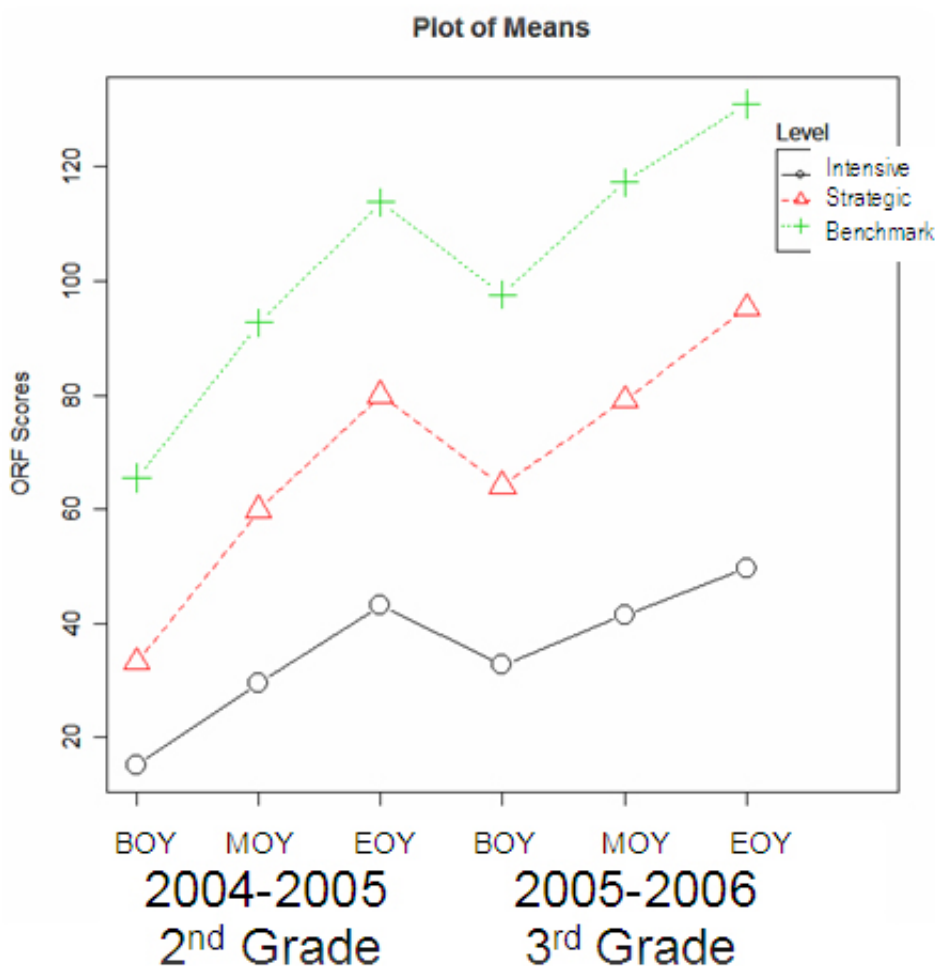
Finally, we are conducting statistical analysis to determine the impact of school-level factors on students outcomes. In preparing for that analysis, the following table was created to start identifying school-level factors that appear to be most related to student outcomes. While the table is not yet complete, it provides a starting place for identifying factors to include in the statistical analysis. Additional factors are being identified based on survey and interview data, and school observations.

In Table 8, the Cohort 1 schools are listed in order of decreasing percent progress for the 2006-2007 school year. The first two columns with checks indicate whether the principal and reading coach have stayed the same during the three years of Reading First implementation. The next column indicates whether the school is using a Direct Instruction program (e.g., Reading Mastery, Corrective Reading) as a replacement core for struggling readers, or whether the WRRFTAC templates and lesson maps are being implemented appropriately (i.e., with students for whom they were designed or with all students, with appropriate amounts of time allocated to each template). The final column lists the reading programs.

The following summarizes the factors that appear to have a substantial positive or negative impact on student outcomes, partly based on the table.

- ❖ Positive
  - Same principal and reading coach both years
  - Engaged leadership (classroom observations)
  - Scripted reading or intervention program (e.g., Reading Mastery, ERI, Voyager Passport), and/or WRRFTAC templates implemented well
- ❖ Negative
  - High staff turnover, particularly at leadership levels
  - Conflict within school
  - Resistance to Reading First or reading programs

**Figure 11: Average ORF Scores by Group over Time**





**Table 8: School-Level Factors Affecting Student Outcomes**

<b>BIE Cohort I Average Percent Progress</b>	<b>67</b>	Same Principal	Same RC	DI Programs/ WRRFTAC Templates	Reading Programs
<b>* Nay Ah Shing</b>	<b>81</b>	✓	✓	✓	RM/RALP
Jemez	<b>80</b>	✓	✓		SFA
Turtle Mountain	<b>75</b>	✓	✓	✓	MH/VP/RM
Dunseith	<b>75</b>	✓		✓	HM/VP
Chief Leschi	<b>74</b>	✓	✓	✓	HB/RM/RW/RN
Hannahville	<b>73</b>	✓	✓	✓	HM/VP
Taos	<b>73</b>	✓	✓	✓	HM/RN/VP
Santa Clara	<b>70</b>		✓	✓	HM
<b>* Alamo Navajo</b>	<b>70</b>	✓	✓-	✓	SF/RN/RM/ERI
Indian Island	<b>66</b>	✓	✓	✓	HM/RM
LCO	<b>65</b>	✓-	✓	✓	HM/RM
<b>* Chinle</b>	<b>61</b>			✓	RM

**\* Started below BIE average for percent Intensive students**

ERI = Early Reading Intervention

HB = Harcourt Brace

HM = Houghton Mifflin

MH = McMillan McGraw Hill

RALP = Reading for All Learners

RM = Reading Mastery

RN = Read Naturally

RW = Read Well

SF = Scott Foresman

SFA = Success for All

VP = Voyager Passports

## **Conclusions**

Overall, the BIE's Reading First program has initiated substantive changes in schools. Observations, interviews, and surveys support claims for the following changes across all BIE Reading First schools.

- Increased engagement of instructional leadership,
- Increased fidelity of implementation,
- Increased progress monitoring with the DIBELS and use of data to adjust intervention,
- Increased knowledge of research-based and effective reading instruction,
- Increased use of effective teaching practices,
- Increased student engagement,
- Increased numbers of students reading at Benchmark or improving based on the DIBELS instructional recommendations and other curriculum-embedded assessments, as well as standardized (e.g., SAT-10) and state-level criterion-referenced tests, and
- Professional development and technical assistance that has changed over time to better meet needs.

Schools in which student outcomes in reading showed the greatest improvement had

- Strong, engaged, and supportive leadership,
- Stable (i.e., high retention), informed and supportive staff with high levels of communication about students and programs,
- High expectations for students,
- Well-implemented research-based reading programs,
- Policies that supported implementation of reading programs,
- Established plans for reading intervention for struggling students, and
- Regular progress monitoring and review of data to plan instruction and intervention.

In schools where student reading outcomes indicated smaller gains,

- Leadership was typically disengaged and teachers felt unsupported,
- Staff turnover was high,
- Resistance to Reading First or implementing reading programs with fidelity was high,
- Many students were observably off-task and/or disruptive,
- Management of student behavior was often negative and punitive, or disruptive behavior was ignored,
- Many activities were observed that were not aligned with the five key components of reading instruction, and/or
- Use of time for reading instruction was inefficient (i.e., lengthy transitions and breaks).

Additionally, the BIE's Reading First schools are better meeting the needs of struggling readers so that they improve reading skills rather than falling further and further behind. Schools report fewer special education referrals and placements, and greatly reduced referrals for behavior.

Some recommendations based on findings are warranted here. First, the BIE's Project Director and Reading Specialists have been quite responsive and reflective in designing staff development activities and recommending targeted technical assistance for individual schools. This process of integrating feedback and reflection has resulted in improved satisfaction over time among staff at Reading First schools, and should be continued as the BIE Reading First Program continues to evolve and improve implementation and integration of Reading First goals and priorities. Additionally, the move towards more targeted assistance to schools, based on evidence from student outcomes and school and classroom observations, strengthened implementation of reading programs across schools, particularly in those schools in which more assistance was needed. Continuing to provide more targeted staff development and technical assistance should help improve student outcomes across schools.

The site visits conducted by both the BIE's Leadership team and the external evaluation team have resulted in changes in school and classroom practices and policies. Feedback to schools from these visits needs to be in writing, with recommendations also reported verbally directly to school administration and staff. During interviews, many administrators, reading coaches, and teachers reported "mixed" messages. While we have evidence that indeed some communication was conflicting and confusing, we also have evidence that the same message delivered by different messengers, or the same message delivered by a single messenger to many staff, results in different interpretations and causes confusion. Again, putting recommendations in writing with distribution to all affected stakeholders is important to reduce "message confusion."

Finally, using high performing schools and effective reading coaches and teachers as models for other schools and teachers seems to have had a strong influence on fidelity of implementation of reading programs and improving teaching practices. In particular, Reading First Cohort 2 staff have expressed appreciation for the modeling of appropriate implementation of reading programs and teaching strategies, systematic and explicit instruction, efficient and effective time management, and positive management of classroom behavior provided by the Cohort 1 coaches.

## Appendix A: BIE Reading First Evaluation Questions

### *BIE Reading First Subgrant Evaluation Questions*

Subgrant-specific program evaluation criteria include those stated in Appendix 1 of the BIE Reading First proposal (BIE/OIEP, 2003, pp. 68-73) and adapted below.

#### **Administration, Organization, & Communication**

Are administrators and instructional leaders implementing practices that help all student meet Reading First goals and objectives and that are consistent with research-based effective schools and leadership practices?

1. To what extent are all ongoing building-level program development and program evaluation documents, including the Comprehensive School Reform Plan, aligned with Reading First program goals, activities, and evaluation criteria listed in these evaluation questions?
2. Have administrators had access to and attended training on the research-based reading curriculum building blocks: phonemic awareness, phonics, fluency, vocabulary development, and text comprehension?
3. Have administrators had access to and attended training on current scientifically-based reading research (e.g., instructional strategies for teaching blending), including those items listed in the scoring rubric for subgrant applications?
4. In what ways are administrators (a) leading staff (including resource specialists and paraprofessionals) to create a comprehensive school-wide plan for reading instruction based on Reading First goals and priorities, (b) instituting practices to support the schools' reading goals, and (c) communicating Reading First goals with teachers, parents, and other stakeholders?
5. To what extent are administrators maximizing and protecting instructional time and organizing resources and personnel to support reading instruction, practice, and assessment?
6. To what extent have differentiated staffing teams been established and supported to analyze student reading outcomes and plan and implement immediate instructional adjustments?
7. To what extent are educators (a) using student outcome data to make instructional decisions that improve the coordination of instruction across time and between instructional providers, and (b) ensuring that receiving teachers have the needed on-going documentation of individual progress and instructionally relevant information on past successful and unsuccessful interventions?
8. To what extent is concurrent instruction (e.g., ESL, Special Education) coordinated with and supplementing regular classroom reading instruction?
9. Through what mechanisms is student performance reported to teachers, parents, and other stakeholders?
10. How often do instructional leaders meet with K-3 teachers to review progress data on all students, evaluate, and if necessary, modify the instructional programs for students not considered "on grade level"?

## Goals, Objectives, & Priorities

Are measurable goals and objectives aligned with Reading First used to guide instruction, assessment, and professional development?

11. Are measurable goals and objectives for each grade level in the K-3 reading program available? Can Reading First staff articulate the goals and objectives for lessons?
12. Are goals and objectives prioritized and organized by the dimensions of reading: phonemic awareness, phonics, fluency, vocabulary, and text comprehension?
13. In what ways is progress in meeting goals and objectives evaluated by an analysis of available data (e.g., DIBELS scores, curriculum-based assessments) on student outcomes?
14. In what ways is progress data being used for instructional planning to ensure goals and objectives are met and to systematically and progressively improve student outcomes?
15. In what ways is student progress data and other classroom-level data being analyzed to provide professional development, including coaching and modeling, to systematically and progressively improve teaching practices to increase progress towards meeting goals and objectives?

## Assessment

Is ongoing and consistent assessment using valid and reliable measures that are appropriately administered used to monitor progress, adjust instruction, and report outcomes?

16. To what extent has a school-wide assessment system been established and maintained for documenting student performance and monitoring progress at individual, classroom, and building levels?
17. To what extent are measures being used to provide formative and summative information on student performance on phonemic awareness, phonics, fluency, vocabulary, and text comprehension?
18. To what extent has the technical adequacy of the measures been established?
19. Have all users had access to and received training and follow-up on measurement administration, scoring, and data interpretation?
20. To what extent are assessments used with all students, as prescribed and defined in the BIE Reading First student assessment plan requirements, to ensure systematic and progressive instructional improvements for all students?
21. Does the building have a “resident” expert to maintain the assessment system and ensure that measures are administered reliably, data are scored and entered accurately, and feedback to instructional teams and administrators is provided in a timely fashion?
22. In keeping with SEA reporting requirements for all BIE schools, are reports of reading achievement data disaggregated by low-income, major racial/ ethnic groups, LEA, and special education for K-3 students?

## **Instructional Programs & Materials**

Was a validated process that involved stakeholders (e.g., administrators, teachers, other Reading First staff) used to select core, supplemental, and intervention programs that are based on scientifically-based reading research and that have documented evidence for improving student achievement?

23. Was a validated process, based on scientifically-based criteria, used to select instructional materials?
24. In what ways were school administrators and staff involved in the selection process? In what ways does their level of involvement affect implementation of the program?
25. Are the core and intervention instructional programs adopted and implemented in K-3 classrooms (a) research-based, (b) aligned with Reading First curriculum goals, (c) aligned with scientifically-based research on effective, preventive reading pedagogy with diverse learners, and (d) showing documented evidence for improving student achievement?
26. Are additional, intensive instructional programs, with documented evidence of improving student achievement for struggling readers, used for intervention and remediation of students who do not demonstrate adequate progress in the core program?

## **Differentiated Instruction, Grouping, Staffing, & Scheduling**

Are school staff, schedules, instruction, and students organized in ways consistent with the effective schools literature and with the goal of helping all students meet Reading First goals and objectives?

27. To what extent are cross-class and cross-grade homogeneous groupings used, when appropriate, to maximize learning opportunities?
28. Are group sizes and membership dynamic and based on continuous monitoring of individual student performance?
29. In what ways are double- and triple-dosing, and after-school tutoring used to supplement (not supplant) explicit teacher-directed instruction?
30. To what extent are instructional programs, group sizes, and instructional time determined by and adjusted according to learner performance (e.g., students with greatest needs are in groups that allow more frequent monitoring and opportunities to respond and receive feedback)?
31. If aides are members of differentiated staffing teams, (a) are they linked to the language, culture, and economy of the local communities, (b) what functions do they perform, and (c) what professional development was provided?
32. Are diagnostic assessment tools and instructional practices that address academic needs and that ensure a timely, successful, instructional experience used with mid-year entry students?

## ***BIE Reading First Program Evaluation Questions***

### **Program Effectiveness**

Is the overall implementation of Reading First goals and objectives within the school effective in improving student achievement for all students, including students learning English and students with special needs?

1. What gains are students making in grades K-3 as a result of Reading First programming?
2. In what ways are professional development activities changing teachers' instructional practices in reading so that they are aligned with research-based reading instruction and effective teaching practices?
3. What reading assessment measures are used and how are students' scores on the assessments used when planning instruction?
4. What specific changes are being made at the classroom and school levels as a result of Reading First formative and summative evaluations?
5. Are referral rates to Special Education changing, particularly referrals linked to reading failure in the K-3 grades?
6. Are promotion rates for students, particularly English Language Learners, changing in Reading First schools?

### **Alignment & Coherence**

Are reading and other content area instruction, teaching practices, staff development practices, and standards/benchmarks changing to be more aligned with Reading First goals and objectives?

7. Are research-based pedagogical approaches (e.g. systematic and explicit instruction) used by Reading First staff during reading instruction?
8. What shifts in instructional practices are occurring based on reading instruction and effective teaching practices?
9. Have changes in existing standards and/or benchmarks occurred that improve alignment to the Reading First goals and objectives?
10. Are shifts in staff development programs occurring that reflect the BIE's Reading First priorities?

### **Capacity Building**

Are Reading First staff able to apply research-based Reading First practices with diverse learners and to help people outside of Reading First learn and apply Reading First practices?

11. Are teachers able to use research-based reading instruction and effective teaching practices to help at-risk learners, ELL students, or students with possible learning disabilities make reading progress?

12. Are Reading First teachers participating in professional activities outside of Reading First programs—providing workshops, attending conferences, or training other teachers and administrators in approaches to reading instruction?



## Appendix B: Research-Based Effective Teaching Practices

Adapted from Hofmeister, A. M., & Lubke, M. (1999). *Research into practice: Implementing effective teaching strategies*. Boston, MA: Allyn & Bacon.

### **Classroom Management**

#### **Skill 1. Instructional Strengths**

Well-planned, strong, durable procedures are used to teach the course content.

##### Evaluation Questions

1. Are time management procedures, such as brisk pacing and effective transitions, used?
2. Are functions, such as checking prerequisites and guided practice, ensuring successful student management?
3. Are effective academic feedback and academic monitoring skills in use?
4. Does the teacher avoid being distracted from instruction by student misbehavior?
5. Do good preparation and planning help maintain instructional momentum?

#### **Skill 2. Setting and Implementing Rules**

The teacher uses rules and related procedures to prevent problems and manage the classroom in a businesslike manner.

##### Evaluation Questions

1. Does the teacher provide a set of rules at the beginning of the course?
2. Do the rules specify behaviors needed for productive, instructional interactions?
3. Does the teacher demonstrate a willingness and ability to act when rules are broken?
4. Does the teacher establish and interpret rules to develop student support for the rules?

#### **Skill 3. Prevention**

The teacher effectively intervenes to prevent and reduce misbehavior.

##### Evaluation Questions

1. Does the teacher demonstrate increased vigilance at appropriate times?
2. Are interventions timed to avoid disrupting the flow of the instruction?
3. Do teacher reactions to misbehavior model appropriate social interactions?
4. Does the teacher effectively use physical placement to monitor students and intervene effectively?

#### **Skill 4. Appropriate Behavior Recognized**

The teacher clearly identifies and recognizes appropriate academic and classroom social interaction skills.

##### Evaluation Questions

1. Are the important classroom social interaction skills clearly identified?
2. Is teacher praise contingent, i.e., clearly linked in time and action with the student behavior?
3. Does teacher praise leave no doubt as to who and what action is being recognized?
4. Is teacher praise delivered in a credible manner?

### **Skill 5. Limited Reprimands**

Reprimands are limited in number and effectively used.

#### Evaluation Questions

1. If reprimands are used, are they contingent, specific, and credible?
2. Is there any evidence to suggest that academic errors are being equated with misbehavior?

## ***Academic Feedback***

### **Skill 1. Feedback Opportunities**

A classroom environment has been created that provides for extensive academic interactions between teacher and students.

#### Evaluation Questions

1. Have procedures been used to ensure that a large amount of time is allocated to academic instruction?
2. Do lessons include appropriate amounts of guided practice and daily reviews?
3. Is new material presented in small steps with large amounts of academic feedback?
4. Depending on content, are appropriate amounts of oral and written feedback used?

### **Skill 2. Questioning**

The questions are consistent with the instructional needs.

#### Evaluation Questions

1. Are student success rates appropriate for the lesson activity?
2. Do the questions support the presentation of new content in small steps?
3. Are questions to individuals posed before the individual is named?
4. Do questioning procedures maintain instructional momentum?

### **Skill 3. Student Responses**

Individual responses, group responses, and written responses are used to ensure high levels of involvement from all students.

#### Evaluation Questions

1. Is the teacher blending choral and individual responses where their use is appropriate?
2. Are all the students being equally involved during individual questioning?
3. When appropriate does the teacher require written responses to the most important skills?

### **Skill 4. Reacting to Student Responses**

Teacher reactions are consistent with student responses to questions.

#### Evaluation Questions

1. Does the hesitant, correct response typically receive stronger praise and a quick review?
2. For incorrect responses for lack of knowledge, does the teacher rephrase the question or reteach?
3. Do correction procedures indicate the use of elegant rules and practical problem-solving strategies?
4. Do teacher's responses to student errors indicate an atmosphere where students are not afraid to make errors?

### **Skill 5. Question Clarity**

Questions are clearly framed and clearly delivered.

#### Evaluation Questions

1. Are questions short and precise, or rambling and disjointed?
2. Are questions delivered clearly and audibly?
3. Are questions clearly aligned with the content focus of the lesson?
4. Is student attention gained before questions are posed?

## ***Teaching Functions***

### **Skill 1. Daily Reviews**

Daily reviews and prerequisite checks are conducted systematically and effectively.

#### Evaluation Questions

1. Do most lessons include a short review of skills taught in previous lessons?
2. Does the review include an assessment of the level of mastery of skills reviewed?
3. Do you often get well into the body of a lesson and find you have to reteach prerequisites?
4. On reviews, do you immediately reteach if less than 80 percent demonstrate mastery?

## **Skill 2. New Content**

A major part of most lessons is set aside for the effective acquisition of new content.

### Evaluation Questions

1. Is extensive guided practice integrated into demonstrations and lectures on new content?
2. Is approximately half the lesson time devoted to guided practice, demonstrations, and presentations of new content?
3. Do you monitor guided practice and conduct reteaching if necessary?
4. Do you revise instruction if students consistently experience difficulty?
5. Is the material taught in small steps, with performance checks after each step?

## **Skill 3. Independent Practice**

Independent practice is consistently and effectively conducted.

### Evaluation Questions

1. Are the students achieving better than 80 percent before moving from guided to independent practice?
2. Is a minimum of 20 minutes of independent practice assigned on most days in each major subject area?
3. Do students receive prompt feedback on independent practice?
4. Is independent practice extensive enough to achieve “overlearning”?

## **Skill 4. Comprehensive Mastery Testing**

Mastery testing is conducted systematically and diagnostically.

### Evaluation Questions

1. Is comprehensive mastery testing conducted at least twice per month?
2. Does some mastery testing facilitate the review of material covered several weeks previously?
3. Is the mastery testing capable of diagnosing specific skill deficits in individual students?
4. Are the mastery testing results being used to guide reteaching?

## **Skill 5. Homework**

Homework is managed appropriately.

### Evaluation Questions

1. Does the homework serve to consolidate skills already taught?
2. Is the student error rate low?
3. Is the homework corrected promptly?
4. Is the homework highly aligned with daily lessons?

## **Academic Monitoring**

### **Skill 1. Assignment Clarification**

Students understand what is expected of them.

#### Evaluation Questions

1. Are students required to demonstrate that they know how to accomplish assignments?
2. Are students taught how to use self-monitoring procedures to evaluate their assignments?
3. Are assignments written out rather than presented orally?
4. Is there a clear procedure for communicating with parents regarding homework assignments?

### **Skill 2. Assignment Follow-up**

Students are expected to complete their work according to established standards and deadlines.

#### Evaluation Questions

1. Are the format requirements of assignments in terms of length, neatness, and accuracy specified?
2. Are the students required to make corrections to inaccurate or incomplete assigned?
3. Does the grading of assignments reward accurate and timely completion?
4. Do assignment correction procedures diagnose subskill deficits?

### **Skill 3. Seatwork Monitoring**

During seatwork, the teacher circulates around the classroom, checking all students' work and providing assistance to students.

#### Evaluation Questions

1. Does the physical layout of the room facilitate easy movement among students' desks?
2. Are random selections of students' work checked during independent seatwork?
3. Are students encouraged to recognize when they need help and ask for it?

### **Skill 4. Daily Monitoring**

Students' work products are monitored on a daily basis and rates of learning and levels of understanding or performance monitored.

#### Evaluation Questions

1. Are cues from students used to modify instruction during a presentation?
2. Are students required to demonstrate understanding during instructional presentations?
3. Does teacher recordkeeping monitor individual skill acquisition?
4. Are worksheets checked in an accurate and timely fashion?

## **Skill 5. Instructional Alignment**

Academic monitoring reflects alignment among curriculum, instruction, and testing.

### Evaluation Questions

1. Do comprehensive tests of student mastery adequately sample the instructional content?
2. Do comprehensive tests of student mastery adequately sample the required curriculum?
3. Are test results used to adjust instructional procedures for the class as a whole?
4. Does the instruction focus on the curriculum?

## ***Time Management***

### **Skill 1. Allocated Time**

The maximum possible time is allocated for direct intensive instruction.

### Evaluation Questions

1. Are the time allocations for class changes and breaks unnecessarily long?
2. Are non-curricular activities taking time could be allocated to curricular activities?
3. Do the time allocations reflect teacher interests rather than student needs?
4. Is allocated time scheduled to ensure that continuity and systematic review are facilitated?

### **Skill 2. Engaged Time**

A high percentage of allocated time is spent “on-task” by students.

### Evaluation Questions

1. Does the lesson start quickly and smoothly?
2. How long after the start of a lesson did it take to have all students on-task?
3. Are large amounts of the allocated time being taken up with managerial tasks?
4. Is transition time for lesson activities excessive?
5. Is there a reduction in instructional intensity near the end of the lesson?

### **Skill 3. Individual Engaged Time**

All students, low and high achievers, are on-task.

### Evaluation Questions

1. Is the teacher moving about the classroom, actively checking on all students?
2. Do the teacher’s questioning procedures suggest a preference for high or low achievers?
3. Does the teacher structure activities so that student nonparticipation is facilitated?
4. Are the high achieving students becoming bored?
5. Are attractive “back-up” activities available for early finishers?

### **Skill 4. Teacher Use of Time**

Teacher practices model a concern for the effective use of instructional time.

#### Evaluation Questions

1. Are all instructional materials and equipment available and operational at the start of the lesson?
2. Is the teacher physically in the room at the start of the lesson?
3. Have assignments been corrected in a timely manner?
4. Is the teacher giving full attention to the instructional tasks?
5. Is the teacher conducting the lesson at a brisk and interesting pace?

### **Skill 5. Academic Learning Time**

Indicators of academic learning time should provide evidence of progressive improvement in instruction.

#### Evaluation Questions

1. Are high, average, and low achievers on-task and successful?
2. Am I aware of the amount of “on-task” behavior of all individual in my class?
3. Am I aware of the actual levels of mastery of individuals in my class?
4. Has individual “on-task” and mastery information been combined to profile instructional strengths and weaknesses?
5. Is information on academic learning time directing efforts to improve instruction?

### Appendix C: School Data from the BIE's Reading First Proposal

School	Cohort	Grades	2001-2002 Total Enrollment	2002-2003 %FRL	2001- 2002 %LEP	LEP Instrument
Alamo Navajo Community School	1	K-12	387	90%	42%	Language Assessment Survey (LAS)
Aneth Community School	1	K-12	387	90%	42%	Primary Language Form
Black Mesa Community School	1	K-8	55	79%	64%	IDEA Proficiency Test
Chief Leschi School	1	K-12	597	79%	47%	Gates McGinitie, 76% if WA test used
Chilchinbeto Community School	1	PK-8	123	98%	73%	QUI-LEP
Chinle Boarding School	1	K-8	472	98%	96%	IPT: Individual Prof. Test
Cibecue Community School	1	K-12	440	94%	80%	Parents indicate primary language of child
Couer d'Alene Tribal School	1	K-8	63	96%	57%	TOLD
Dunseith Day School	1	K-6	174	100%	56%	BINL, Woodcock Munoz
Greasewood Springs Community School	1	K-8	333	91%	73%	LAS
Hannahville Indian School	1	K-12	136	78%	100%	Terra Nova
Indian Island School	1	PK-8	100	64%	0%	Maine State Parent Survey, Teacher Referral
Jemez Day School	1	K-6	152	100%	99%	IDEA (IPT), SPA Intense
Kaibeto Boarding School	1	K-8	355	100%	46%	LEP based on ISEP prog.
Kayenta Community School	1	K-9	482	89%	76%	Window Rock Oral Language
Lac Courte Oreilles Ojibwa School	1	PK-12	296	80%	82%	Intense Bilingual Certification Form, Teacher Referral, Ojibwemowin Rubric
Lower Brule Day School	1	K-12	215	100%	85%	Grades 4-6 DAC; 7-1 1 -SAT
Lukachukai Community School	1	K-8	374	98%	100%	Test of Language Development
Nay-ah-Shing School	1	K-12	280	77%	64%	Ojibwe Lang.Assessment, LAS, DIAL
Pinon Community School	1	K	63	unknown	100%	IDEA(ipt)NES;LES; FES;WROL
Rough Rock Community School	1	K-12	546	90%	77%	ITP, IDEA for all K-12
Santa Clara Day School	1	K-6	129	70%	40%	Parent Survey
Taos Day School	1	K-8	162	100%	34%	Parent Affidavit
Turtle Mountain Elementary School	1	K-5	622	76%	72%	BINL, Woodcock Munoz, TN
Atsa'Biya'a'zh Community School	2	K-12	320	84%	63%	Parent Survey, LAS, Tchr Evaluation
Beatrice Rafferty School	2	K-8	133	80%	0%	Teacher Referral
Cottonwood Day School	2	K-8	280	100%	65%	IPT I & IPT II
Gila Crossing Community School	2	PK-8	324	100%	100%	Peabody Picture Voc Test & Stanford 9 scores



Indian Township School	2	PK-8	132	66%	91%	Student Record; Teacher Interview, Parent Info, Home Lang. Survey, SPED, Teacher Observation, Student Grades, Informal Assessment
Keams Canyon Elementary School	2	K-6	92	85%	35%	Parent/Teacher Language Survey
Mandaree Day School	2	K-12	226	93%	52%	Woodcock Munoz, ITBS
Menominee Tribal School	2	K-8	262	100%	56%	LAS
Pueblo Pintado Community School	2	K-8	349	95%	92%	LAS
Takini School	2	K-12	182	97%	100%	LAS
Theodore Jamerson Elementary School	2	K-6	116	86%	97%	BINL
Tuba City Boarding School	2	K-8	1026	82%	40%	Parent Survey
Twin Buttes Day School	2	K-8	57	96%	58%	BINL, LAS

## Appendix D: Results of RDD Analyses

**Cohort I:** DIBELS ORF Outcome

**3rd**

**Grade**

Model 1:  $Z\_EORF \sim Z\_BORF + Tx$

	Estimate	SE	t value	P-value
Intercept	0.83	0.08	NA	NA
BORF	0.91	0.04	23.78	<0.001
Tx	-0.33	0.07	-4.45	<0.001

R-sq	0.711	Adj R-sq	0.709
F=554	(2, 451)	<0.001	

Model 2:  $Z\_EORF \sim Z\_BORF + Tx + Tx:Z\_BORF$

	Estimate	SE	t value	P-value
Intercept	-1.87	0.41	NA	NA
BORF	-0.63	0.23	-2.73	0.007
TX	2.34	0.4	5.8	<0.001
Tx:BORF	1.5	0.22	6.72	<0.001

R-Sq	0.737	Adj R-sq	0.735
F=420	(3, 450)	<0.001	

**2nd**

**Grade**

Model 1:  $Z\_EORF \sim Z\_BORF + Tx$

	Estimate	SE	t value	P-value
Intercept	0.11	0.03	NA	NA
BORF	0.74	0.03	23.77	<0.001
Tx	-0.33	0.08	-4.43	<0.001

R-sq	0.711	Adj R-sq	0.709
F=553.4	(2, 451)	<0.001	

Model 2:  $Z\_EORF \sim Z\_BORF + Tx + Tx:Z\_BORF$

	Estimate	SE	t value	P-value
Intercept	-0.88	0.22	NA	NA
BORF	-0.54	0.19	-2.81	0.005
TX	0.998	0.21	4.78	<0.001
Tx:BORF	1.25	0.18	6.79	<0.001

R-Sq	0.737	Adj R-sq	0.736
F=421	(3, 450)	<0.001	

**1st Grade**

Model 1: Z\_EORF ~ Z\_BNWF + Tx

	Estimate	SE	t value	P-value
Intercept	0.08	0.05	NA	NA
BNWF	0.85	0.05	16.63	<0.001
Tx	-0.45	0.16	-2.75	0.006

R-sq	0.442	Adj R-sq	0.44
F=206	(2, 520)	<0.001	

Model 2: Z\_EORF ~ Z\_BORF + Tx + Tx:Z\_BORF

	Estimate	SE	t value	P-value
Intercept	0.08	0.05	NA	NA
BORF	0.85	0.05	16.6	<0.001
TX	-0.84	0.67	-1.25	0.21
Tx: BORF	-0.31	0.51	-0.6	0.55

R-Sq	0.443	Adj R-sq	0.439
F=137	(3, 519)	<0.001	

## Cohort 1: 3rd Grade Vocabulary

	Estimate	SE	t value	P-value
Intercept	0.93		0.09 NA	NA
BORF	0.45		0.08	5.33 <0.001
Treatment	-0.57		0.21	-2.73 0.007
R-sq	0.511	Adj R-sq		0.507
F-value	125.2	238		P<0.001
	Estimate	SE	t value	P-value
Intercept	0.89		0.11 NA	NA
BORF	0.39		0.09	4.3 <0.001
Treatment	0.64		0.67	0.96 0.339
Tx: BORF	0.47		0.25	1.9 0.058
R-sq	0.519	Adj R-sq		0.513
F-value	85.1	3, 237		P<0.001

## 3rd Grade Fluency

	Estimate	SE	t value	P-value
Intercept	0.78		0.09 NA	NA
BORF	0.37		0.09	4.19 <0.001
Treatment	-0.54		0.22	-2.47 0.014
R-sq	0.417	Adj R-sq		0.412
F-value	85.2	238		P<0.001
	Estimate	SE	t value	P-value
Intercept	0.73		0.09 NA	NA
BORF	0.3		0.09	3.19 0.002
Treatment	0.82		0.7	1.17 0.242
Tx: BORF	0.53		0.26	2.05 0.042
R-sq	0.427	Adj R-sq		0.419
F-value	58.8	3, 237		P<0.001

## 3rd Grade Comprehension

	Estimate	SE	t value	P-value
Intercept	0.91		0.08 NA	NA
BORF	0.48		0.08	5.93 <0.001
Treatment	-0.39		0.2	-1.96 0.051
R-sq	0.503	Adj R-sq		0.499
F-value	121.2	238		P<0.001
	Estimate	SE	t value	P-value
Intercept	0.88		0.09 NA	NA
BORF	0.43		0.09	4.95 <.001
Treatment	0.59		0.65	0.92 0.36
Tx: BORF	0.38		0.24	1.61 0.11
R-sq	0.509	Adj R-sq		0.502
F-value	81.8	3, 237		P<0.001

## Cohort 2: 3rd Grade Vocabulary

	Estimate	SE	t value	P-value
Intercept	1.07		0.12 NA	NA
BORF	0.76		0.12	6.46 <0.001
Treatment	0.41		0.29	1.4 0.162
R-sq	0.438	Adj R-sq		0.432
F-value	72.1	2, 185	P<0.001	
	Estimate	SE	t value	P-value
Intercept	0.99		0.15 NA	NA
BORF	0.58		0.22	2.67 0.008
Treatment	0.71		0.41	1.72 0.088
Tx: BORF	0.26		0.26	1.03 0.306
R-sq	0.441	Adj R-sq		0.432
F-value	48.4	3, 184	P<0.001	

## 3rd Grade Fluency

	Estimate	SE	t value	P-value
Intercept	1.17		0.12 NA	NA
BORF	0.8		0.12	6.87 <0.001
Treatment	0.45		0.29	1.56 0.12
R-sq	0.463	Adj R-sq		0.457
F-value	79.8	2, 185	P<0.001	
	Estimate	SE	t value	P-value
Intercept	0.97		0.14 NA	NA
BORF	0.37		0.21	1.77 0.078
Treatment	1.14		0.4	2.84 0.005
Tx: BORF	0.61		0.25	2.43 0.016
R-sq	0.48	Adj R-sq		0.471
F-value	56.6	3, 184	P<0.001	

## 3rd Grade Comprehension

	Estimate	SE	t value	P-value
Intercept	0.99		0.12 NA	NA
BORF	0.71		0.12	6.08 <0.001
Treatment	0.4		0.29	1.37 0.17
R-sq	0.404	Adj R-sq		0.398
F-value	62.7	2, 185	P<0.001	
	Estimate	SE	t value	P-value
Intercept	0.76		0.14 NA	NA
BORF	0.22		0.21	1.06 0.291
Treatment	1.19		0.4	2.95 0.004
Tx: BORF	0.7		0.25	2.78 0.006
R-sq	0.428	Adj R-sq		0.419
F-value	45.9	3, 184	P<0.001	

## Cohort 1: 2nd Grade Vocabulary

	Estimate	SE	t value	P-value
Intercept	0.32		0.07 NA	NA
BORF	0.22		0.05	3.94 <0.001
Treatment	-0.65		0.19	-3.43 <0.001
R-sq	0.345	Adj R-sq		0.34
F-value	63.2	239	P<0.001	
	Estimate	SE	t value	P-value
Intercept	0.32		0.07 NA	NA
BORF	0.21		0.06	3.8 <0.001
Treatment	-0.5		0.61	-0.82 0.415
Tx: BORF	0.08		0.29	0.27 0.79
R-sq	0.345	Adj R-sq		0.337
F-value	41.9	3, 238	P<0.001	

## 2nd Grade Fluency

	Estimate	SE	t value	P-value
Intercept	0.25		0.07 NA	NA
BORF	0.25		0.05	4.84 <0.001
Treatment	-0.64		0.18	-3.53 <0.001
R-sq	0.405	Adj R-sq		0.4
F-value	81.5	2, 239	P<0.001	
	Estimate	SE	t value	P-value
Intercept	0.27		0.07 NA	NA
BORF	0.23		0.05	4.26 <0.001
Treatment	0.93		0.58	1.62 0.107
Tx: BORF	0.79		0.27	2.88 0.004
R-sq	0.425	Adj R-sq		0.418
F-value	58.7	3, 238	P<0.001	

## 2nd Grade Comprehension

	Estimate	SE	t value	P-value
Intercept	0.25		0.07 NA	NA
BORF	0.23		0.06	4 <0.001
Treatment	-0.52		0.2	-2.66 0.008
R-sq	0.302	Adj R-sq		0.296
F-value	51.7	2, 239	P<0.001	
	Estimate	SE	t value	P-value
Intercept	0.26		0.07 NA	NA
BORF	0.2		0.06	3.52 <.001
Treatment	0.79		0.62	1.27 0.205
Tx: BORF	0.65		0.29	2.22 0.027
R-sq	0.316	Adj R-sq		0.308
F-value	36.7	3, 238	P<0.001	

## Cohort 2: 2nd Grade Vocabulary

	Estimate	SE	t value	P-value
Intercept	0.67		0.1 NA	NA
BORF	0.42		0.11	3.81 <0.001
Treatment	-0.31		0.27	-1.18 0.241
R-sq	0.36	Adj R-sq		0.353
F-value	52.3	2, 186		P<0.001
	Estimate	SE	t value	P-value
Intercept	0.69		0.1 NA	NA
BORF	0.27		0.12	2.21 0.028
Treatment	0.64		0.47	1.37 0.173
Tx: BORF	0.64		0.26	2.46 0.015
R-sq	0.381	Adj R-sq		0.37
F-value	37.9	3, 185		P<0.001

## 2nd Grade Fluency

	Estimate	SE	t value	P-value
Intercept	0.72		0.11 NA	NA
BORF	0.34		0.12	2.83 0.005
Treatment	-0.5		0.29	-1.77 0.079
R-sq	0.32	Adj R-sq		0.313
F-value	43.7	2, 186		P<0.001
	Estimate	SE	t value	P-value
Intercept	0.73		0.11 NA	NA
BORF	0.224		0.13	1.67 0.097
Treatment	0.22		0.51	0.43 0.665
Tx: BORF	0.49		0.28	1.73 0.085
R-sq	0.331	Adj R-sq		0.32
F-value	30.5	3, 185		P<0.001

## 2nd Grade Comprehension

	Estimate	SE	t value	P-value
Intercept	0.63		0.11 NA	NA
BORF	0.4		0.12	3.38 <0.001
Treatment	-0.24		0.28	-0.86 0.39
R-sq	0.29	Adj R-sq		0.282
F-value	38	2, 186		P<0.001
	Estimate	SE	t value	P-value
Intercept	0.64		0.11 NA	NA
BORF	0.3		0.13	2.26 0.025
Treatment	0.38		0.5	0.76 0.449
Tx: BORF	0.42		0.28	1.5 0.135
R-sq	0.298	Adj R-sq		0.287
F-value	26.2	3, 185		P<0.001

## Cohort 1: 1st Vocabulary

	Estimate	SE	t value	P-value
Intercept	0.25		0.06 NA	NA
BNWF	0.22		0.06	3.5 <0.001
Treatment	-0.45		0.19	-2.37 0.018
R-sq	0.127	Adj R-sq		0.121
F-value	21.2	2, 290	P<0.001	
	Estimate	SE	t value	P-value
Intercept	0.25		0.06 NA	NA
BNWF	0.21		0.06	3.32 0.001
Treatment	1.94		1.25	1.55 0.122
Tx: BNWF	1.25		0.65	1.94 0.054
R-sq	0.138	Adj R-sq		0.129
F-value	15.4	3, 289	P<0.001	

## 1st Grade Fluency

	Estimate	SE	t value	P-value
Intercept	0.31		0.05 NA	NA
BNWF	0.23		0.06	3.79 <0.001
Treatment	-0.91		0.18	-4.96 <0.001
R-sq	0.243	Adj R-sq		0.237
F-value	46.5	2, 290	P<0.001	
	Estimate	SE	t value	P-value
Intercept	0.31		0.05 NA	NA
BNWF	0.22		0.06	3.65 <0.001
Treatment	0.73		1.21	0.6 0.548
Tx: BNWF	0.86		0.62	1.38 0.17
R-sq	0.248	Adj R-sq		0.24
F-value	31.7	3, 289	P<0.001	

## 1st Grade Comprehension

	Estimate	SE	t value	P-value
Intercept	0.34		0.05 NA	NA
BNWF	0.22		0.06	3.82 <0.001
Treatment	-0.77		0.18	-4.3 <0.001
R-sq	0.215	Adj R-sq		0.21
F-value	39.7	2, 290	P<0.001	
	Estimate	SE	t value	P-value
Intercept	0.33		0.06 NA	NA
BNWF	0.21		0.06	3.65 <.001
Treatment	1.2		1.17	1.02 0.307
Tx: BNWF	1.03		0.6	1.7 0.09
R-sq	0.223	Adj R-sq		0.215
F-value	27.6	3, 289	P<0.001	



## Cohort 2: 1st grade Vocabulary

	Estimate	SE	t value	P-value
Intercept	0.46		0.09 NA	NA
BNWF	0.18		0.11 1.74	0.083
Treatment	-0.5		0.25 -2.02	0.045
R-sq	0.178	Adj R-sq	0.168	
F-value	18.9 2,	175	P<0.001	
	Estimate	SE	t value	P-value
Intercept	0.46		0.09 NA	NA
BNWF	0.19		0.11 1.74	0.084
Treatment	-0.66		0.64 -1.03	0.302
Tx: BNWF	-0.1		0.35 -0.28	0.782
R-sq	0.178	Adj R-sq	0.164	
F-value	12.6 3,	174	P<0.001	

## 1st Grade Fluency

	Estimate	SE	t value	P-value
Intercept	0.41		0.09 NA	NA
BNWF	0.29		0.1 2.88	0.005
Treatment	-0.3		0.24 -1.28	0.202
R-sq	0.211	Adj R-sq	0.202	
F-value	23.5 2,	175	P<0.001	
	Estimate	SE	t value	P-value
Intercept	0.41		0.09 NA	NA
BNWF	0.27		0.11 2.54	0.012
Treatment	0.04		0.61 0.06	0.952
Tx: BNWF	0.2		0.34 0.6	0.548
R-sq	0.213	Adj R-sq	0.2	
F-value	15.7 3,	174	P<0.001	

## 1st Grade Comprehension

	Estimate	SE	t value	P-value
Intercept	0.47		0.08 NA	NA
BNWF	0.33		0.1 3.45	<0.001
Treatment	-0.51		0.22 -2.28	0.024
R-sq	0.334	Adj R-sq	0.327	
F-value	44 2,	175	P<0.001	
	Estimate	SE	t value	P-value
Intercept	0.47		0.08 NA	NA
BNWF	0.28		0.1 2.81	0.006
Treatment	0.29		0.58 0.5	0.615
Tx: BNWF	0.48		0.32 1.5	0.134
R-sq	0.343	Adj R-sq	0.332	
F-value	30.3 3,	174	P<0.001	

## **Appendix E: Predictive Ability of DIBELS When Compared to SAT-10**

# BIE Reading First DIBELS and SAT-10 Data: 2004-2006 School Years (24 Cohort I Schools)

## Numbers of Students Assessed

This report summarizes results of DIBELS and SAT-10 data for the Bureau of Indian Education's (BIE) 24 Cohort I Reading First schools. The first full year of Reading First implementation occurred during the 2004-2005 school year. The data analyzed for this report includes data from both the 2004-2005 and 2005-2006 school years.

The number of students in each analysis reported in this document varies based on completeness of data (e.g., some students did not have beginning-of-year or BOY scores but had both end-of-year or EOY scores and SAT-10 scores, or some students attended BIE Reading First schools during one year but not the other, or not all schools administered the DIBELS benchmark to students who were absent during EndVision's 2004-2005 site visits, etc.). The numbers of students for whom scores exist at each time point are included in the following table.

### Number of Students with Scores by Test, Time Point, and Grade

	<b>DIBELS</b>				<b>SAT-10</b>	
	<b>2004-2005 BOY</b>	<b>2004-2005 EOY</b>	<b>2005-2006 BOY</b>	<b>2005-2006 EOY</b>	<b>2004-2005</b>	<b>2005-2006</b>
<b>Kindergarten</b>	684	692	691	694	333	334
<b>1<sup>st</sup> Grade</b>	559	566	625	619	508	596
<b>2<sup>nd</sup> Grade</b>	615	630	582	587	604	560
<b>3<sup>rd</sup> Grade</b>	607	599	576	548	541	541
<b>TOTAL</b>	2465	2487	2474	2448	1986	2031

# **BIE Reading First DIBELS and SAT-10 Data: 2004-2006 School Years (24 Cohort I Schools)**

## **Relationships between 2005-2006 BOY and EOY DIBELS Instructional Recommendations**

The relationships between BOY and EOY DIBELS Instructional Recommendations provide an indication of the amount of progress students made. For example, we would hope that most Benchmark students remained at Benchmark, and that fewer students who were Intensive at the beginning of the year remained Intensive at the end of the year. Additionally, we want students who were Strategic and Intensive at the beginning of the year to “move up” in Instructional Recommendation.

In the following tables, the **RED** highlights indicate the percentage of students who were Benchmark at the beginning of the year who remained Benchmark at the end of the year, and the percentage of students who were Intensive who “moved up.” The **BLUE** highlights indicate the percentage of students who were Strategic at the beginning of the year who “moved up.”

During the 2004-2005 school year, the percentages of students with paired scores who started at Benchmark and stayed at Benchmark ranged from 64% to 80% across grades, with 74% of students staying at Benchmark overall. However, the percentages of students with paired scores who “moved up” from Intensive or Strategic were smaller than during the second full year of BIE Reading First implementation.

During the 2005-2006 school year, the percentages of students with paired scores who started at Benchmark and stayed at Benchmark were 90% or higher in all but 1<sup>st</sup> grade. In 1<sup>st</sup> grade, only 76% of students who started at Benchmark remained at Benchmark. Overall, 85% of students who started at Benchmark read on grade level as indicated by the DIBELS at the end of the year.

The percentages of students with paired scores who began the year Intensive but who “moved up” varied between grades, with kindergarten showing 78% of students “moving up” while in 2<sup>nd</sup> grade, only 23% of students “moved up.” Surprisingly, 43% of 3<sup>rd</sup> grade students who were Intensive at the beginning of the year “moved up” by the end of the year!

# BIE Reading First DIBELS and SAT-10 Data: 2004-2006 School Years (24 Cohort I Schools)

## Relationships between 2004-2005 BOY and EOY DIBELS Instructional Recommendations

<b>Kindergarten 2004-2005</b>		<b>Kindergarten EOY DIBELS Instructional Recommendation (number of paired scores)</b>				<b>Kindergarten EOY DIBELS Instructional Recommendation (percent of paired scores)</b>		
		<b>Benchmark</b>	<b>Strategic</b>	<b>Intensive</b>	<b>TOTAL</b>	<b>Benchmark</b>	<b>Strategic</b>	<b>Intensive</b>
Kindergarten BOY DIBELS Instructional Recommendation	<b>Benchmark</b>	85	15	6	106	<b>80%</b>	14%	6%
	<b>Strategic</b>	149	64	60	273	<b>55%</b>	23%	22%
	<b>Intensive</b>	75	56	133	264	<b>28%</b>	<b>21%</b>	50%
	<b>TOTAL</b>	309	135	199	643			

<b>1<sup>st</sup> Grade 2004-2005</b>		<b>1<sup>st</sup> Grade EOY DIBELS Instructional Recommendation (number of paired scores)</b>				<b>1<sup>st</sup> Grade EOY DIBELS Instructional Recommendation (percent of paired scores)</b>		
		<b>Benchmark</b>	<b>Strategic</b>	<b>Intensive</b>	<b>TOTAL</b>	<b>Benchmark</b>	<b>Strategic</b>	<b>Intensive</b>
1 <sup>st</sup> Grade BOY DIBELS Instructional Recommendation	<b>Benchmark</b>	127	62	9	198	<b>64%</b>	31%	5%
	<b>Strategic</b>	62	68	32	162	<b>38%</b>	42%	20%
	<b>Intensive</b>	14	51	104	169	<b>8%</b>	<b>30%</b>	62%
	<b>TOTAL</b>	203	181	145	529			

<b>2<sup>nd</sup> Grade 2004-2005</b>		<b>2<sup>nd</sup> Grade EOY DIBELS Instructional Recommendation (number of paired scores)</b>				<b>2<sup>nd</sup> Grade EOY DIBELS Instructional Recommendation (percent of paired scores)</b>		
		<b>Benchmark</b>	<b>Strategic</b>	<b>Intensive</b>	<b>TOTAL</b>	<b>Benchmark</b>	<b>Strategic</b>	<b>Intensive</b>
2 <sup>nd</sup> Grade BOY DIBELS Instructional Recommendation	<b>Benchmark</b>	163	46	7	216	<b>75%</b>	21%	3%
	<b>Strategic</b>	30	64	63	157	<b>19%</b>	41%	40%
	<b>Intensive</b>	8	17	191	216	<b>4%</b>	<b>8%</b>	88%
	<b>TOTAL</b>	201	127	261	589			

<b>3<sup>rd</sup> Grade 2004-2005</b>		<b>3<sup>rd</sup> Grade EOY DIBELS Instructional Recommendation (number of paired scores)</b>				<b>3<sup>rd</sup> Grade EOY DIBELS Instructional Recommendation (percent of paired scores)</b>		
		<b>Benchmark</b>	<b>Strategic</b>	<b>Intensive</b>	<b>TOTAL</b>	<b>Benchmark</b>	<b>Strategic</b>	<b>Intensive</b>
3 <sup>rd</sup> Grade BOY DIBELS Instructional Recommendation	<b>Benchmark</b>	114	30	1	145	<b>79%</b>	21%	1%
	<b>Strategic</b>	36	106	24	166	<b>22%</b>	64%	14%
	<b>Intensive</b>	5	45	212	262	<b>2%</b>	<b>17%</b>	81%
	<b>TOTAL</b>	155	181	237	573			

<b>Grades K-3 2004-2005</b>		<b>Total Grades K-3 EOY DIBELS Instructional Recommendation (number of paired scores)</b>				<b>Total Grades K-3 EOY DIBELS Instructional Recommendation (percent of paired scores)</b>		
		<b>Benchmark</b>	<b>Strategic</b>	<b>Intensive</b>	<b>TOTAL</b>	<b>Benchmark</b>	<b>Strategic</b>	<b>Intensive</b>
Total Grades K-3 BOY DIBELS Instructional Recommendation	<b>Benchmark</b>	489	153	23	665	<b>74%</b>	23%	3%
	<b>Strategic</b>	277	302	179	758	<b>37%</b>	40%	24%
	<b>Intensive</b>	102	169	640	911	<b>11%</b>	<b>19%</b>	70%
	<b>TOTAL</b>	868	624	842	2334			

# BIE Reading First DIBELS and SAT-10 Data: 2004-2006 School Years (24 Cohort I Schools)

## Relationships between 2005-2006 BOY and EOY DIBELS Instructional Recommendations

<b>Kindergarten 2005-2006</b>		<b>Kindergarten EOY DIBELS Instructional Recommendation (number of paired scores)</b>				<b>Kindergarten EOY DIBELS Instructional Recommendation (percent of paired scores)</b>		
		<b>Benchmark</b>	<b>Strategic</b>	<b>Intensive</b>	<b>TOTAL</b>	<b>Benchmark</b>	<b>Strategic</b>	<b>Intensive</b>
Kindergarten BOY DIBELS Instructional Recommendation	<b>Benchmark</b>	66	3	1	70	<b>94%</b>	4%	1%
	<b>Strategic</b>	200	25	25	250	<b>80%</b>	10%	10%
	<b>Intensive</b>	172	68	66	306	<b>56%</b>	<b>22%</b>	22%
	<b>TOTAL</b>	438	96	92	626			

<b>1<sup>st</sup> Grade 2005-2006</b>		<b>1<sup>st</sup> Grade EOY DIBELS Instructional Recommendation (number of paired scores)</b>				<b>1<sup>st</sup> Grade EOY DIBELS Instructional Recommendation (percent of paired scores)</b>		
		<b>Benchmark</b>	<b>Strategic</b>	<b>Intensive</b>	<b>TOTAL</b>	<b>Benchmark</b>	<b>Strategic</b>	<b>Intensive</b>
1 <sup>st</sup> Grade BOY DIBELS Instructional Recommendation	<b>Benchmark</b>	224	60	9	293	<b>76%</b>	20%	3%
	<b>Strategic</b>	53	73	23	149	<b>36%</b>	49%	15%
	<b>Intensive</b>	17	45	61	123	<b>14%</b>	<b>37%</b>	50%
	<b>TOTAL</b>	294	178	93	565			

<b>2<sup>nd</sup> Grade 2005-2006</b>		<b>2<sup>nd</sup> Grade EOY DIBELS Instructional Recommendation (number of paired scores)</b>				<b>2<sup>nd</sup> Grade EOY DIBELS Instructional Recommendation (percent of paired scores)</b>		
		<b>Benchmark</b>	<b>Strategic</b>	<b>Intensive</b>	<b>TOTAL</b>	<b>Benchmark</b>	<b>Strategic</b>	<b>Intensive</b>
2 <sup>nd</sup> Grade BOY DIBELS Instructional Recommendation	<b>Benchmark</b>	172	14	1	187	<b>92%</b>	7%	1%
	<b>Strategic</b>	60	68	26	154	<b>39%</b>	44%	17%
	<b>Intensive</b>	12	32	149	193	<b>6%</b>	<b>17%</b>	77%
	<b>TOTAL</b>	244	114	176	534			

<b>3<sup>rd</sup> Grade 2005-2006</b>		<b>3<sup>rd</sup> Grade EOY DIBELS Instructional Recommendation (number of paired scores)</b>				<b>3<sup>rd</sup> Grade EOY DIBELS Instructional Recommendation (percent of paired scores)</b>		
		<b>Benchmark</b>	<b>Strategic</b>	<b>Intensive</b>	<b>TOTAL</b>	<b>Benchmark</b>	<b>Strategic</b>	<b>Intensive</b>
3 <sup>rd</sup> Grade BOY DIBELS Instructional Recommendation	<b>Benchmark</b>	139	15	0	154	<b>90%</b>	10%	0%
	<b>Strategic</b>	67	87	6	160	<b>42%</b>	54%	4%
	<b>Intensive</b>	14	76	120	210	<b>7%</b>	<b>36%</b>	57%
	<b>TOTAL</b>	220	178	126	524			

<b>Grades K-3 2005-2006</b>		<b>Total Grades K-3 EOY DIBELS Instructional Recommendation (number of paired scores)</b>				<b>Total Grades K-3 EOY DIBELS Instructional Recommendation (percent of paired scores)</b>		
		<b>Benchmark</b>	<b>Strategic</b>	<b>Intensive</b>	<b>TOTAL</b>	<b>Benchmark</b>	<b>Strategic</b>	<b>Intensive</b>
Total Grades K-3 BOY DIBELS Instructional Recommendation	<b>Benchmark</b>	601	92	11	704	<b>85%</b>	13%	2%
	<b>Strategic</b>	380	253	80	713	<b>53%</b>	35%	11%
	<b>Intensive</b>	215	221	396	832	<b>26%</b>	<b>27%</b>	48%
	<b>TOTAL</b>	1196	566	487	2249			

# BIE Reading First DIBELS and SAT-10 Data: 2004-2006 School Years (24 Cohort I Schools)

## Relationships between 2004-2005 and 2005-2006 SAT-10 Levels of Proficiency

The relationships between 2004-2005 and 2005-2006 SAT-10 levels of proficiency also provide an indication of the amount of progress students made. For example, we would hope that most students At Grade Level (AGL) remained AGL the following year, and that fewer students who Need Substantial Intervention (NSI) the first year stayed at that level the following year. Additionally, we want students who Need Additional Intervention (NAI) to “move up” to AGL the second year. Additionally, the tables show the relationships between those who were proficient (i.e., at or above the 40<sup>th</sup> percentile) or not proficient.

In the following tables, the **RED** highlights indicate the percentage of students who were AGL at the end of the first full year of BIE Reading First implementation who remained AGL at the end of the second year, and the percentage of students who were NSI who “moved up.” The **BLUE** highlights indicate the percentage of students who were NAI at the end of the first year and who “moved up” by the end of the second year. The **MAGENTA** indicates students who were not proficient at the end of the first year based on SAT-10 scores and who remained not proficient at the end of the second year.

A few points need to be mentioned.

- First, **the SAT-10 was not required to be administered in kindergarten, so not all schools administered it.** Hence, the number of kindergarten students for whom SAT-10 scores are available is approximately half of what might be expected, or approximately 230 of more than 530 students.
- Additionally, one school submitted scores without student names the first year, so those scores cannot be matched with student scores from the second year. This school was one of the highest performing BIE schools (implying that the numbers reported here are slightly deflated) although the number of Reading First students at that school was small (approximately 40 students).
- Also, these numbers compare only matched samples of students who attended BIE Reading First Cohort I schools during the first two full years of implementation. Given the high mobility rate of students at some schools, **the numbers should be interpreted with caution.**

## BIE Reading First DIBELS and SAT-10 Data: 2004-2006 School Years (24 Cohort I Schools)

### Kindergarten to 1<sup>st</sup> Grade: Relationships between 2004-2005 & 2005-2006 SAT-10

K to 1 <sup>st</sup> Phonemic Awareness		2005-2006 1 <sup>st</sup> Grade SAT-10 Phonemic Awareness (number of paired scores)				(percent of paired scores)		
		AGL	NAI	NSI	TOTAL	AGL	NAI	NSI
		Kindergarten 2004-2005 SAT-10 Phonemic Awareness	AGL	128	17	7	152	84%
	NAI	8	3	2	13	62%	23%	15%
	NSI	19	15	6	40	48%	38%	15%
	TOTAL	155	35	15	205			

K to 1 <sup>st</sup> Phonics		2005-2006 1 <sup>st</sup> Grade SAT-10 Phonics (number of paired scores)				(percent of paired scores)		
		AGL	NAI	NSI	TOTAL	AGL	NAI	NSI
		Kindergarten 2004-2005 SAT-10 Phonics	AGL	52	58	28	138	38%
	NAI	10	19	11	40	25%	48%	28%
	NSI	5	13	9	27	19%	48%	33%
	TOTAL	67	90	48	205			

K to 1 <sup>st</sup> Vocabulary		2005-2006 1 <sup>st</sup> Grade SAT-10 Vocabulary (number of paired scores)				(percent of paired scores)		
		AGL	NAI	NSI	TOTAL	AGL	NAI	NSI
		Kindergarten 2004-2005 SAT-10 Vocabulary	AGL	36	9	11	56	64%
	NAI	39	13	47	99	39%	13%	47%
	NSI	14	14	22	50	28%	28%	44%
	TOTAL	89	36	80	205			

K to 1 <sup>st</sup> Fluency		2005-2006 1 <sup>st</sup> Grade SAT-10 Fluency (number of paired scores)				(percent of paired scores)		
		AGL	NAI	NSI	TOTAL	AGL	NAI	NSI
		Kindergarten 2004-2005 SAT-10 Fluency	AGL	66	30	15	111	59%
	NAI	11	4	10	25	44%	16%	40%
	NSI	19	17	33	69	28%	25%	48%
	TOTAL	96	51	58	205			

K to 1 <sup>st</sup> Comprehension		2005-2006 1 <sup>st</sup> Grade SAT-10 Comprehension (number of paired scores)				(percent of paired scores)		
		AGL	NAI	NSI	TOTAL	AGL	NAI	NSI
		Kindergarten 2004-2005 SAT-10 Comprehension	AGL	44	12	5	61	72%
	NAI	58	19	11	88	66%	22%	13%
	NSI	17	24	14	55	31%	44%	25%
	TOTAL	119	55	30	204			

K to 1 <sup>st</sup> SAT-10 Proficiency		2005-2006 1 <sup>st</sup> Grade SAT-10 Proficiency (number of paired scores)			(percentages)	
		Proficient	Not Proficient	TOTAL	Proficient	Not Proficient
		Kindergarten 2004- 2005 SAT-10 Proficiency	Proficient	74	26	100
	Not Proficient	22	78	100	22%	78%
	TOTAL	96	104	200		



## BIE Reading First DIBELS and SAT-10 Data: 2004-2006 School Years (24 Cohort I Schools)

### 1<sup>st</sup> to 2<sup>nd</sup> Grade: Relationships between 2004-2005 and 2005-2006 SAT-10 Levels

1 <sup>st</sup> to 2 <sup>nd</sup> Phonemic Awareness		2005-2006 2 <sup>nd</sup> Grade SAT-10 Phonemic Awareness (number of paired scores)				(percent of paired scores)		
		AGL	NAI	NSI	TOTAL	AGL	NAI	NSI
		1 <sup>st</sup> Grade 2004-2005 SAT-10 Phonemic Awareness	AGL	260	1	0	261	100%
	NAI	68	3	2	73	93%	4%	3%
	NSI	28	5	3	36	78%	14%	8%
	TOTAL	356	9	5	370			

1 <sup>st</sup> to 2 <sup>nd</sup> Phonics		2005-2006 2 <sup>nd</sup> Grade SAT-10 Phonics (number of paired scores)				(percent of paired scores)		
		AGL	NAI	NSI	TOTAL	AGL	NAI	NSI
		1 <sup>st</sup> Grade 2004-2005 SAT-10 Phonics	AGL	74	10	21	105	70%
	NAI	60	20	67	147	41%	14%	46%
	NSI	30	11	77	118	25%	9%	65%
	TOTAL	164	41	165	370			

1 <sup>st</sup> to 2 <sup>nd</sup> Vocabulary		2005-2006 2 <sup>nd</sup> Grade SAT-10 Vocabulary (number of paired scores)				(percent of paired scores)		
		AGL	NAI	NSI	TOTAL	AGL	NAI	NSI
		1 <sup>st</sup> Grade 2004-2005 SAT-10 Vocabulary	AGL	94	24	18	136	69%
	NAI	31	25	40	96	32%	26%	42%
	NSI	15	18	105	138	11%	13%	76%
	TOTAL	140	67	163	370			

1 <sup>st</sup> to 2 <sup>nd</sup> Fluency		2005-2006 2 <sup>nd</sup> Grade SAT-10 Fluency (number of paired scores)				(percent of paired scores)		
		AGL	NAI	NSI	TOTAL	AGL	NAI	NSI
		1 <sup>st</sup> Grade 2004-2005 SAT-10 Fluency	AGL	90	35	36	161	56%
	NAI	26	26	48	100	26%	26%	48%
	NSI	26	30	53	109	24%	28%	49%
	TOTAL	142	91	137	370			

1 <sup>st</sup> to 2 <sup>nd</sup> Comprehension		2005-2006 2 <sup>nd</sup> Grade SAT-10 Comprehension (number of paired scores)				(percent of paired scores)		
		AGL	NAI	NSI	TOTAL	AGL	NAI	NSI
		1 <sup>st</sup> Grade 2004-2005 SAT-10 Comprehension	AGL	89	63	71	223	40%
	NAI	10	16	42	68	15%	24%	62%
	NSI	5	16	58	79	6%	20%	73%
	TOTAL	104	95	171	370			

1 <sup>st</sup> to 2 <sup>nd</sup> SAT-10 Proficiency		2005-2006 2 <sup>nd</sup> Grade SAT-10 Proficiency (number of paired scores)			(percentages)	
		Proficient	Not Proficient	TOTAL	Proficient	Not Proficient
		1 <sup>st</sup> Grade 2004-2005 SAT-10 Proficiency	Proficient	129	33	162
	Not Proficient	40	160	200	20%	80%
	TOTAL	169	193	362		

## BIE Reading First DIBELS and SAT-10 Data: 2004-2006 School Years (24 Cohort I Schools)

### 2<sup>nd</sup> to 3<sup>rd</sup> Grade: Relationships between 2004-2005 and 2005-2006 SAT-10 Levels

2 <sup>nd</sup> to 3 <sup>rd</sup> Phonemic Awareness		2005-2006 3 <sup>rd</sup> Grade SAT-10 Phonemic Awareness (number of paired scores)				(percent of paired scores)		
		AGL	NAI	NSI	TOTAL	AGL	NAI	NSI
		2 <sup>nd</sup> Grade 2004-2005 SAT-10 Phonemic Awareness	AGL	216	70	127	413	52%
	NAI	4	7	11	22	18%	32%	50%
	NSI	1	3	5	9	11%	33%	56%
	TOTAL	221	80	143	444			

2 <sup>nd</sup> to 3 <sup>rd</sup> Phonics		2005-2006 3 <sup>rd</sup> Grade SAT-10 Phonics (number of paired scores)				(percent of paired scores)		
		AGL	NAI	NSI	TOTAL	AGL	NAI	NSI
		2 <sup>nd</sup> Grade 2004-2005 SAT-10 Phonics	AGL	128	12	22	162	79%
	NAI	30	10	11	51	59%	20%	22%
	NSI	90	37	103	230	39%	16%	45%
	TOTAL	248	59	136	443			

2 <sup>nd</sup> to 3 <sup>rd</sup> Vocabulary		2005-2006 3 <sup>rd</sup> Grade SAT-10 Vocabulary (number of paired scores)				(percent of paired scores)		
		AGL	NAI	NSI	TOTAL	AGL	NAI	NSI
		2 <sup>nd</sup> Grade 2004-2005 SAT-10 Vocabulary	AGL	79	52	27	158	50%
	NAI	22	26	26	74	30%	35%	35%
	NSI	19	66	126	211	9%	31%	60%
	TOTAL	120	144	179	443			

2 <sup>nd</sup> to 3 <sup>rd</sup> Fluency		2005-2006 3 <sup>rd</sup> Grade SAT-10 Fluency (number of paired scores)				(percent of paired scores)		
		AGL	NAI	NSI	TOTAL	AGL	NAI	NSI
		2 <sup>nd</sup> Grade 2004-2005 SAT-10 Fluency	AGL	119	26	10	155	77%
	NAI	53	30	33	116	46%	26%	28%
	NSI	41	35	96	172	24%	20%	56%
	TOTAL	213	91	139	443			

2 <sup>nd</sup> to 3 <sup>rd</sup> Comprehension		2005-2006 3 <sup>rd</sup> Grade SAT-10 Comprehension (number of paired scores)				(percent of paired scores)		
		AGL	NAI	NSI	TOTAL	AGL	NAI	NSI
		2 <sup>nd</sup> Grade 2004-2005 SAT-10 Comprehension	AGL	49	56	20	125	39%
	NAI	29	32	24	85	34%	38%	28%
	NSI	13	73	147	233	6%	31%	63%
	TOTAL	91	161	191	443			

2 <sup>nd</sup> to 3 <sup>rd</sup> SAT-10 Proficiency		2005-2006 3 <sup>rd</sup> Grade SAT-10 Proficiency (number of paired scores)			(percentages)	
		Proficient	Not Proficient	TOTAL	Proficient	Not Proficient
		2 <sup>nd</sup> Grade 2004-2005 SAT-10 Proficiency	Proficient	118	55	173
	Not Proficient	44	221	265	17%	83%
	TOTAL	162	276	438		

## BIE Reading First DIBELS and SAT-10 Data: 2004-2006 School Years (24 Cohort I Schools)

### Correlations between EOY DIBELS ORF Scores and SAT-10 Scores

We typically use correlations between scores on two or more tests as one method to assess the concurrent (or criterion-related) validity of a measure. In our case, we want correlations between EOY DIBELS ORF scores and SAT-10 scores to be as high as possible (and statistically significant). With such a large sample of students, the correlations presented here are statistically significant just because of the large sample size.

However, how high does the correlation need to be in order to be meaningful? While various “rules of thumb” can be located in the literature, another way to determine “meaningful” is to consider the magnitude of shared variance between the tests, or the square of the correlation. For example, a correlation of  $r = .5$  indicates that the tests have 25% of their variability in scores in common. With a correlation of  $r = .7$ , the shared variance is nearly 50%.

In tests of concurrent validity, the literature rarely reports Pearson correlations higher than  $r = .60$ . Correlations of  $r = .50$  are typically reported as moderate, so for the purposes of this report, correlations of .50 or higher were highlighted in **RED**. Because we are comparing EOY DIBELS ORF scores with SAT-10 scores, we hope to see substantial amounts of **RED** in the first column of correlations—or the column that contains EOY ORF correlations (otherwise in **BLUE**) with SAT-10 scores.

DIBELS ORF scores differ in magnitude between 1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup> grades (e.g., we would expect a student in 3<sup>rd</sup> grade to read more words than a student in 1<sup>st</sup> grade, and hence, have a higher ORF score), but the SAT-10 subtests are scaled to have similar scores in each grade (e.g., a high performing 1<sup>st</sup> grade student would have a score similar to a high performing 3<sup>rd</sup> grade student). Because of the differences in scales across years on the DIBELS, correlations for the three grades are presented in separate tables.

Note that in all three grades, the correlations between EOY DIBELS ORF scores and SAT-10 subtests for fluency and comprehension are at least moderate. Additionally, in both 1<sup>st</sup> and 2<sup>nd</sup> grades, the correlations between EOY DIBELS ORF scores and SAT-10 vocabulary subtest scores are at least moderate.

Does the lower correlation between EOY DIBELS ORF scores and SAT-10 vocabulary subtest scores for 3<sup>rd</sup> grade students provide additional evidence for the problem that the BIE Reading First schools and leadership team have already identified: that 3<sup>rd</sup> grade students struggle with comprehension due to insufficient vocabulary? Can we make the claim that fluency is less affected in 3<sup>rd</sup> grade due to the strong focus on phonics of many of the intervention or replacement core reading programs used in 2<sup>nd</sup> and 3<sup>rd</sup> grades in many of the BIE’s Reading First schools?

# BIE Reading First DIBELS and SAT-10 Data: 2004-2006 School Years (24 Cohort I Schools)

## 1<sup>st</sup> Grade: Correlations between 2005-2006 EOY DIBELS ORF Scores and SAT-10 Scores

<b>1<sup>st</sup> Grade 2005-2006</b>		<b>1<sup>st</sup> EOY ORF</b>	SAT 2006 Phonemic Awareness Number Correct	SAT 2006 Phonics Number Correct	SAT 2006 Vocabulary Number Correct	SAT 2006 Fluency Number Correct	SAT 2006 Comprehension Number Correct
SAT 2006 Phonemic Number Correct	Pearson Correlation	<b>.538(**)</b>					
	Significance (2-tailed)	.000					
	Number of students	595					
SAT 2006 Phonics Number Correct	Pearson Correlation	.431(**)	.439(**)				
	Significance (2-tailed)	.000	.000				
	Number of students	595	597				
SAT 2006 Vocabulary Number Correct	Pearson Correlation	<b>.587(**)</b>	<b>.521(**)</b>	.465(**)			
	Significance (2-tailed)	.000	.000	.000			
	Number of students	596	597	597			
SAT 2006 Fluency Number Correct	Pearson Correlation	<b>.599(**)</b>	<b>.519(**)</b>	.429(**)	<b>.613(**)</b>		
	Significance (2-tailed)	.000	.000	.000	.000		
	Number of students	596	597	597	598		
SAT 2006 Comprehension Number Correct	Pearson Correlation	<b>.624(**)</b>	<b>.559(**)</b>	.456(**)	<b>.636(**)</b>	<b>.673(**)</b>	
	Significance (2-tailed)	.000	.000	.000	.000	.000	
	Number of students	596	597	597	598	598	
SAT 2006 Total Score	Pearson Correlation	.419(**)	.494(**)	.353(**)	.400(**)	.417(**)	.483(**)
	Significance (2-tailed)	.000	.000	.000	.000	.000	.000
	Number of students	596	596	596	596	596	596

\*\*Correlation is significant at the 0.01 level (2-tailed).

Blue indicates correlations between EOY DIBELS ORF Scores and SAT-10 Scores

Red indicates correlations that exceed .500

## BIE Reading First DIBELS and SAT-10 Data: 2004-2006 School Years (24 Cohort I Schools)

### 2<sup>nd</sup> Grade: Correlations between 2005-2006 EOY DIBELS ORF Scores and SAT-10 Scores

<b>2<sup>nd</sup> Grade 2005-2006</b>		<b>2<sup>nd</sup> EOY ORF</b>	SAT 2006 Phonemic Awareness Number Correct	SAT 2006 Phonics Number Correct	SAT 2006 Vocabulary Number Correct	SAT 2006 Fluency Number Correct	SAT 2006 Comprehension Number Correct
SAT 2006 Phonemic Number Correct	Pearson Correlation	.379(**)					
	Significance (2-tailed)	.000					
	Number of students	559					
SAT 2006 Phonics Number Correct	Pearson Correlation	.464(**)	.293(**)				
	Significance (2-tailed)	.000	.000				
	Number of students	559	561				
SAT 2006 Vocabulary Number Correct	Pearson Correlation	<b>.626(**)</b>	.349(**)	<b>.636(**)</b>			
	Significance (2-tailed)	.000	.000	.000			
	Number of students	558	560	560			
SAT 2006 Fluency Number Correct	Pearson Correlation	<b>.548(**)</b>	.391(**)	.480(**)	<b>.665(**)</b>		
	Significance (2-tailed)	.000	.000	.000	.000		
	Number of students	559	561	561	560		
SAT 2006 Comprehension Number Correct	Pearson Correlation	<b>.535(**)</b>	.334(**)	<b>.553(**)</b>	<b>.660(**)</b>	<b>.700(**)</b>	
	Significance (2-tailed)	.000	.000	.000	.000	.000	
	Number of students	559	561	561	560	561	
SAT 2006 Total Score	Pearson Correlation	.417(**)	.362(**)	<b>.574(**)</b>	<b>.594(**)</b>	<b>.513(**)</b>	<b>.565(**)</b>
	Significance (2-tailed)	.000	.000	.000	.000	.000	.000
	Number of students	558	560	560	560	560	560

\*\*Correlation is significant at the 0.01 level (2-tailed).

Blue indicates correlations between EOY DIBELS ORF Scores and SAT-10 Scores

Red indicates correlations that exceed .500

## BIE Reading First DIBELS and SAT-10 Data: 2004-2006 School Years (24 Cohort I Schools)

### 3<sup>rd</sup> Grade: Correlations between 2005-2006 EOY DIBELS ORF Scores and SAT-10 Scores

<b>3<sup>rd</sup> Grade 2005-2006</b>		<b>3<sup>rd</sup> EOY ORF</b>	SAT 2006 Phonemic Awareness Number Correct	SAT 2006 Phonics Number Correct	SAT 2006 Vocabulary Number Correct	SAT 2006 Fluency Number Correct	SAT 2006 Comprehension Number Correct
SAT 2006 Phonemic Number Correct	Pearson Correlation	.385(**)					
	Significance (2-tailed)	.000					
	Number of students	532					
SAT 2006 Phonics Number Correct	Pearson Correlation	.344(**)	.429(**)				
	Significance (2-tailed)	.000	.000				
	Number of students	532	541				
SAT 2006 Vocabulary Number Correct	Pearson Correlation	.445(**)	.495(**)	.439(**)			
	Significance (2-tailed)	.000	.000	.000			
	Number of students	532	541	541			
SAT 2006 Fluency Number Correct	Pearson Correlation	.500(**)	.407(**)	.485(**)	.572(**)		
	Significance (2-tailed)	.000	.000	.000	.000		
	Number of students	532	541	541	541		
SAT 2006 Comprehension Number Correct	Pearson Correlation	.520(**)	.443(**)	.504(**)	.610(**)	.681(**)	
	Significance (2-tailed)	.000	.000	.000	.000	.000	
	Number of students	532	541	541	541	541	
SAT 2006 Total Score	Pearson Correlation	.376(**)	.500(**)	.466(**)	.563(**)	.560(**)	.569(**)
	Significance (2-tailed)	.000	.000	.000	.000	.000	.000
	Number of students	532	541	541	541	541	541

\*\*Correlation is significant at the 0.01 level (2-tailed).

Blue indicates correlations between EOY DIBELS ORF Scores and SAT-10 Scores

Red indicates correlations that exceed .500

## **BIE Reading First DIBELS and SAT-10 Data: 2004-2006 School Years (24 Cohort I Schools)**

### **Relationships between 2005-2006 EOY DIBELS Instructional Recommendations and Proficiency on SAT-10 (e.g., 40<sup>th</sup> percentile or higher)**

**OR**

#### **Do EOY DIBELS Instructional Recommendations Predict SAT-10 Outcomes?**

Another measure of test validity is the ability of one test to predict outcomes on another test. For example, do EOY DIBELS Instructional Recommendations for those who are reading at grade level (i.e., Benchmark) or far below grade level (i.e., Intensive) predict proficiency on the SAT-10? Particularly because the EOY DIBELS outcomes are used to make funding decisions, we would hope that the EOY DIBELS results would substantively predict (or be related) to SAT-10 student outcomes.

The following tables show the extent to which the EOY DIBELS instructional recommendations of Benchmark and Intensive were related to whether students were proficient (i.e., at the 40<sup>th</sup> percentile or higher) or not proficient on the SAT-10. The tables show the number of paired scores (included for completeness, as percentages give no indication of the magnitude of sample size or cell sizes) followed by percentages of paired scores. The percentages that are important to note are indicated in **RED** (e.g., Benchmark students attaining proficiency) or **MAGENTA** (e.g., Intensive students who were not proficient).

Note that nearly 90% or more of all students whose instructional recommendation on the EOY DIBELS assessments was Intensive scored below proficiency on the SAT-10, Reading First version. In 1<sup>st</sup> and 2<sup>nd</sup> grades, and the overall total, almost 80% or more of those who were Benchmark on the DIBELS met proficiency on the SAT-10, while only 67% of 3<sup>rd</sup> grade students who were at Benchmark met proficiency on the SAT-10.

Finally, one of the arguments against the DIBELS measures (see Pressley, 2006) is that the correlations with other tests of reading are low. Yet, as the data in this report show, correlations between the DIBELS ORF scores and SAT-10 subtests were at least moderate. Additionally, as subsequent tables show, nearly 80% of students who were at Benchmark on the DIBELS were proficient on the SAT-10, while over 90% of students who were Intensive on the DIBELS were NOT proficient on the SAT-10. These results refute Pressley's claims that were based on data from only 3<sup>rd</sup> grade students and a much smaller sample than is presented here (i.e., a convenience sample from an urban mid-west school with an original n of 250, but a sample size for analysis of less than 130 students that included those who returned permission letters to participate).

## BIE Reading First DIBELS and SAT-10 Data: 2004-2006 School Years (24 Cohort I Schools)

### Relationships between EOY DIBELS Instructional Recommendations and Proficiency on SAT-10 (e.g., 40<sup>th</sup> percentile or higher)

<b>1<sup>st</sup> Grade 2005-2006</b>		<b>SAT-10 (number of paired scores)</b>			<b>SAT-10 (percentages)</b>	
		<b>Proficient</b>	<b>Not Proficient</b>	<b>TOTAL</b>	<b>Proficient</b>	<b>Not Proficient</b>
1 <sup>st</sup> Grade EOY DIBELS Instructional Recommendation	<b>Benchmark</b>	261	44	305	<b>86%</b>	14%
	<b>Strategic</b>	53	125	178	30%	70%
	<b>Intensive</b>	11	92	103	11%	<b>89%</b>
	<b>TOTAL</b>	325	261	586		

<b>2<sup>nd</sup> Grade 2005-2006</b>		<b>SAT-10 (number of paired scores)</b>			<b>SAT-10 (percentages)</b>	
		<b>Proficient</b>	<b>Not Proficient</b>	<b>TOTAL</b>	<b>Proficient</b>	<b>Not Proficient</b>
2 <sup>nd</sup> Grade EOY DIBELS Instructional Recommendation	<b>Benchmark</b>	202	50	252	<b>80%</b>	20%
	<b>Strategic</b>	52	70	122	43%	57%
	<b>Intensive</b>	15	166	181	8%	<b>92%</b>
	<b>TOTAL</b>	269	286	555		

<b>3<sup>rd</sup> Grade 2005-2006</b>		<b>SAT-10 (number of paired scores)</b>			<b>SAT-10 (percentages)</b>	
		<b>Proficient</b>	<b>Not Proficient</b>	<b>TOTAL</b>	<b>Proficient</b>	<b>Not Proficient</b>
3 <sup>rd</sup> Grade EOY DIBELS Instructional Recommendation	<b>Benchmark</b>	147	73	220	<b>67%</b>	33%
	<b>Strategic</b>	45	132	177	25%	75%
	<b>Intensive</b>	7	121	128	5%	<b>95%</b>
	<b>TOTAL</b>	199	326	525		

<b>Grades 1-3 2005-2006</b>		<b>SAT-10 (number of paired scores)</b>			<b>SAT-10 (percentages)</b>	
		<b>Proficient</b>	<b>Not Proficient</b>	<b>TOTAL</b>	<b>Proficient</b>	<b>Not Proficient</b>
Total Grades 1-3 EOY DIBELS Instructional Recommendation	<b>Benchmark</b>	610	167	777	<b>79%</b>	21%
	<b>Strategic</b>	150	327	477	31%	69%
	<b>Intensive</b>	33	379	412	8%	<b>92%</b>
	<b>TOTAL</b>	793	873	1666		



# **BIE Reading First DIBELS and SAT-10 Data: 2004-2006 School Years (24 Cohort I Schools)**

## **Relationships between EOY DIBELS Instructional Recommendations and SAT-10 Outcomes by Subtest**

The following tables show the extent to which the EOY DIBELS instructional recommendations of Benchmark and Intensive were related to whether students were At Grade Level (AGL) or Need Substantial Intervention (NSI) on SAT-10 subtests of the Reading First version.

Again, the tables show the number of paired scores followed by percentages of paired scores. The percentages that are important to note are indicated in **RED** (e.g., Benchmark students attaining proficiency) or **MAGENTA** (e.g., Intensive students who were not proficient).

While the highlighted percentages vary, most indicate a moderate to high relationship, with the percentages of students with paired scores who ended the year at Benchmark based on DIBELS scores and were AGL on the SAT-10 subtests ranging from 42% to 99%, depending on grade level and SAT-10 subtests. The percentages of students with paired scores who ended the year Intensive on the DIBELS and NSI on SAT-10 subtests ranged from 2% (yes, you read that correctly! Just 2% on the Phonemic Awareness subtest in 2<sup>nd</sup> grade) to 77%.

# BIE Reading First DIBELS and SAT-10 Data: 2004-2006 School Years (24 Cohort I Schools)

## Relationships between EOY DIBELS Instructional Recommendations and SAT-10 PHONEMIC AWARENESS

<b>1<sup>st</sup> Grade 2005-2006 Phonemic Awareness</b>		<b>SAT-10 Phonemic Awareness (number of paired scores)</b>				<b>SAT-10 Phonemic Awareness (percent of paired scores)</b>		
		<b>AGL</b>	<b>NAI</b>	<b>NSI</b>	<b>TOTAL</b>	<b>AGL</b>	<b>NAI</b>	<b>NSI</b>
1 <sup>st</sup> Grade EOY DIBELS Instructional Recommendation	<b>Benchmark</b>	288	18	1	307	<b>94%</b>	6%	0%
	<b>Strategic</b>	148	32	5	185	80%	17%	3%
	<b>Intensive</b>	40	36	26	102	39%	35%	<b>25%</b>
	<b>TOTAL</b>	476	86	32	594			

<b>2<sup>nd</sup> Grade 2005-2006 Phonemic Awareness</b>		<b>SAT-10 Phonemic Awareness (number of paired scores)</b>				<b>SAT-10 Phonemic Awareness (percent of paired scores)</b>		
		<b>AGL</b>	<b>NAI</b>	<b>NSI</b>	<b>TOTAL</b>	<b>AGL</b>	<b>NAI</b>	<b>NSI</b>
2 <sup>nd</sup> Grade EOY DIBELS Instructional Recommendation	<b>Benchmark</b>	252	2	0	254	<b>99%</b>	1%	0%
	<b>Strategic</b>	116	5	2	123	94%	4%	2%
	<b>Intensive</b>	165	12	4	181	91%	7%	<b>2%</b>
	<b>TOTAL</b>	533	19	6	558			

<b>3<sup>rd</sup> Grade 2005-2006 Phonemic Awareness</b>		<b>SAT-10 Phonemic Awareness (number of paired scores)</b>				<b>SAT-10 Phonemic Awareness (percent of paired scores)</b>		
		<b>AGL</b>	<b>NAI</b>	<b>NSI</b>	<b>TOTAL</b>	<b>AGL</b>	<b>NAI</b>	<b>NSI</b>
3 <sup>rd</sup> Grade EOY DIBELS Instructional Recommendation	<b>Benchmark</b>	142	36	46	224	<b>63%</b>	16%	21%
	<b>Strategic</b>	83	39	58	180	46%	22%	32%
	<b>Intensive</b>	35	20	73	128	27%	16%	<b>57%</b>
	<b>TOTAL</b>	260	95	177	532			

<b>Grades 1-3 2005-2006 Phonemic Awareness</b>		<b>SAT-10 Phonemic Awareness (number of paired scores)</b>				<b>SAT-10 Phonemic Awareness (percent of paired scores)</b>		
		<b>AGL</b>	<b>NAI</b>	<b>NSI</b>	<b>TOTAL</b>	<b>AGL</b>	<b>NAI</b>	<b>NSI</b>
Total Grades 1-3 EOY DIBELS Instructional Recommendation	<b>Benchmark</b>	682	56	47	785	<b>87%</b>	7%	6%
	<b>Strategic</b>	347	76	65	488	71%	16%	13%
	<b>Intensive</b>	240	68	103	411	58%	17%	<b>25%</b>
	<b>TOTAL</b>	1269	200	215	1684			

## BIE Reading First DIBELS and SAT-10 Data: 2004-2006 School Years (24 Cohort I Schools)

### Relationships between EOY DIBELS Instructional Recommendations and SAT-10 PHONICS

1 <sup>st</sup> Grade 2005-2006 Phonics		SAT-10 Phonics (number of paired scores)				SAT-10 Phonics (percent of paired scores)		
		AGL	NAI	NSI	TOTAL	AGL	NAI	NSI
		Benchmark	150	113	44	307	49%	37%
EOY DIBELS Instructional Recommendation	Strategic	32	91	62	185	17%	49%	34%
	Intensive	7	40	55	102	7%	39%	54%
	TOTAL	189	244	161	594			

2 <sup>nd</sup> Grade 2005-2006 Phonics		SAT-10 Phonics (number of paired scores)				SAT-10 Phonics (percent of paired scores)		
		AGL	NAI	NSI	TOTAL	AGL	NAI	NSI
		Benchmark	163	27	63	253	64%	11%
EOY DIBELS Instructional Recommendation	Strategic	59	13	51	123	48%	11%	41%
	Intensive	40	17	124	181	22%	9%	69%
	TOTAL	262	57	238	557			

3 <sup>rd</sup> Grade 2005-2006 Phonics		SAT-10 Phonics (number of paired scores)				SAT-10 Phonics (percent of paired scores)		
		AGL	NAI	NSI	TOTAL	AGL	NAI	NSI
		Benchmark	158	21	45	224	71%	9%
EOY DIBELS Instructional Recommendation	Strategic	92	31	57	180	51%	17%	32%
	Intensive	44	28	56	128	34%	22%	44%
	TOTAL	294	80	158	532			

Grades 1-3 2005-2006 Phonics		SAT-10 Phonics (number of paired scores)				SAT-10 Phonics (percent of paired scores)		
		AGL	NAI	NSI	TOTAL	AGL	NAI	NSI
		Benchmark	471	161	152	784	60%	21%
Total Grades 1-3 EOY DIBELS Instructional Recommendation	Strategic	183	135	170	488	38%	28%	35%
	Intensive	91	85	235	411	22%	21%	57%
	TOTAL	745	381	557	1683			

# BIE Reading First DIBELS and SAT-10 Data: 2004-2006 School Years (24 Cohort I Schools)

## Relationships between EOY DIBELS Instructional Recommendations and SAT-10 VOCABULARY

1 <sup>st</sup> Grade 2005-2006 Vocabulary		SAT-10 Vocabulary (number of paired scores)				SAT-10 Vocabulary (percent of paired scores)		
		AGL	NAI	NSI	TOTAL	AGL	NAI	NSI
1 <sup>st</sup> Grade EOY DIBELS Instructional Recommendation	<b>Benchmark</b>	218	58	31	307	<b>71%</b>	19%	10%
	<b>Strategic</b>	50	46	89	185	27%	25%	48%
	<b>Intensive</b>	17	18	67	102	17%	18%	<b>66%</b>
	<b>TOTAL</b>	285	122	187	594			

2 <sup>nd</sup> Grade 2005-2006 Vocabulary		SAT-10 Vocabulary (number of paired scores)				SAT-10 Vocabulary (percent of paired scores)		
		AGL	NAI	NSI	TOTAL	AGL	NAI	NSI
2 <sup>nd</sup> Grade EOY DIBELS Instructional Recommendation	<b>Benchmark</b>	169	45	40	254	<b>67%</b>	18%	16%
	<b>Strategic</b>	43	29	51	123	35%	24%	41%
	<b>Intensive</b>	13	32	136	181	7%	18%	<b>75%</b>
	<b>TOTAL</b>	225	106	227	558			

3 <sup>rd</sup> Grade 2005-2006 Vocabulary		SAT-10 Vocabulary (number of paired scores)				SAT-10 Vocabulary (percent of paired scores)		
		AGL	NAI	NSI	TOTAL	AGL	NAI	NSI
3 <sup>rd</sup> Grade EOY DIBELS Instructional Recommendation	<b>Benchmark</b>	101	74	49	224	<b>45%</b>	33%	22%
	<b>Strategic</b>	37	69	74	180	21%	38%	41%
	<b>Intensive</b>	15	31	82	128	12%	24%	<b>64%</b>
	<b>TOTAL</b>	153	174	205	532			

Grades 1-3 2005-2006 Vocabulary		SAT-10 Vocabulary (number of paired scores)				SAT-10 Vocabulary (percent of paired scores)		
		AGL	NAI	NSI	TOTAL	AGL	NAI	NSI
Total Grades 1-3 EOY DIBELS Instructional Recommendation	<b>Benchmark</b>	488	177	120	785	<b>62%</b>	23%	15%
	<b>Strategic</b>	130	144	214	488	27%	30%	44%
	<b>Intensive</b>	45	81	285	411	11%	20%	<b>69%</b>
	<b>TOTAL</b>	663	402	619	1684			

# BIE Reading First DIBELS and SAT-10 Data: 2004-2006 School Years (24 Cohort I Schools)

## Relationships between EOY DIBELS Instructional Recommendations and SAT-10 FLUENCY

<b>1<sup>st</sup> Grade 2005-2006 Fluency</b>		<b>SAT-10 Fluency (number of paired scores)</b>				<b>SAT-10 Fluency (percent of paired scores)</b>		
		<b>AGL</b>	<b>NAI</b>	<b>NSI</b>	<b>TOTAL</b>	<b>AGL</b>	<b>NAI</b>	<b>NSI</b>
1 <sup>st</sup> Grade EOY DIBELS Instructional Recommendation	<b>Benchmark</b>	238	43	26	307	<b>78%</b>	14%	8%
	<b>Strategic</b>	58	67	60	185	31%	36%	32%
	<b>Intensive</b>	19	29	54	102	19%	28%	<b>53%</b>
	<b>TOTAL</b>	<b>315</b>	<b>139</b>	<b>140</b>	<b>594</b>			

<b>2<sup>nd</sup> Grade 2005-2006 Fluency</b>		<b>SAT-10 Fluency (number of paired scores)</b>				<b>SAT-10 Fluency (percent of paired scores)</b>		
		<b>AGL</b>	<b>NAI</b>	<b>NSI</b>	<b>TOTAL</b>	<b>AGL</b>	<b>NAI</b>	<b>NSI</b>
2 <sup>nd</sup> Grade EOY DIBELS Instructional Recommendation	<b>Benchmark</b>	161	56	37	254	<b>63%</b>	22%	15%
	<b>Strategic</b>	41	45	37	123	33%	37%	30%
	<b>Intensive</b>	26	33	122	181	14%	18%	<b>67%</b>
	<b>TOTAL</b>	<b>228</b>	<b>134</b>	<b>196</b>	<b>558</b>			

<b>3<sup>rd</sup> Grade 2005-2006 Fluency</b>		<b>SAT-10 Fluency (number of paired scores)</b>				<b>SAT-10 Fluency (percent of paired scores)</b>		
		<b>AGL</b>	<b>NAI</b>	<b>NSI</b>	<b>TOTAL</b>	<b>AGL</b>	<b>NAI</b>	<b>NSI</b>
3 <sup>rd</sup> Grade EOY DIBELS Instructional Recommendation	<b>Benchmark</b>	156	36	32	224	<b>70%</b>	16%	14%
	<b>Strategic</b>	80	39	61	180	44%	22%	34%
	<b>Intensive</b>	25	33	70	128	20%	26%	<b>55%</b>
	<b>TOTAL</b>	<b>261</b>	<b>108</b>	<b>163</b>	<b>532</b>			

<b>Grades 1-3 2005-2006 Fluency</b>		<b>SAT-10 Fluency (number of paired scores)</b>				<b>SAT-10 Fluency (percent of paired scores)</b>		
		<b>AGL</b>	<b>NAI</b>	<b>NSI</b>	<b>TOTAL</b>	<b>AGL</b>	<b>NAI</b>	<b>NSI</b>
Total Grades 1-3 EOY DIBELS Instructional Recommendation	<b>Benchmark</b>	555	135	95	785	<b>71%</b>	17%	12%
	<b>Strategic</b>	179	151	158	488	37%	31%	32%
	<b>Intensive</b>	70	95	246	411	17%	23%	<b>60%</b>
	<b>TOTAL</b>	<b>804</b>	<b>381</b>	<b>499</b>	<b>1684</b>			

## BIE Reading First DIBELS and SAT-10 Data: 2004-2006 School Years (24 Cohort I Schools)

### Relationships between EOY DIBELS Instructional Recommendations and SAT-10 COMPREHENSION

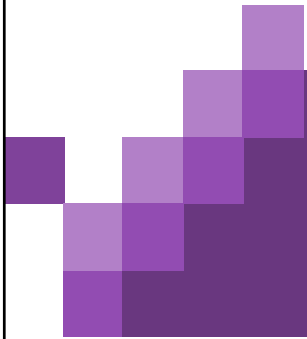
1 <sup>st</sup> Grade 2005-2006 Comprehension		SAT-10 Comprehension (number of paired scores)				SAT-10 Comprehension (percent of paired scores)		
		AGL	NAI	NSI	TOTAL	AGL	NAI	NSI
1 <sup>st</sup> Grade EOY DIBELS Instructional Recommendation	<b>Benchmark</b>	274	26	7	307	<b>89%</b>	8%	2%
	<b>Strategic</b>	92	65	28	185	50%	35%	15%
	<b>Intensive</b>	21	33	48	102	21%	32%	<b>47%</b>
	<b>TOTAL</b>	<b>387</b>	<b>124</b>	<b>83</b>	<b>594</b>			

2 <sup>nd</sup> Grade 2005-2006 Comprehension		SAT-10 Comprehension (number of paired scores)				SAT-10 Comprehension (percent of paired scores)		
		AGL	NAI	NSI	TOTAL	AGL	NAI	NSI
2 <sup>nd</sup> Grade EOY DIBELS Instructional Recommendation	<b>Benchmark</b>	127	69	58	254	<b>50%</b>	27%	23%
	<b>Strategic</b>	27	41	55	123	22%	33%	45%
	<b>Intensive</b>	12	30	139	181	7%	17%	<b>77%</b>
	<b>TOTAL</b>	<b>166</b>	<b>140</b>	<b>252</b>	<b>558</b>			

3 <sup>rd</sup> Grade 2005-2006 Comprehension		SAT-10 Comprehension (number of paired scores)				SAT-10 Comprehension (percent of paired scores)		
		AGL	NAI	NSI	TOTAL	AGL	NAI	NSI
3 <sup>rd</sup> Grade EOY DIBELS Instructional Recommendation	<b>Benchmark</b>	94	79	51	224	<b>42%</b>	35%	23%
	<b>Strategic</b>	24	78	78	180	13%	43%	43%
	<b>Intensive</b>	1	34	93	128	1%	27%	<b>73%</b>
	<b>TOTAL</b>	<b>119</b>	<b>191</b>	<b>222</b>	<b>532</b>			



Grades 1-3 2005-2006 Comprehension		SAT-10 Comprehension (number of paired scores)				SAT-10 Comprehension (percent of paired scores)		
		AGL	NAI	NSI	TOTAL	AGL	NAI	NSI
Total Grades 1-3 EOY DIBELS Instructional Recommendation	<b>Benchmark</b>	495	174	116	785	<b>63%</b>	22%	15%
	<b>Strategic</b>	143	184	161	488	29%	38%	33%
	<b>Intensive</b>	34	97	280	411	8%	24%	<b>68%</b>
	<b>TOTAL</b>	<b>672</b>	<b>455</b>	<b>557</b>	<b>1684</b>			

## **Appendix F: 2006-2006 Reading First Evaluation Report to the BIE's Leadership Team**



# BIE Reading First Leadership Meeting October 12, 2007

EndVision Research & Evaluation, LLC  
Catherine Callow-Heusser




## Evaluation Data Sources

- DIBELS Benchmark Scores
- Interviews and surveys
  - Principals or Assistant Principals
  - Reading Coaches
  - Teachers
  - Paraprofessionals
- Classroom observations
  - Use of effective teaching practices
  - Fidelity of implementation
- Parent meetings
- SAT-10 RF scores
- Wireless Generation and University of Oregon Websites
  - *Student demographics*

Measure ...

- Successes
- Challenges
- Training Needs
- School Climate
- Demographics
- Outcomes



2



## Classroom Observations

- 15-20 minute “snapshots”
- Effective teaching practices
  - Instruction
  - Teaching functions
  - Academic monitoring
  - Time management
  - Classroom or behavior management
- Reading program fidelity of implementation
- Student engagement



3

## Fidelity of Implementation: Classroom

- High: Systematic, explicit instruction was consistently observed
  - Core/supplemental reading program materials were used,
  - Activities were aligned with the critical elements,
  - Instructional time was used efficiently, and
  - The classroom learning environment
    - Maximized opportunities for learning,
    - Included appropriate modeling and instructional feedback,
    - Was positive and appropriate behavior was acknowledged/rewarded,
    - Included frequent opportunities for all students in the instructional group to demonstrate mastery of concepts, and
    - Included observably high levels of student engagement.



4

## Fidelity of Implementation: Classroom

- Fair/Moderate: Reading instruction occurred but
  - Instruction was not consistently systematic and explicit,
  - Instructional time was used inefficiently, and/or
  - Some activities were not aligned with the critical elements of reading.
- Poor: Instruction related to reading occurred but
  - Activities were not aligned with the critical elements,
  - Instruction was not systematic and explicit,
  - Academic learning time was limited, and/or
  - Activities included use of non-core/supplemental program materials.
- None: No reading instruction was observed.



5

## Classroom Observation Purpose


- Formative Evaluation → *Improve Programs*
  - Identify what's working
  - Recommend improvements
- Summative Evaluation → *Determine Outcomes, Factors Related to Outcomes*
  - Measure changes in teaching practices and student engagement
  - Identify factors related to student outcomes
    - For example, 50% of variability in student scores related to ratings of effective teaching and fidelity of implementation



6

**Praise Form**

**Immediate Feedback**



EndVision Research & Evaluation  
dibels@endvision.net

When visiting your classroom today, I saw...

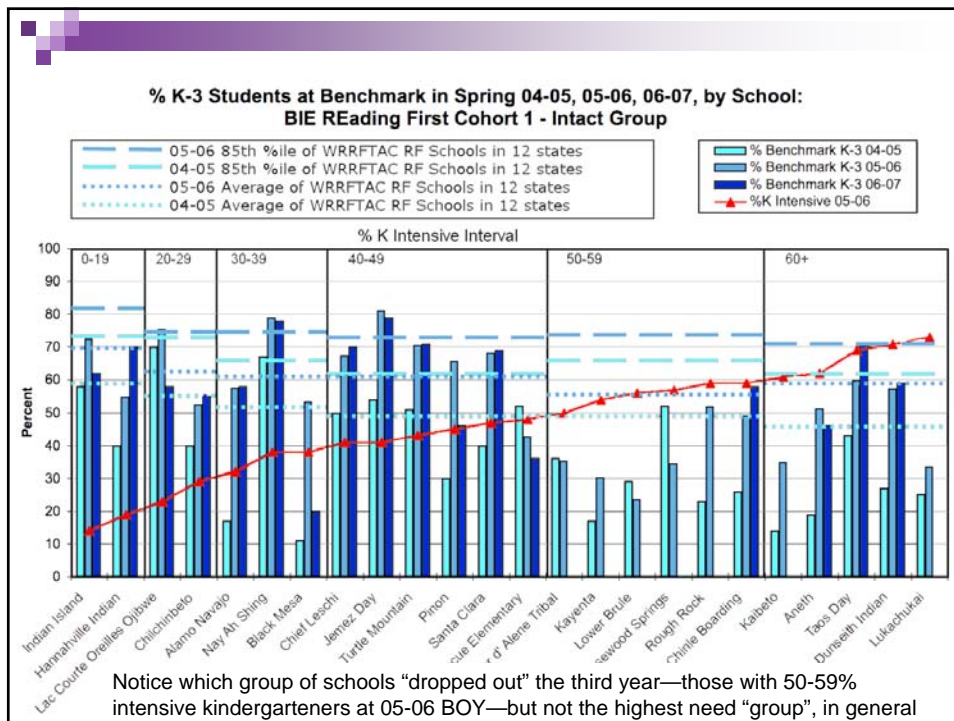
- ☺ *Excellent use of modeling, e.g., "My turn, your turn"*
- ☺ *ALL students responding during group responses, with appropriate pacing and management of student behavior*
- ☺ *Well prepared lesson with quick pacing, many short transitions*

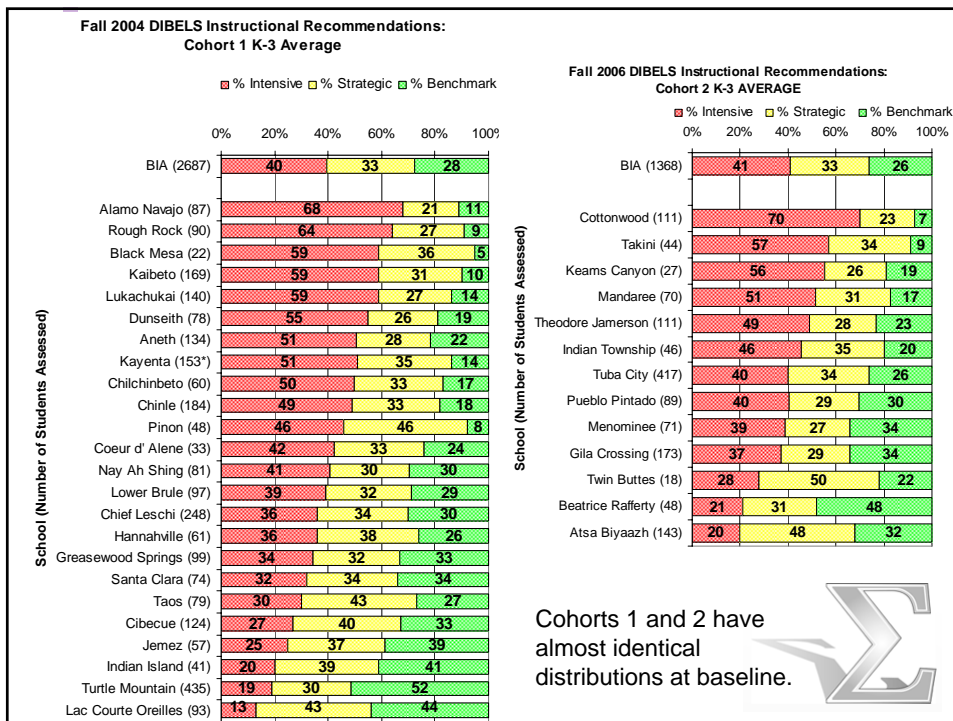
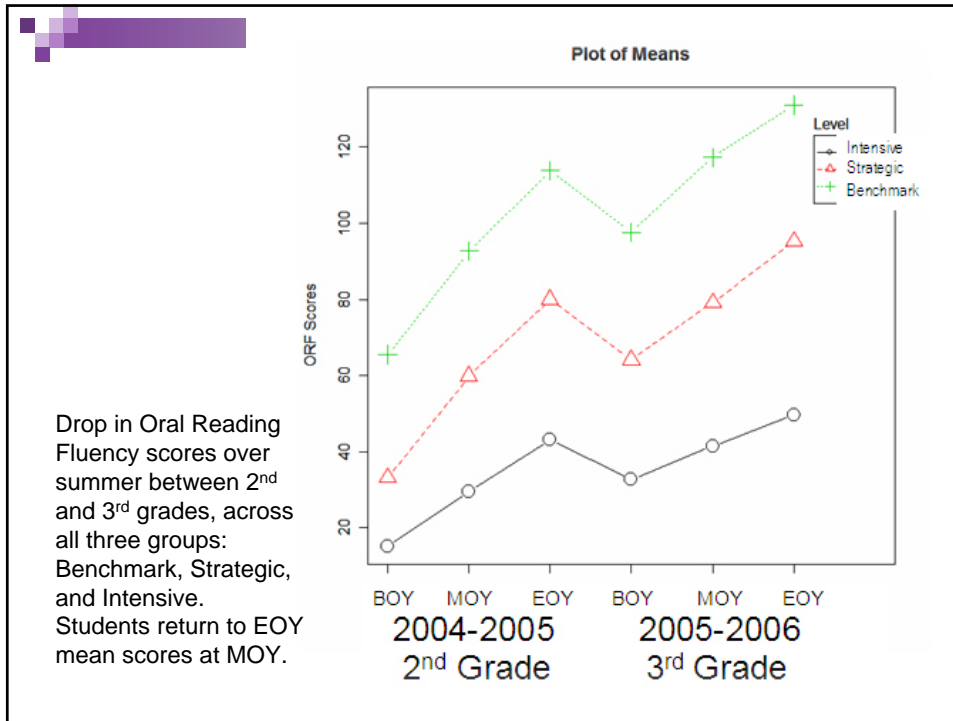
Opportunities for students to learn can be increased by...

*Praising appropriate behavior to set expectation for all. "I like the way Dakota is sitting quietly with her eyes on me and listening to my instructions."*

Catherine Heusser  
Signature

08-24-06  
Date





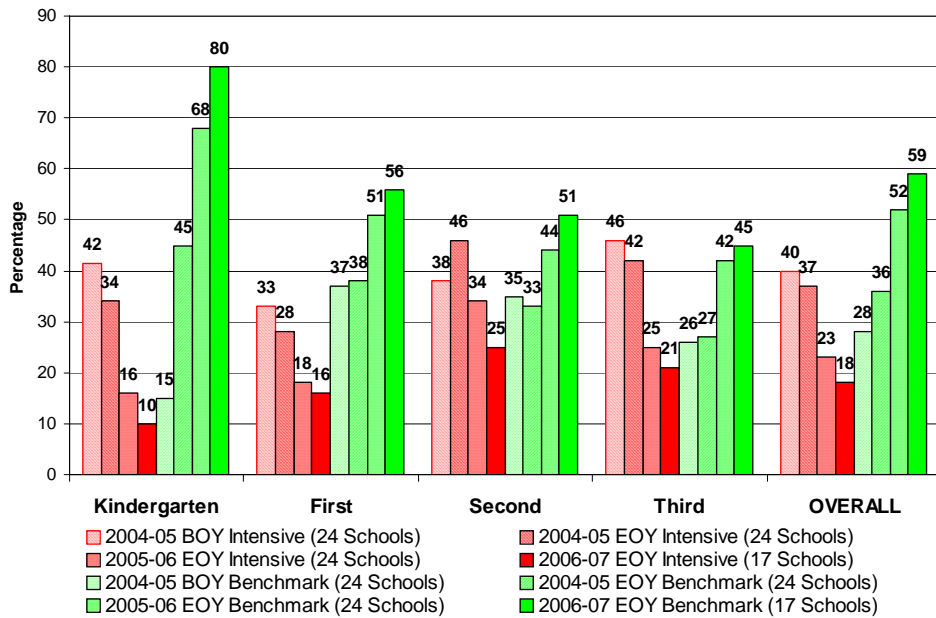
## Percent Progress

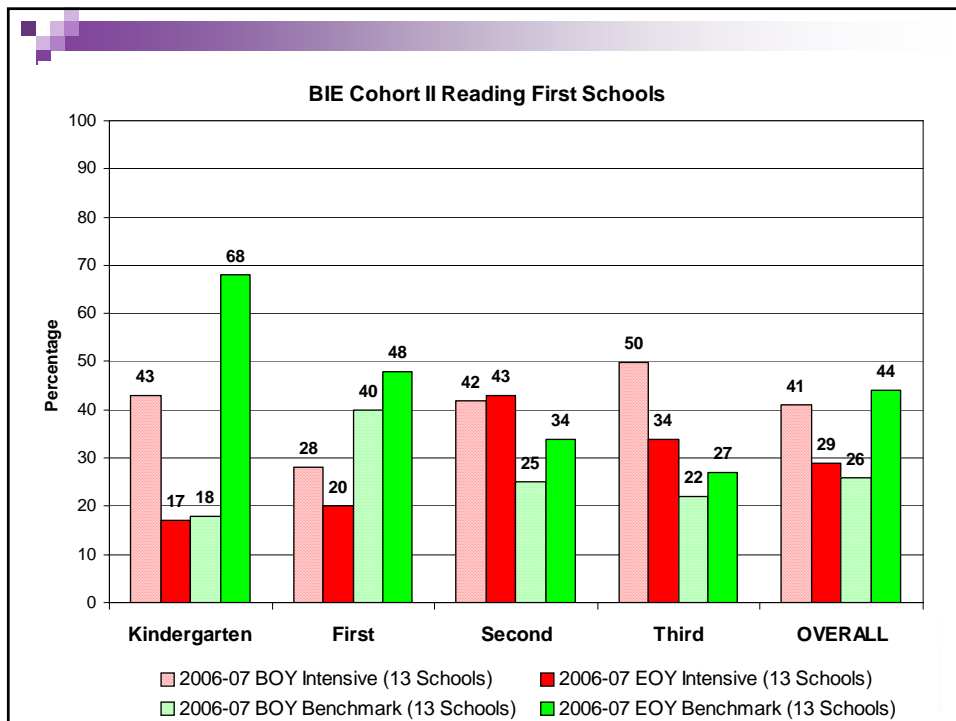
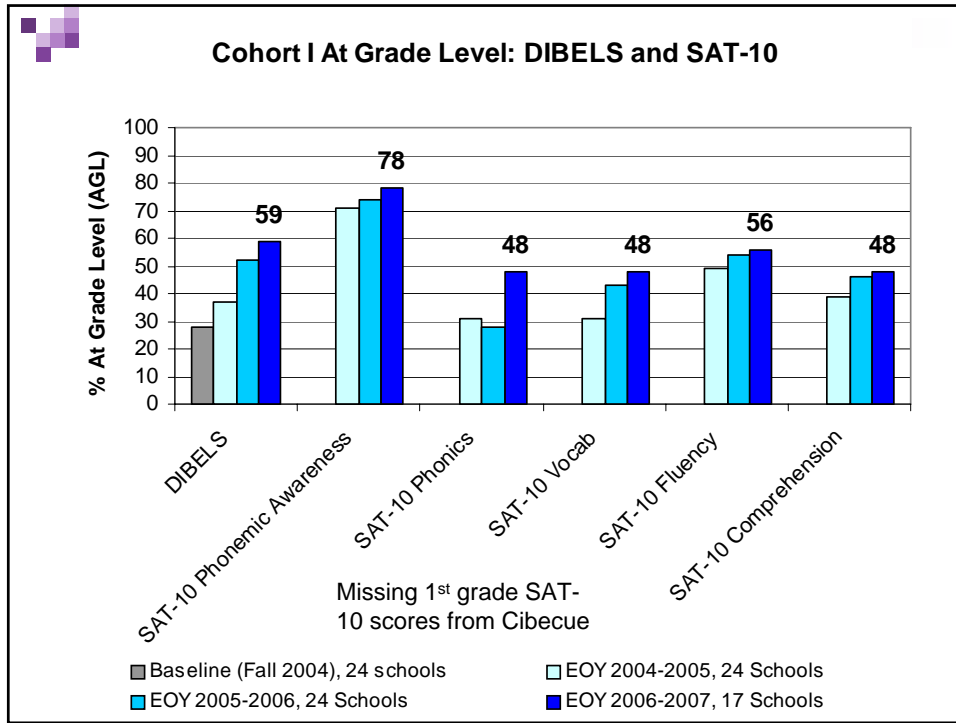
- Based on DIBELS Instructional Recommendations
  - Students who started and ended at Benchmark
  - Students who moved from Strategic to Benchmark
  - Students who moved from Intensive to Benchmark
  - Students who moved from Intensive to Strategic
- 2006-2007 Percent Progress
  - BIE Cohort I: **67%**
    - Top Performing: Nay Ah Shing 81%, Jemez 80%
    - Lowest Performing: Cibecue 33%
  - BIE Cohort II: **55%**
    - Cohort I in Year 1: 45%
    - Top Performing: Mandaree 74%, Theodore Jamerson 65%
    - Lowest Performing: Keams Canyon 46%

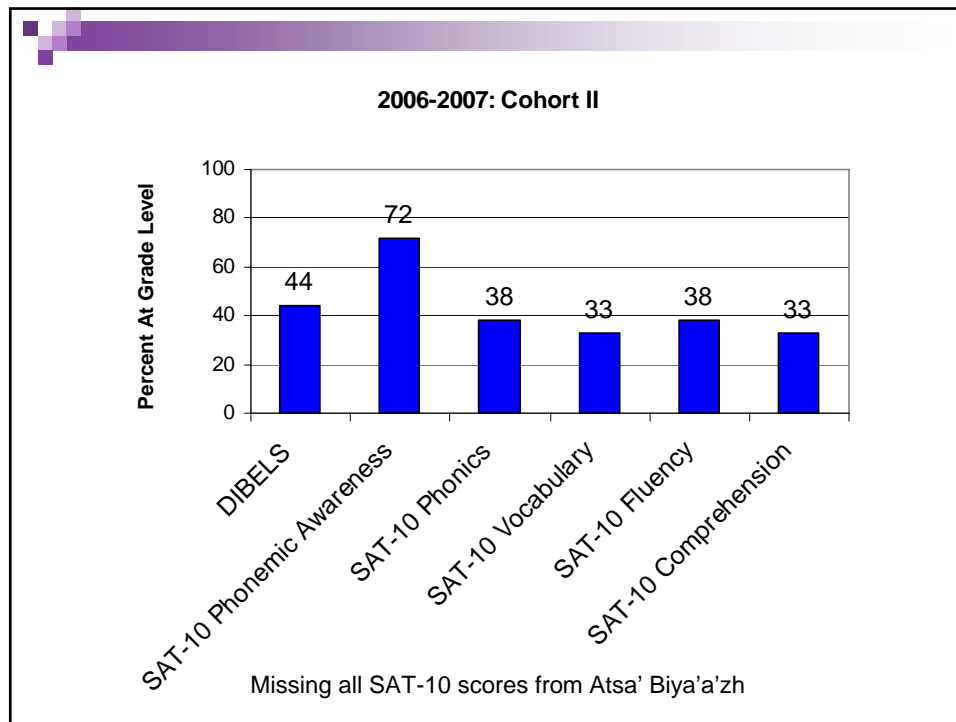


11

BIE Cohort I Reading First Schools







## Reading First Successes

- Increased engagement of instructional leadership
- Increased fidelity of implementation
- Increased progress monitoring with the DIBELS and use of data to adjust intervention
- Increased knowledge of research-based and effective reading instruction
- Increased use of effective teaching practices
- Increased student engagement
- Increased numbers of students reading at Benchmark or improving based on the DIBELS instructional recommendations
- Professional development and technical assistance that has changed over time to better meet needs



## “Spill-Over” Successes

- Spread of reading programs and DIBELS assessments to 4<sup>th</sup>-6<sup>th</sup> or 4<sup>th</sup>-12<sup>th</sup> grades
- Increased use of intervention programs and intervention models in upper grades
- Spread of “Reading First”-like goals, programs, instructional strategies, and assessments to other content areas, e.g., writing/language arts, math, science
- Increased instructional leadership and use of coaches in upper grades

17



## Cohort I—Turtle Mountain

- Special Education Referrals, Grades K-5

	Assessed	Placed
2003-2004	52	31
2004-2005	31	13
2005-2006	13	4

18





## Challenges

- Reading First staff who are resistant or not committed
  - Poor fidelity of implementation
  - Limited academic learning time
  - Increased staff absences
  - Staff turnover
- “Timely” technical assistance and professional development with reduced time away from school
- Understanding and using data appropriately
- Communication

19



## Additional Training Needs

- Implementing WRRFTAC templates with fidelity
- Recently implemented programs
  - Differentiated instruction (ESL, disabilities)
  - Efficient and effective use of time during reading block
- Behavior management
- Additional intervention (e.g., double/triple dosing)
- Using progress data to plan instruction

20



## Factors Affecting Outcomes

### ■ Positive

- Same principal and reading coach both years
- Engaged leadership (classroom observations)
- Scripted reading or intervention program (e.g., Reading Mastery, ERI, Voyager Passport), and/or WRRFTAC templates implemented well

### ■ Negative

- High staff turnover, particularly at leadership levels
- Conflict within school
- Resistance to Reading First or reading programs



21

## Schools Above BIE Average Percent Progress

<b>BIE Cohort I</b>	<b>67</b>	Same Principal	Same RC	DI/Templates	Reading Programs
* Nay Ah Shing	<b>81</b>	✓	✓	✓	RM
Jemez	<b>80</b>	✓	✓		SFA
Turtle Mountain	<b>75</b>	✓	✓	✓	MH/VP/RM
Dunseith	<b>75</b>	✓		✓	HM/VP
Chief Leschi	<b>74</b>	✓	✓	✓	HB/RM/RW/RN
Hannahville	<b>73</b>	✓	✓	✓	HM/VP
Taos	<b>73</b>	✓	✓	✓	HM/RN/VP
Santa Clara	<b>70</b>		✓	✓	HM
* Alamo Navajo	<b>70</b>	✓	✓-	✓	SF/RN/RM/ERI
Indian Island	<b>66</b>	✓	✓	✓	HM/RM
LCO	<b>65</b>	✓-	✓	✓	HM
* Chinle	<b>61</b>			✓	RM

\* Started below BIE average for percent Intensive students

## Cohort II Overall Findings

- In general, Cohort II schools are at the same place Cohort I schools were after 1.5 years
  - Better early implementation, including templates
  - Increased use of effective teaching practices
  - Double and triple dosing are occurring
  - Replacement core programs already being implemented
  - Consistent progress monitoring occurring, and schools are using data
  - Less resistance, more buy-in
  - Increased collaboration, planning
  - On-site technical assistance working
  - Excitement about student successes, improved teaching
- Challenges
  - Training, technical assistance needs
  - Staff turnover

23



## Cohort I and Cohort II Differences

- Timing (i.e., NCLB, 3<sup>rd</sup> year of BIE Reading First)
- Time to prepare
- Cohort II heard Cohort I stories
- BIE Reading First Leadership “Lessons Learned”
  - Professional development
  - Technical assistance (including Cohort I coaches)
  - Reading programs
  - Timing of replacement core, intervention programs
  - Timing of tiered reading instruction, double and triple dosing

24

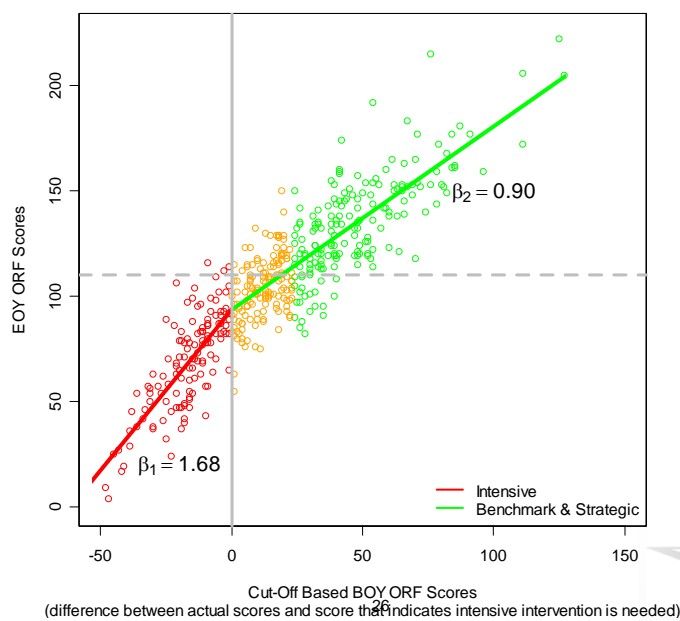


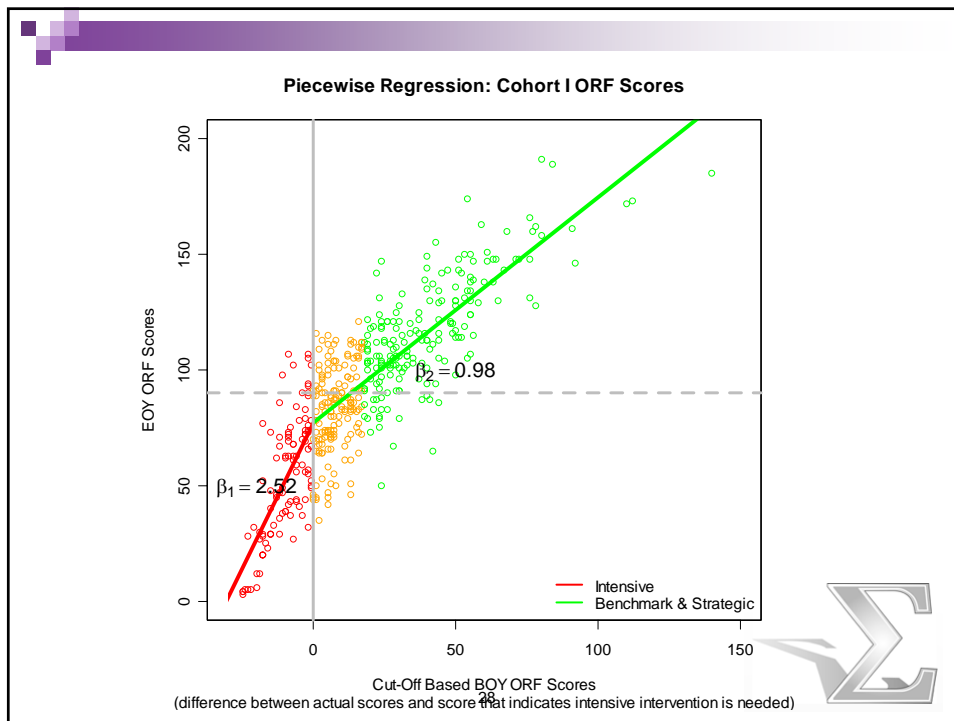
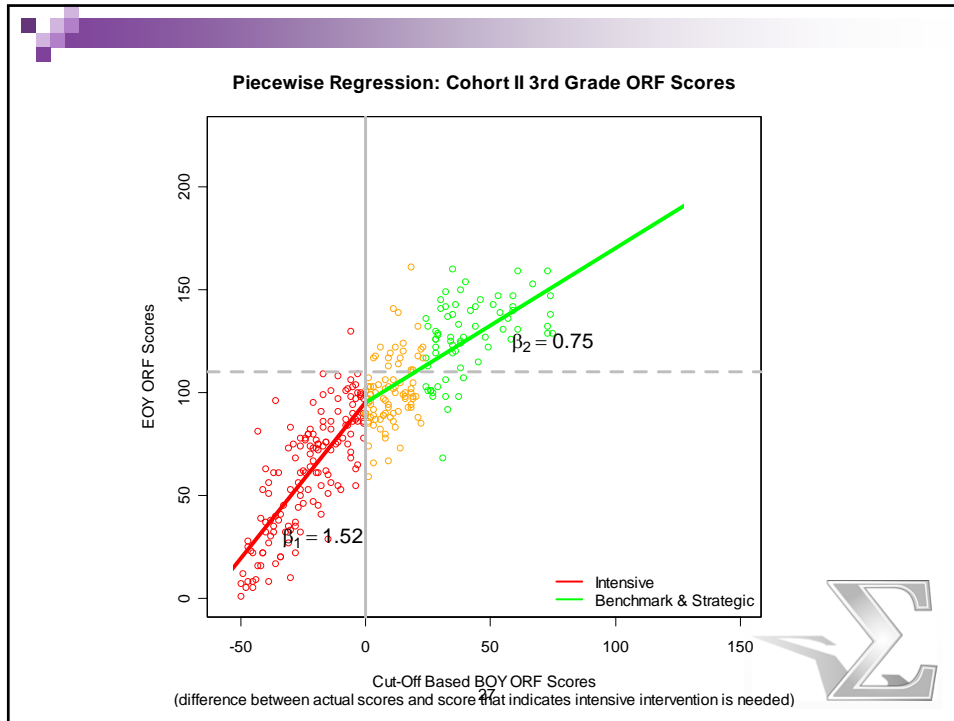
Do students who begin the school year INTENSIVE on the DIBELS receive intervention that is 'different' than students who are Strategic or Benchmark?

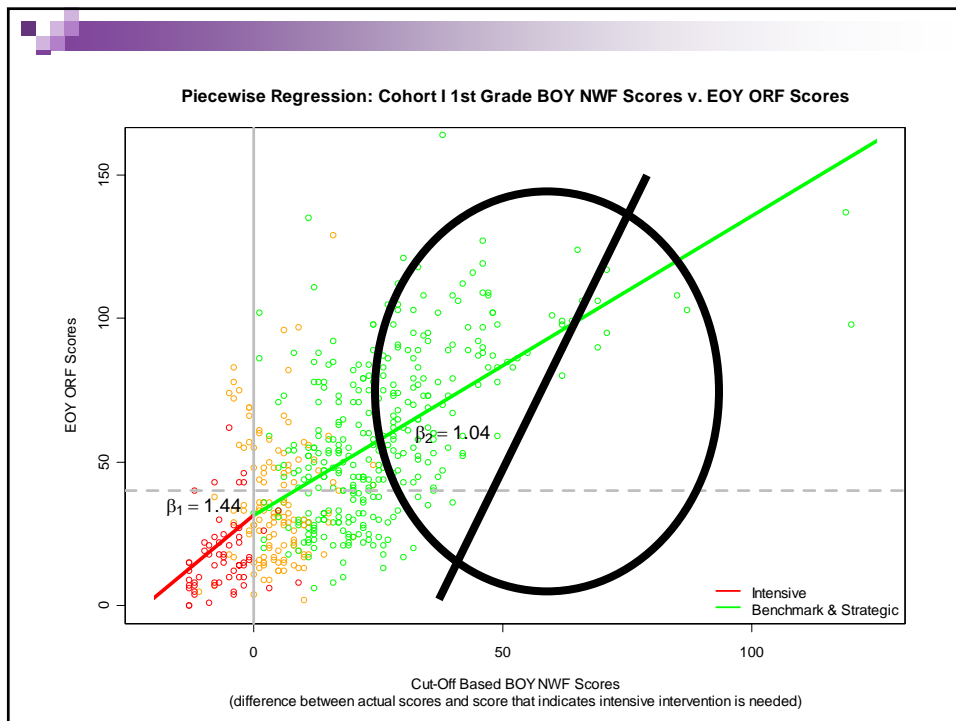
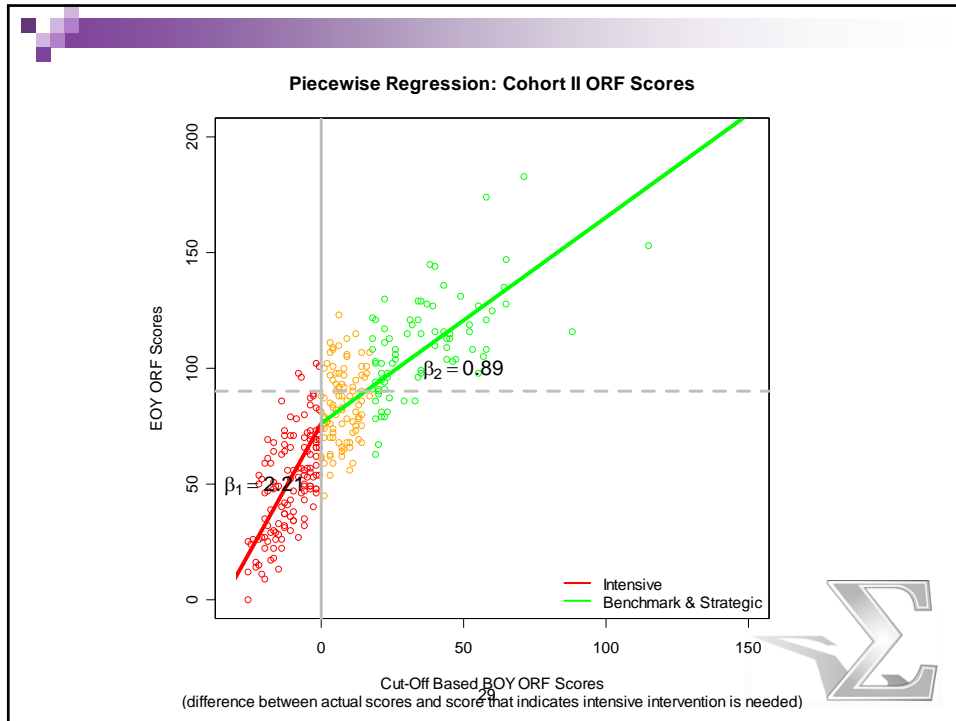
25

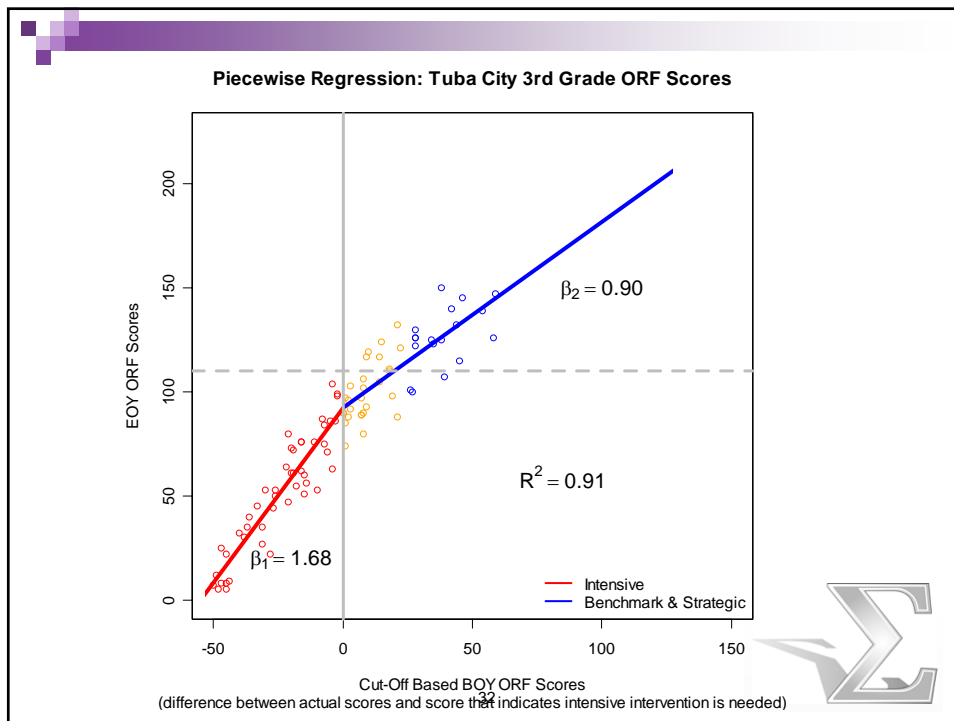
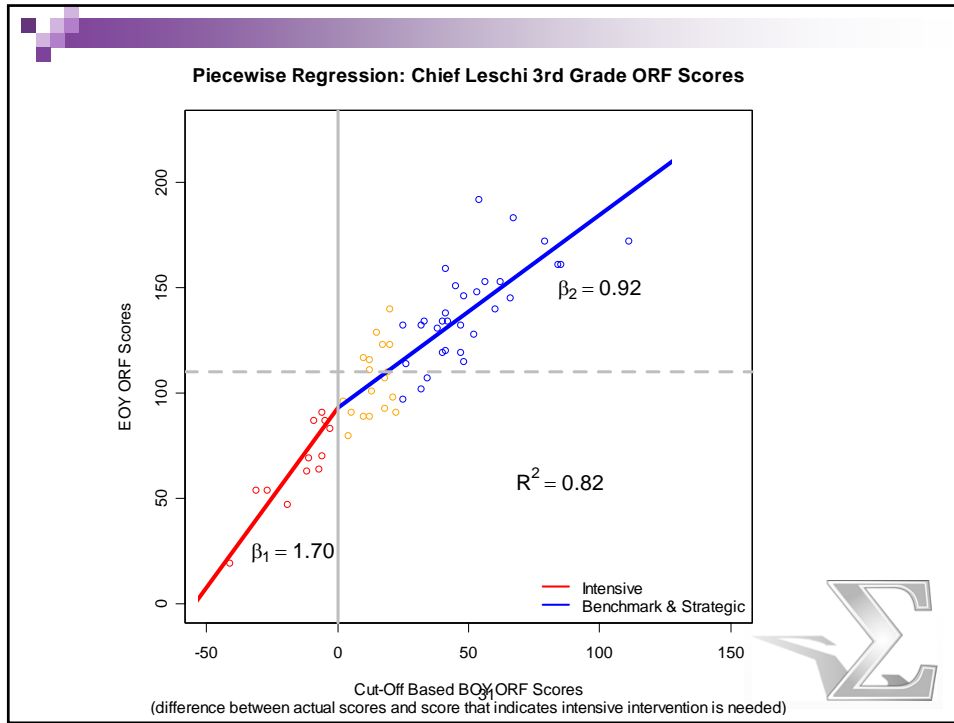


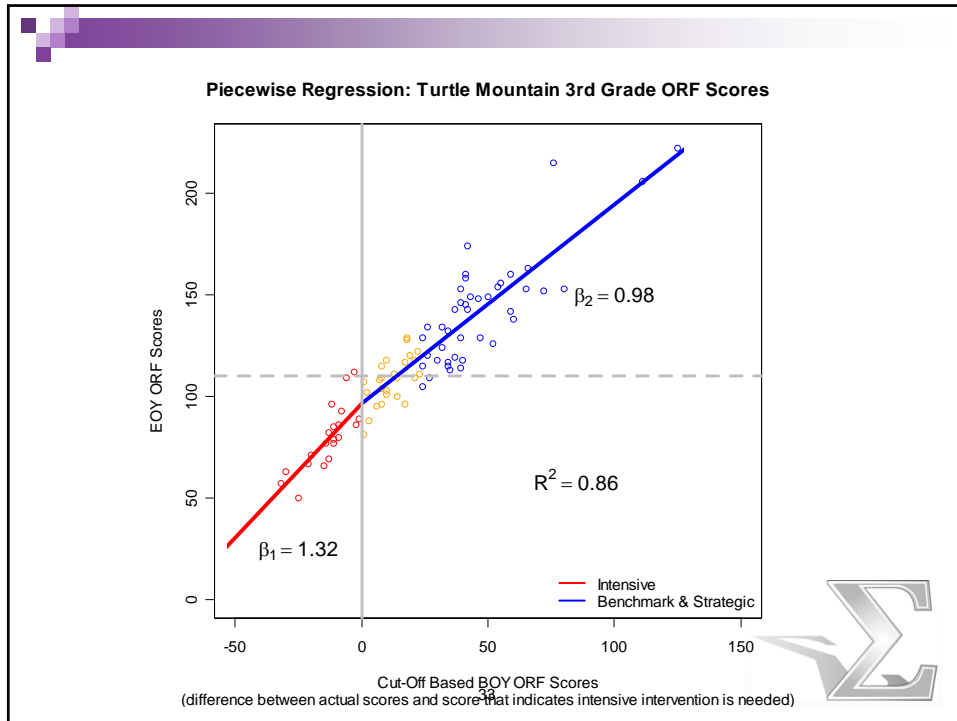
Piecewise Regression: Cohort I 3rd Grade ORF Scores





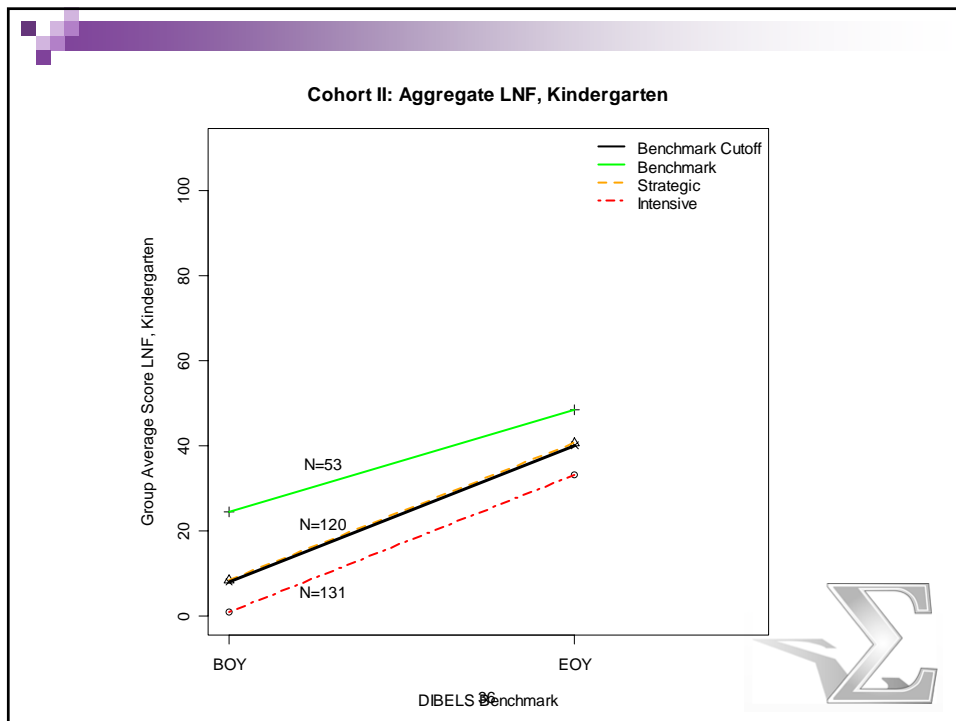
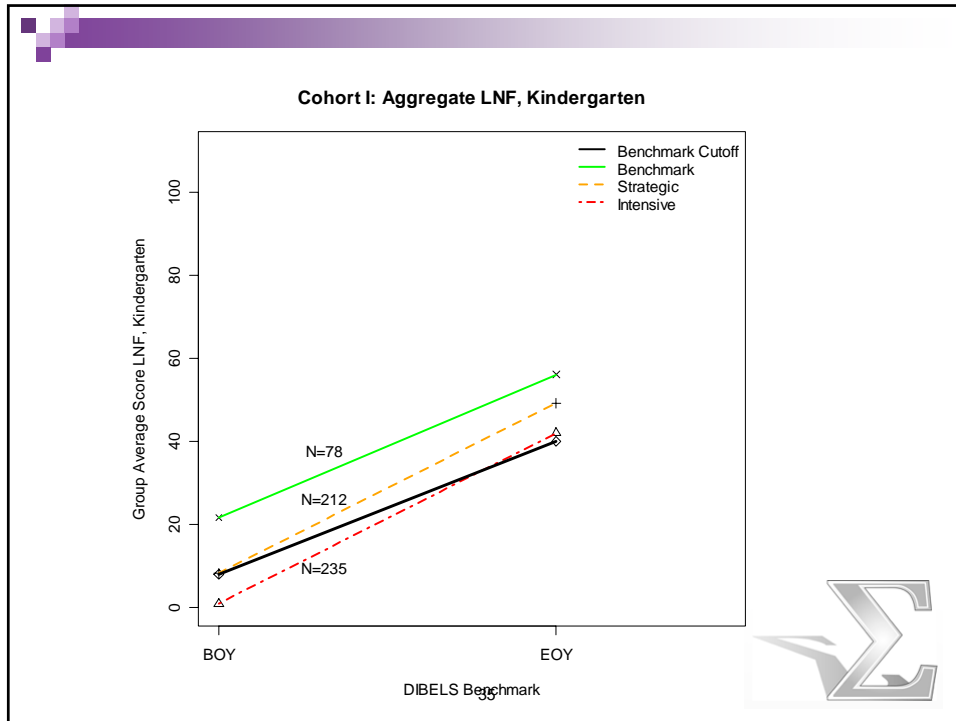


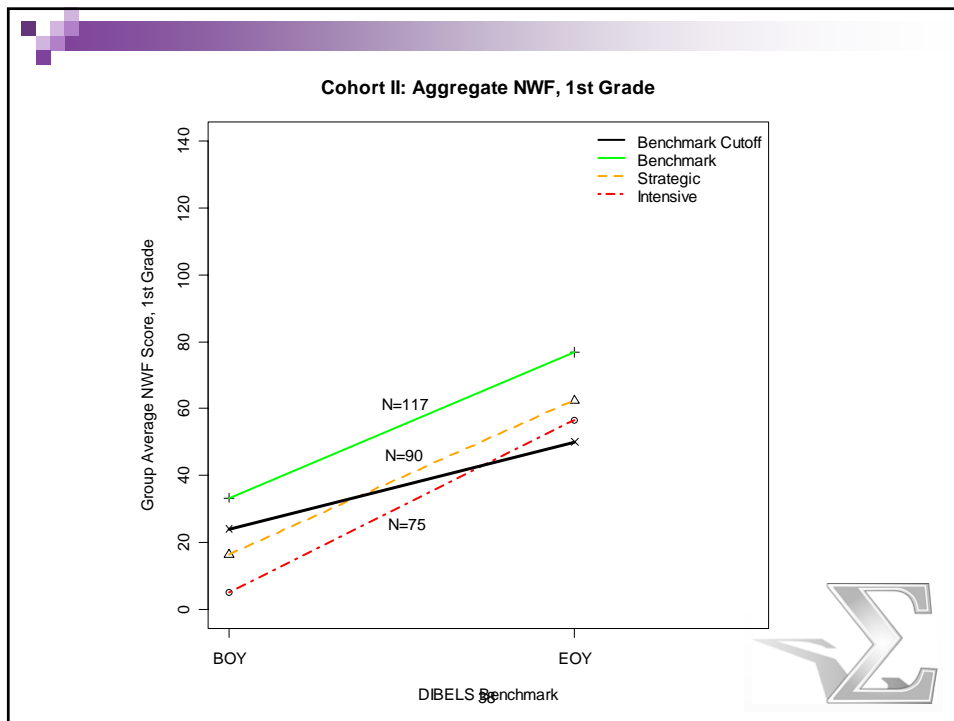
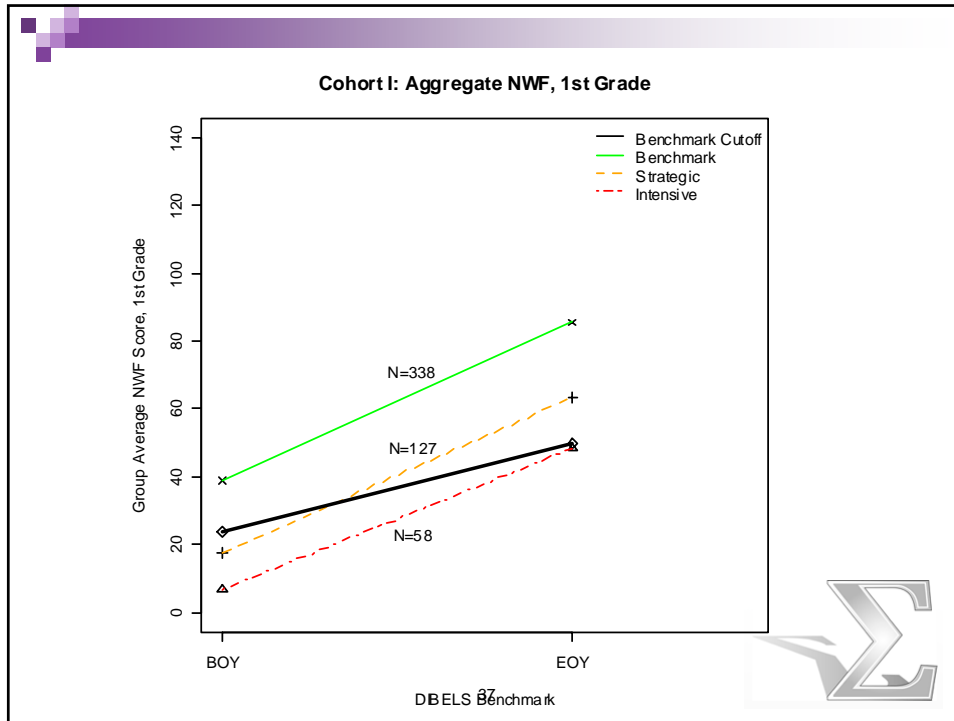


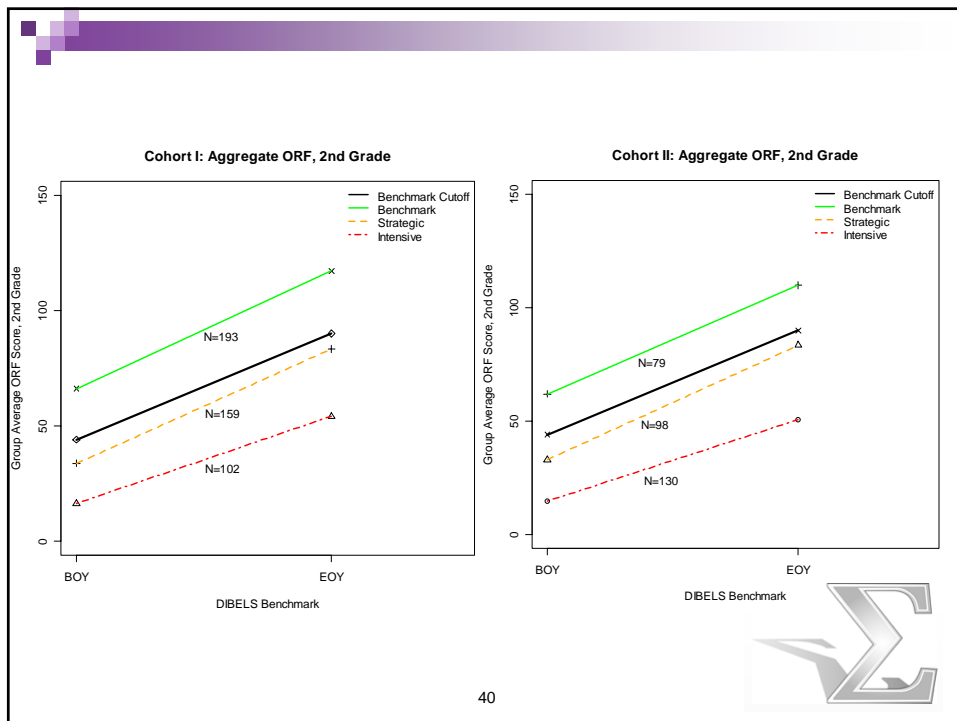
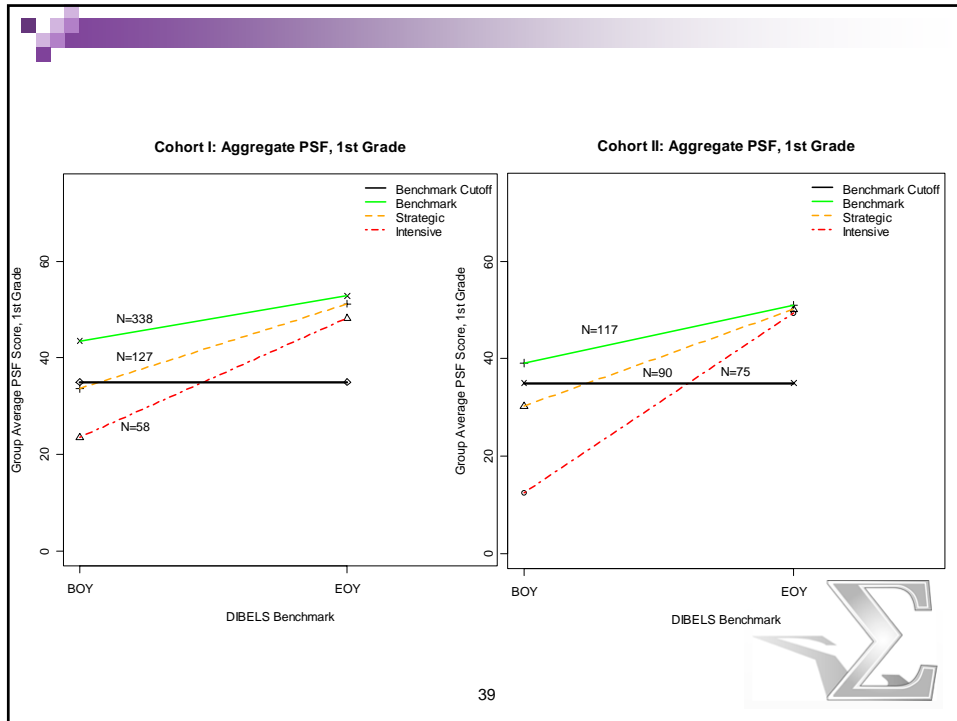


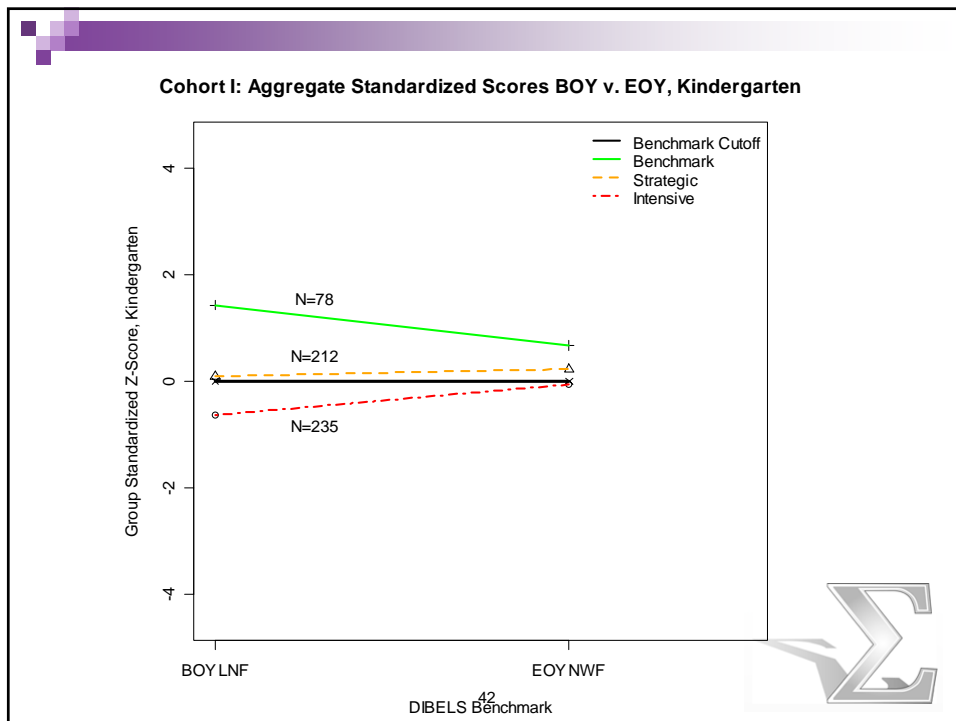
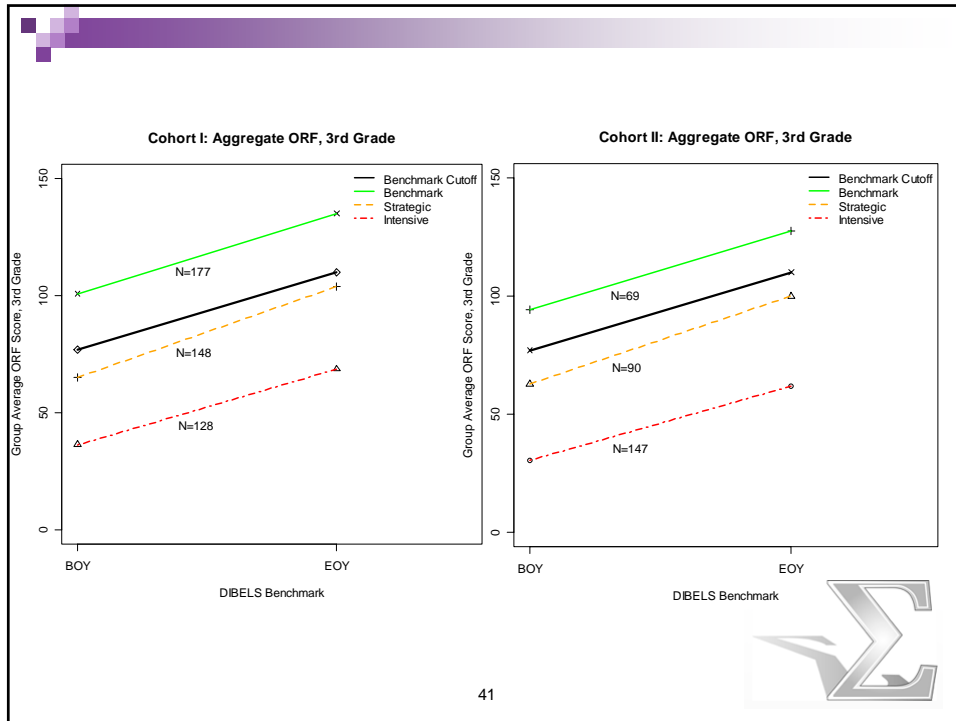
Is the intervention strong enough to  
'catch up' those who begin the year  
INTENSIVE?"

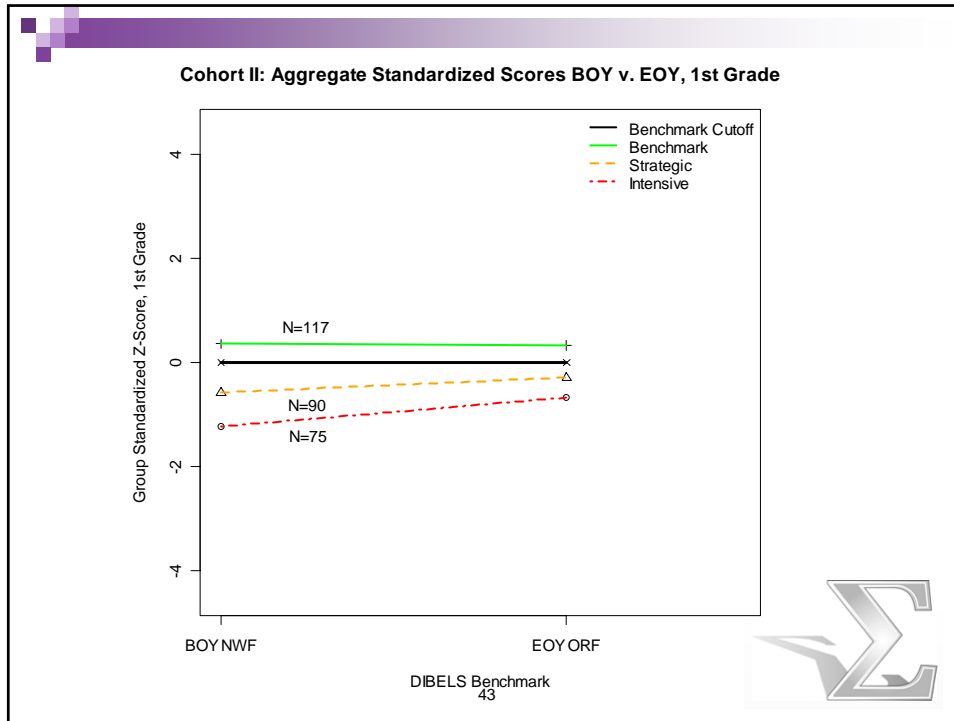












**DIBELS and SAT-10 Alignment**

DIBELS Instructional Recommendation: <b>BENCHMARK</b>	<b>PROFICIENT</b> on SAT-10 (40 <sup>th</sup> percentile)
1-3 Average	<b>79%</b>
1 <sup>st</sup>	<b>86%</b>
2 <sup>nd</sup>	<b>80%</b>
3 <sup>rd</sup>	<b>67%</b>

## DIBELS and SAT-10 Alignment

DIBELS Instructional Recommendation: INTENSIVE	NOT PROFICIENT on SAT-10
1-3 Average	92%
1 <sup>st</sup>	89%
2 <sup>nd</sup>	92%
3 <sup>rd</sup>	95%

45



## Comprehension Measures

- Administered because of concern about CL's push for reading with speed (EOY2006)
  - Students were clearly not reading with understanding
- Determine relationship between fluency and comprehension
- Provide evidence (beyond SAT-10) for adding a comprehension measure after progress monitoring with ORF
- Provide evidence for easy-to-administer measures with higher reliability and "face validity" than DIBELS retell

46



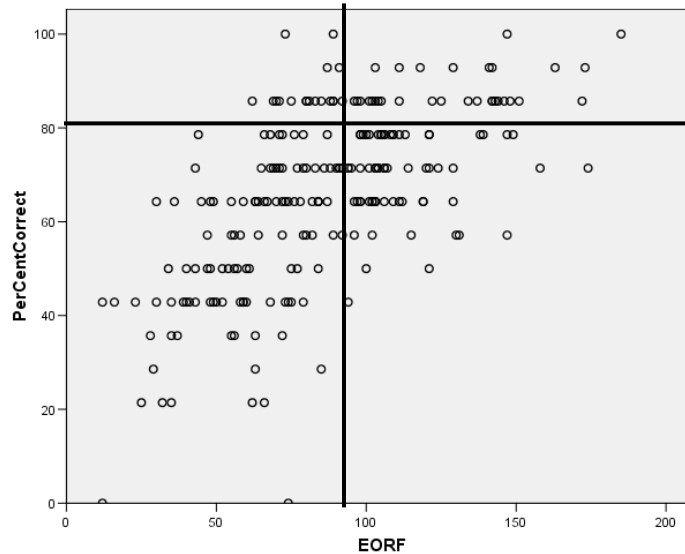
## Comprehension Measures

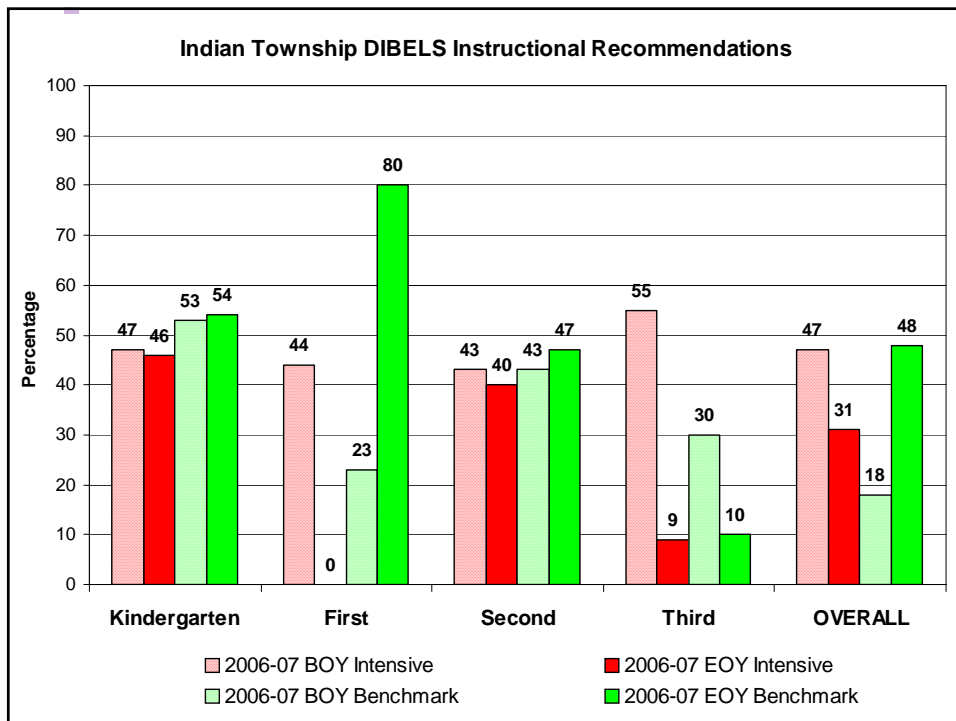
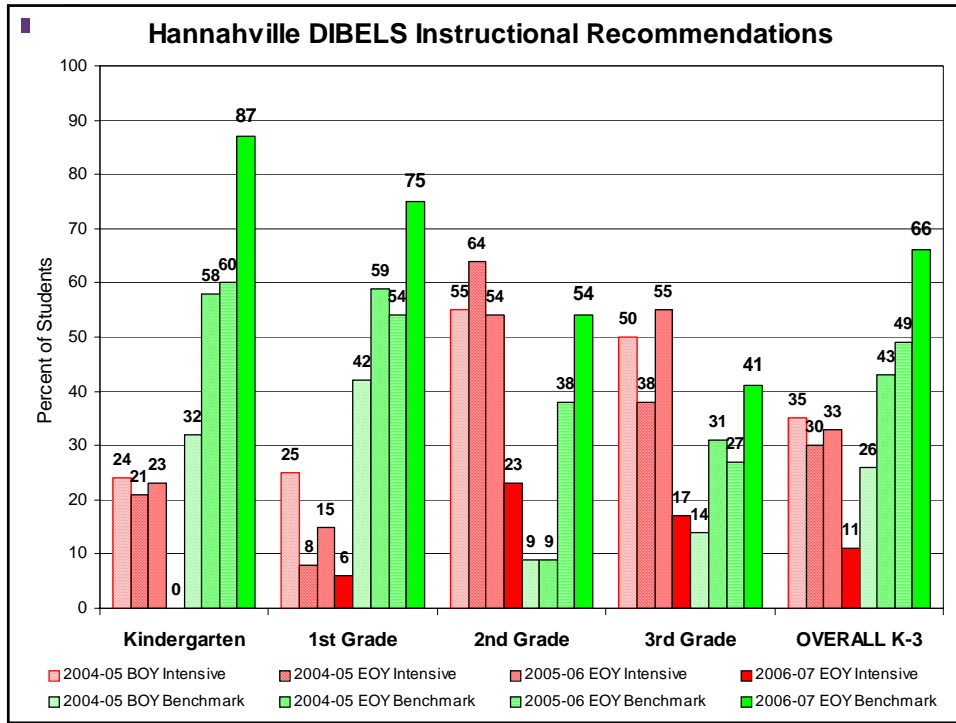
- Measure reliabilities
  - Coefficient alphas range from 0.31 (retell) to 0.65 (cloze)
- Retell: Lowest correlations with DIBELS
  - Pearson-r ranged from 0.11 in 1<sup>st</sup> grade to 0.37 in 3<sup>rd</sup> grade
- Cloze
  - Pearson-r ranged from 0.42 in 1<sup>st</sup> grade to 0.67 in 2<sup>nd</sup> grade
- Questions (recall, infer)
  - Pearson-r ranged from 0.42 to 0.63 (1<sup>st</sup> and 3<sup>rd</sup>)



47

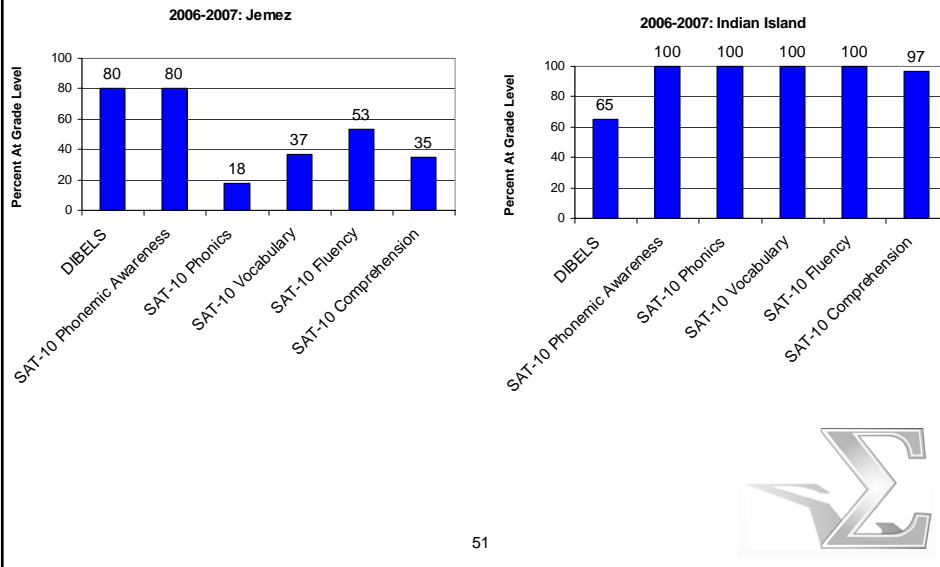
## 2<sup>nd</sup> Grade DIBELS and Cloze (EOY 2007)







## Discrepancy between DIBELS and SAT-10



## Discrepancies...

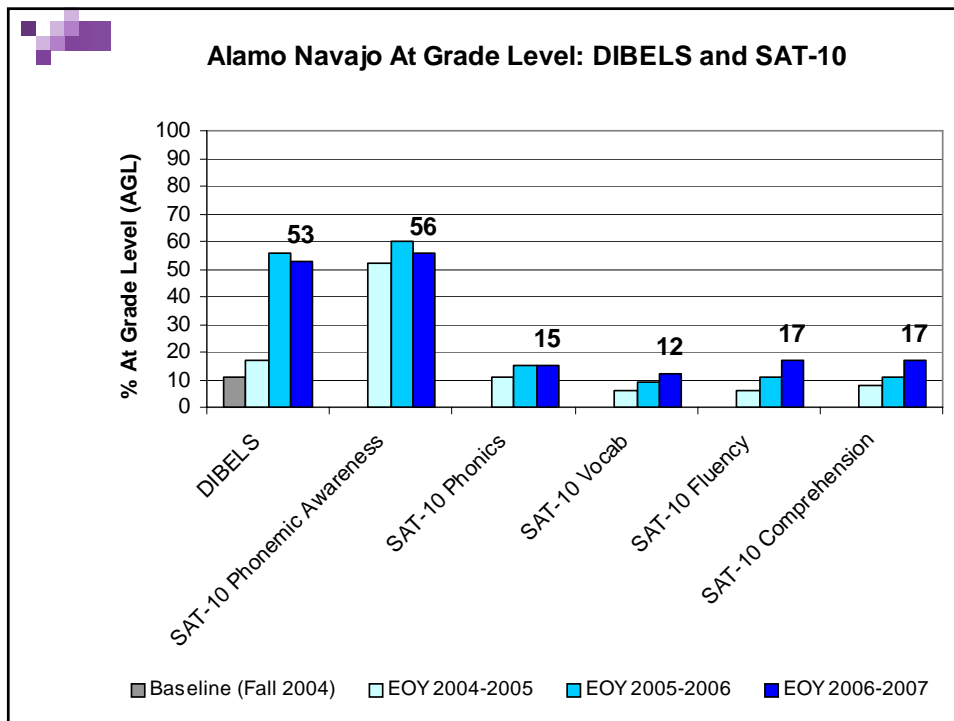
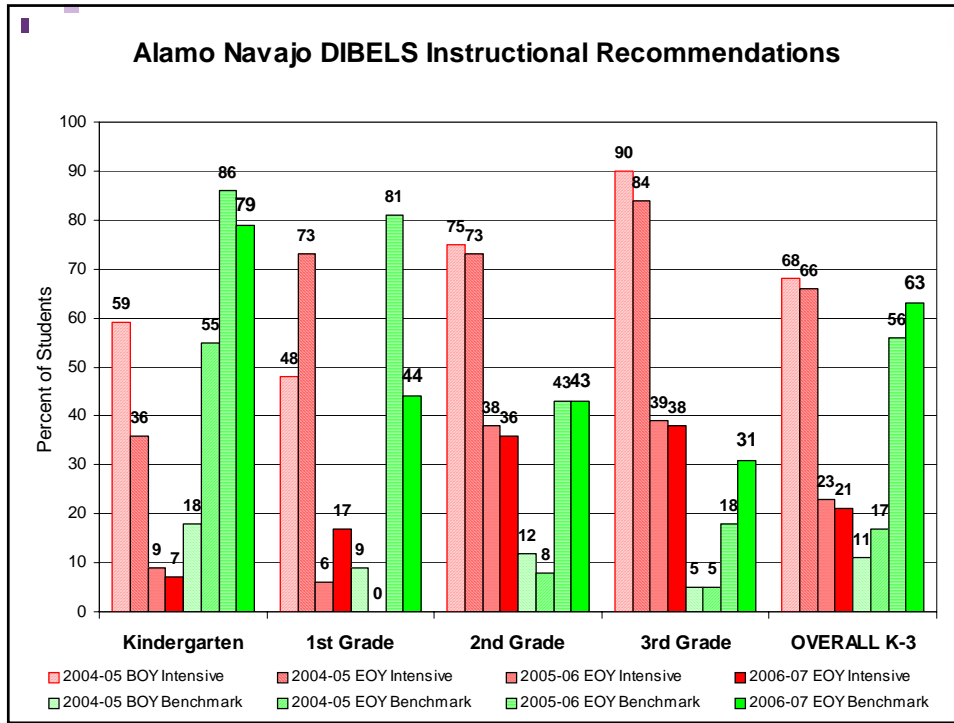
- Between reading coach and staff reporting
  - Length of scheduled reading block
  - Additional intervention provided to struggling readers

52



## Alamo Navajo

- New reading coach who was previously mentor teacher
- High staff turnover
  - Many 3<sup>rd</sup> grade teachers in 2006-2007
  - Coach's sister is RF teacher
- Two teachers with increased use of effective teaching practices
- Principal has very high expectations of staff, which causes frustration among RF staff



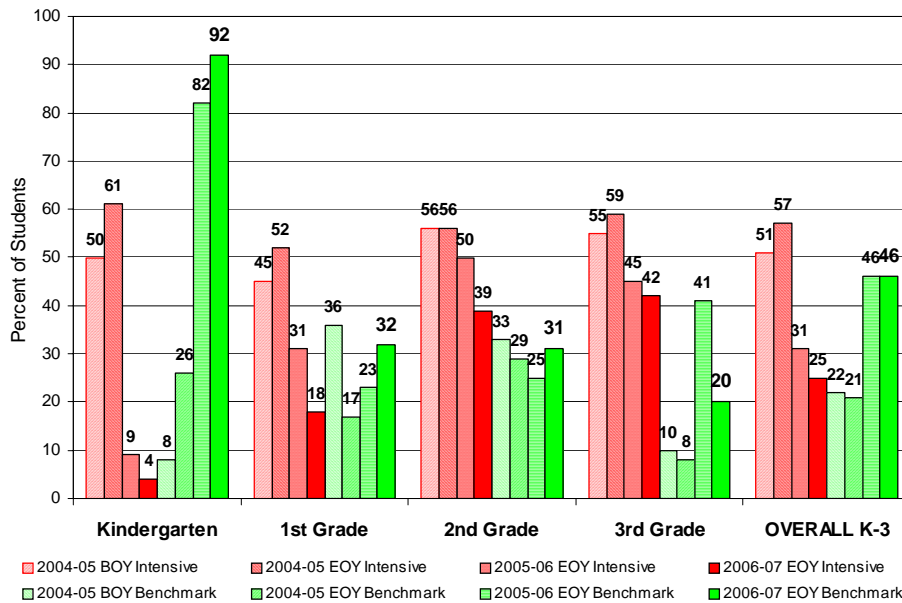
## Aneth

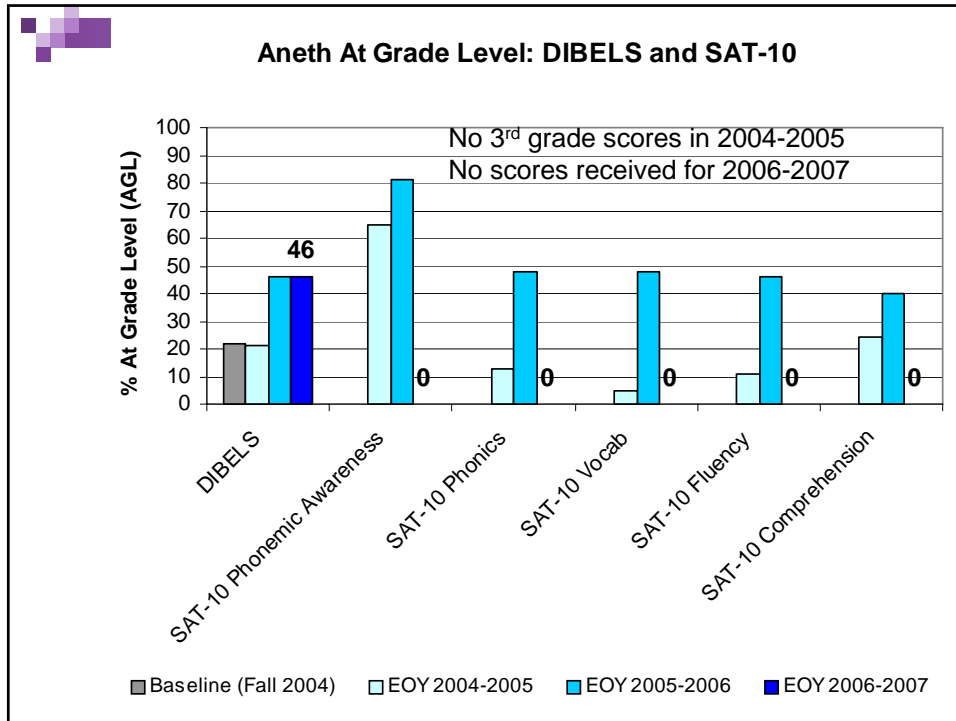
- Principal was 'detailed' mid-year
- Starting January, reading coach also served as administrator
- Staff turnover
- Use of WRRFTAC templates



57

**Aneth DIBELS Instructional Recommendations**





## Black Mesa

- Small geographically isolated school
- Several administrators
- Use of effective teaching practices needs improvement



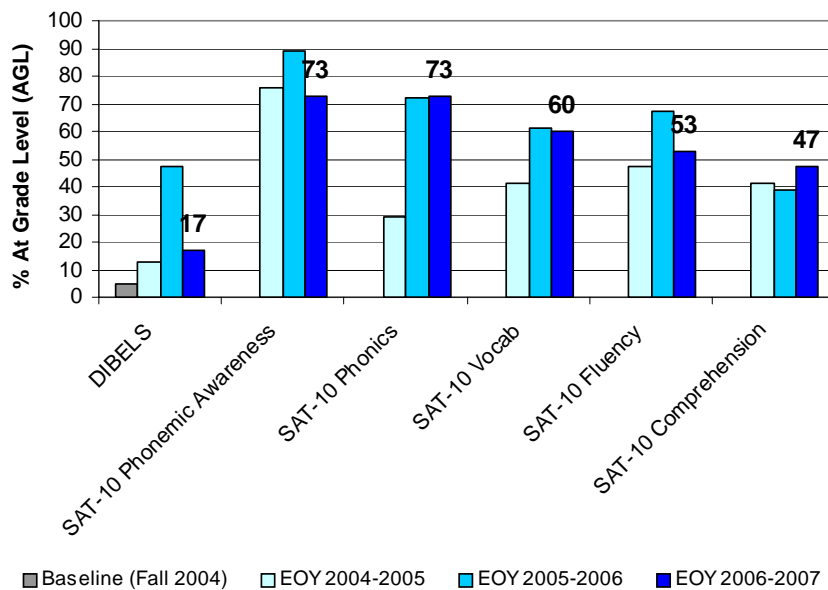
## Black Mesa

- No DIBELS graphs
- Numbers of students are too small
- Numbers of students per grade range from 0-6 across years



61

**Black Mesa At Grade Level: DIBELS and SAT-10**

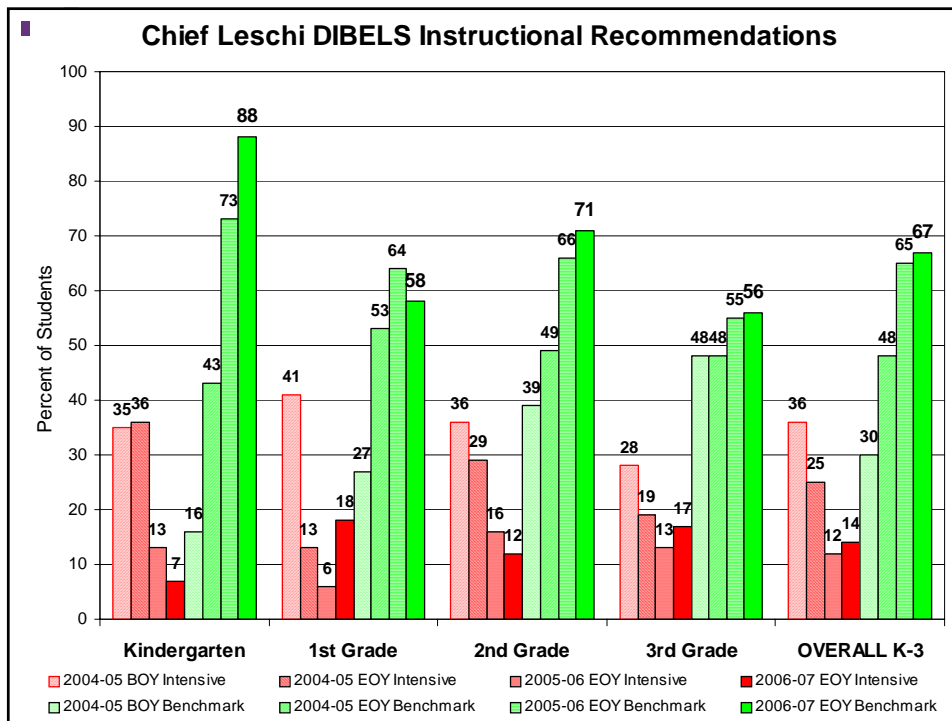


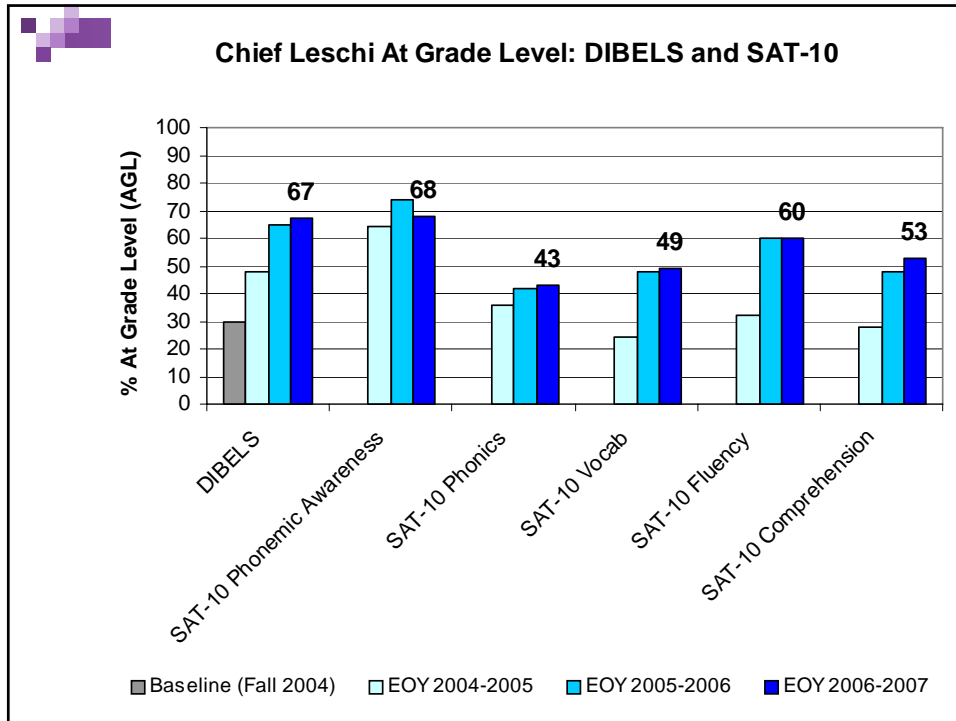
## Chief Leschi

- A model school
- Likely to have a sustainable program
- Reading First models used in K-12, math
- High use of effective teaching practices
- School is clearly “data-driven”



63



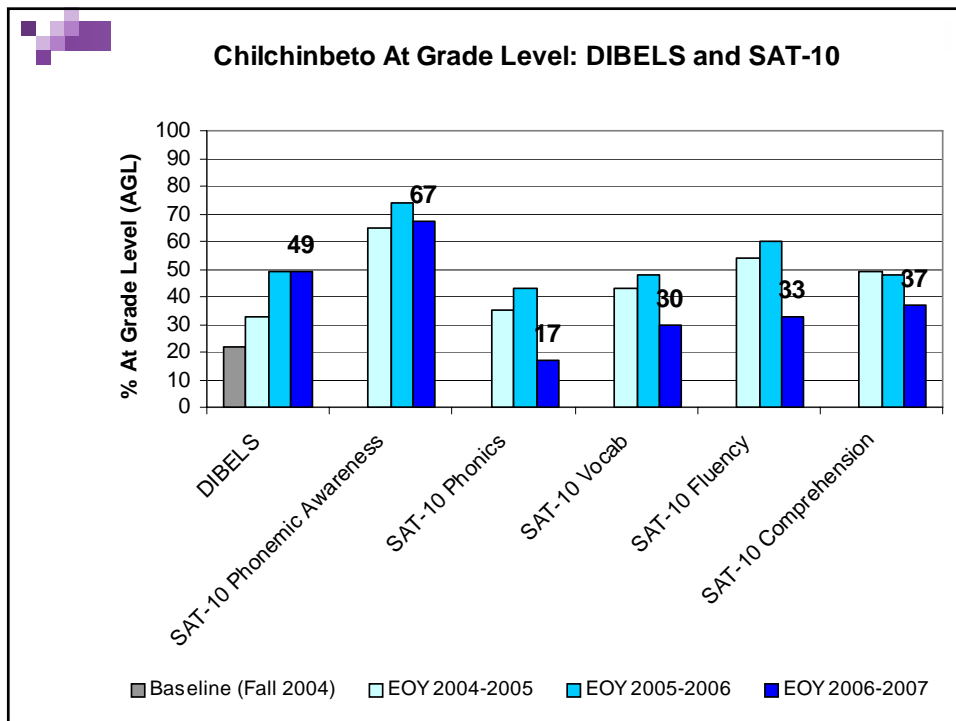
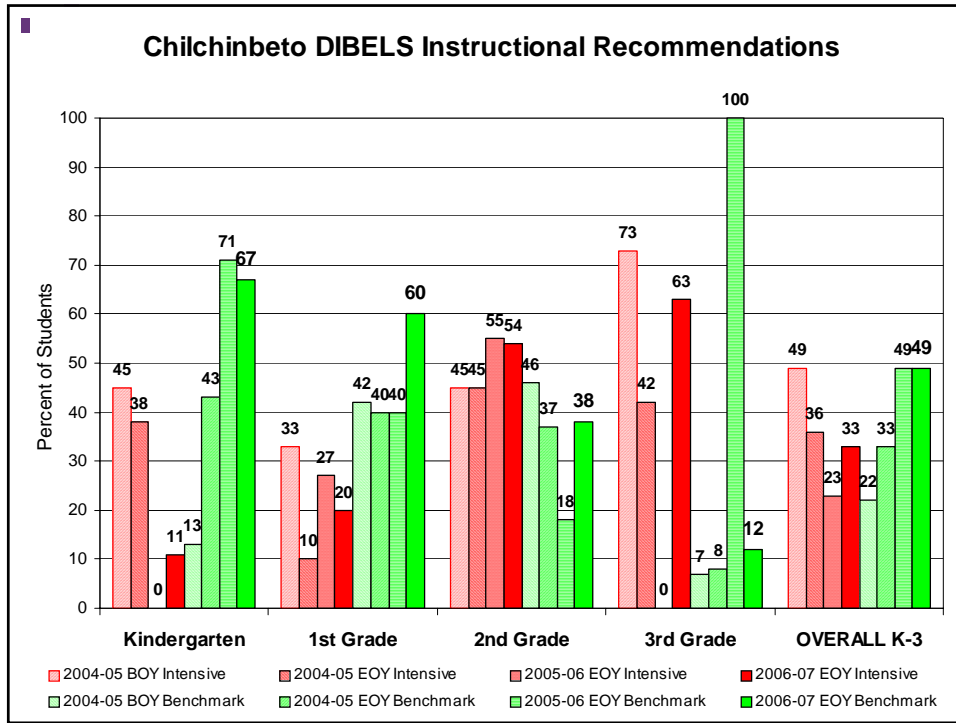


## Chilchinbeto

- Admin and reading coach changes
- Staff turnover





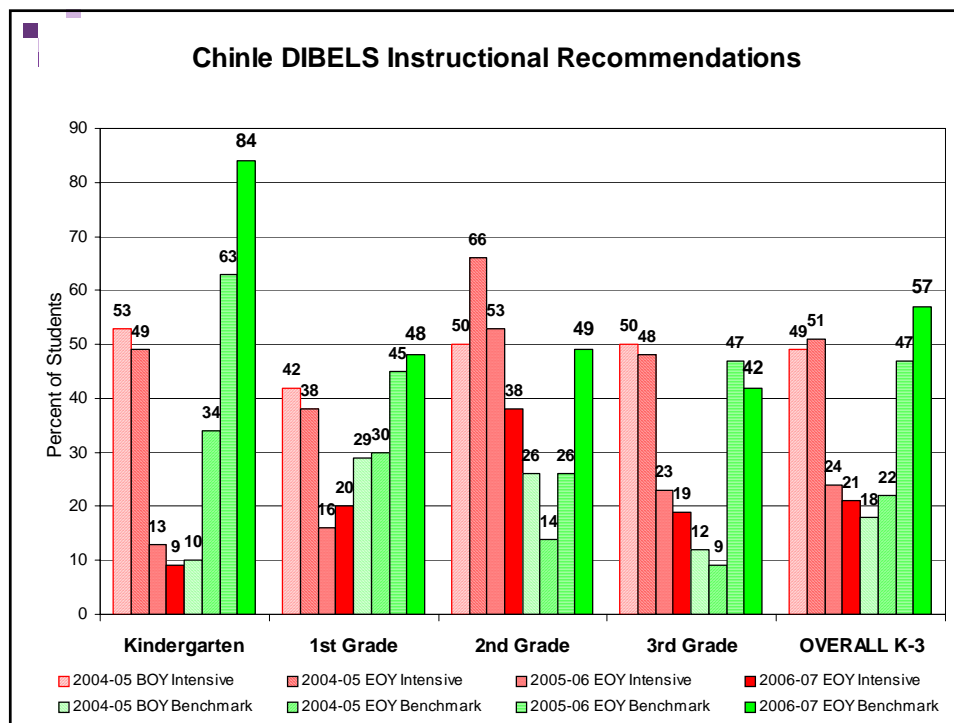


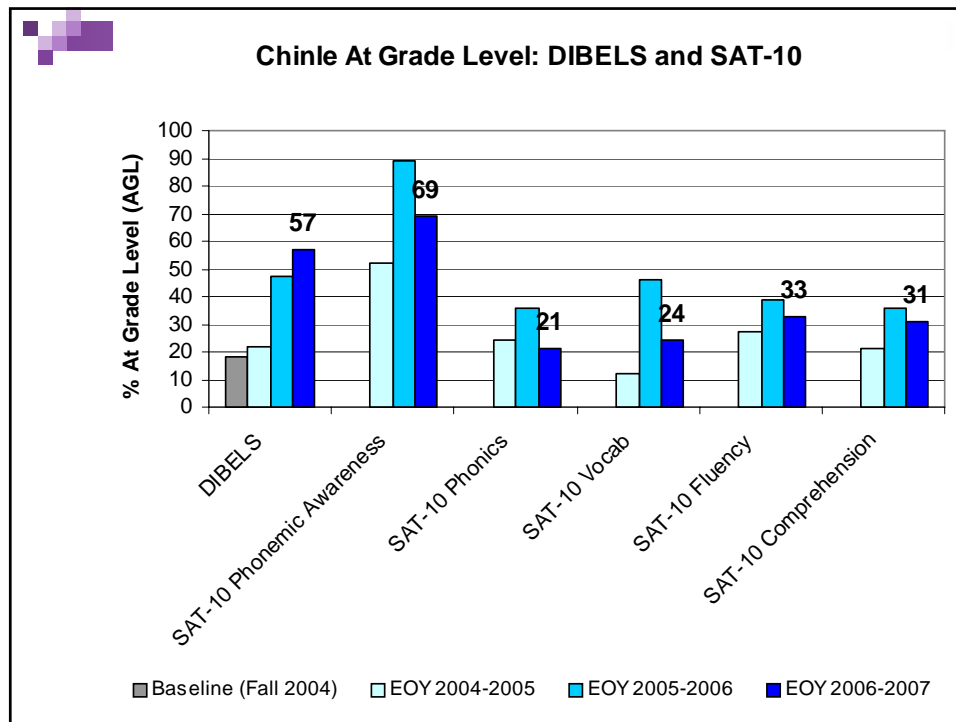
## Chinle

- Admin isolated
- Reading coach changes, now with younger, less experienced reading coach
- Mentor teacher gone (2005-2006), some routines, practices still visible but fading
- Fidelity of implementation decreased
- Reading coach(es), staff requested technical assistance



69

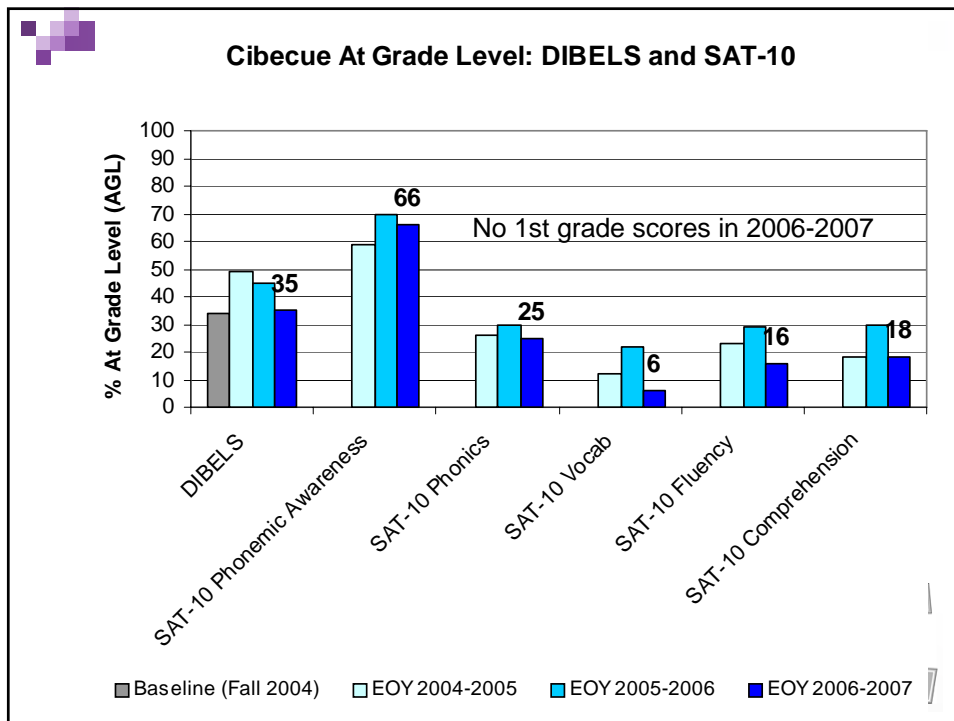
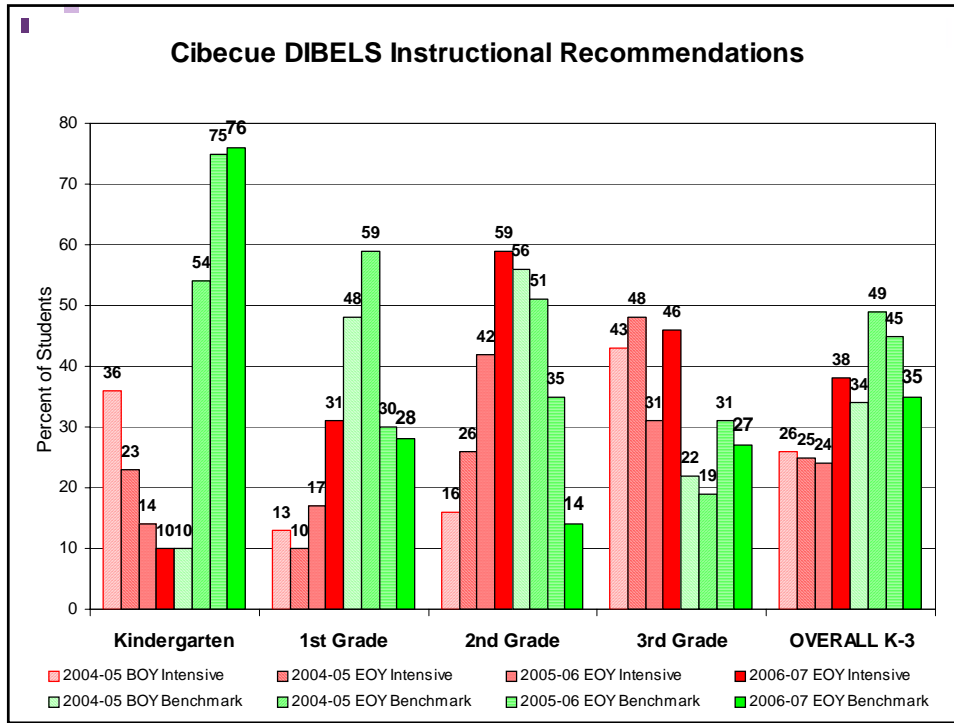




## Cibecue

- Change in reading program  
(Reading Mastery to Open Court)
  - Open Court has not demonstrated effectiveness in BIE schools
- High staff absences
- Need for more effective teaching practices, more effective use of time



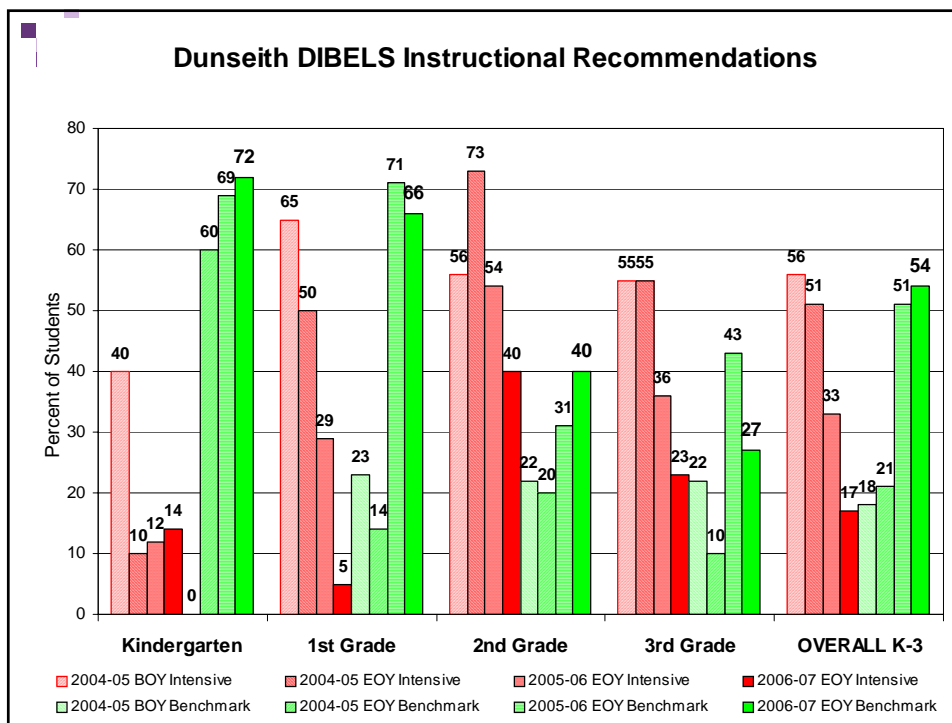


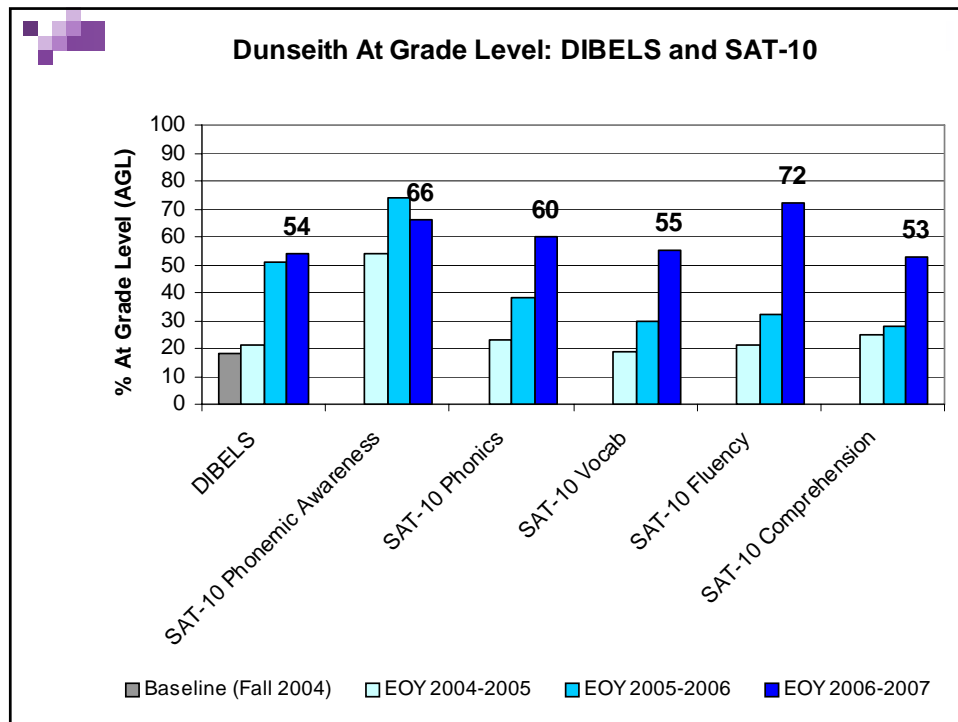
## Dunseith

- New reading coach last year
- Increased use of effective teaching practices



75

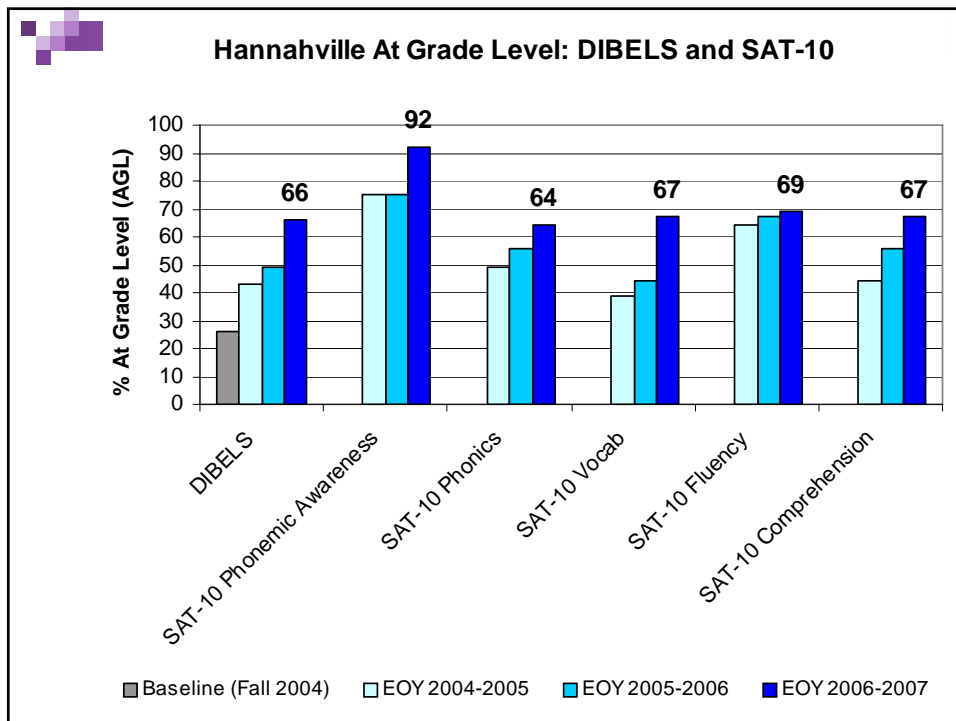
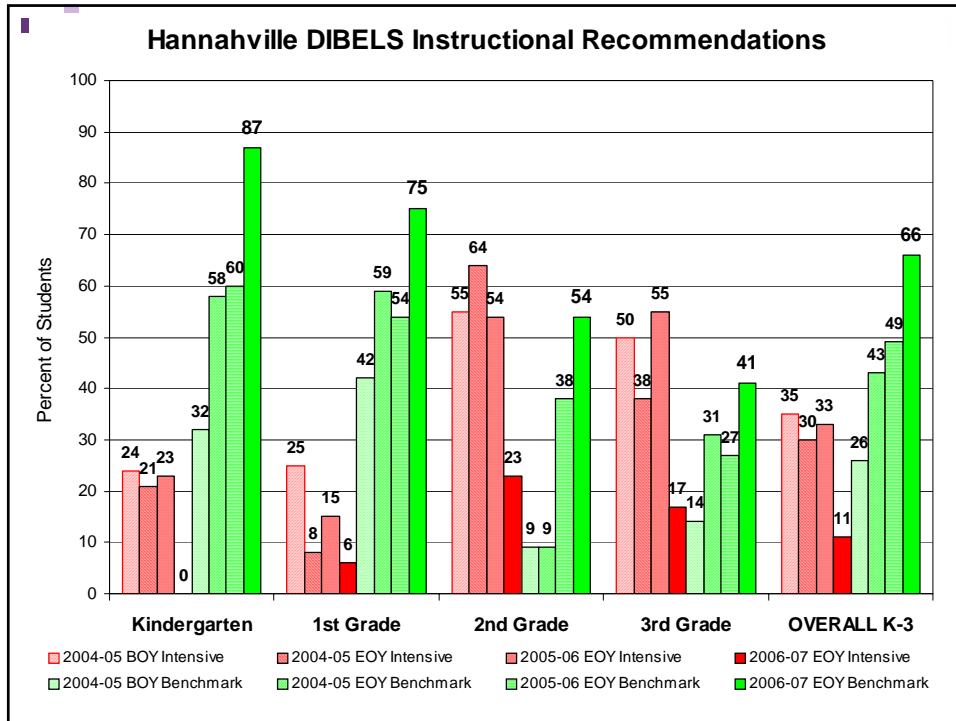




## Hannahville

- WRRFTAC templates made obvious difference
- FACE coordinator now admin for RF program
- Mostly stable staff, admin



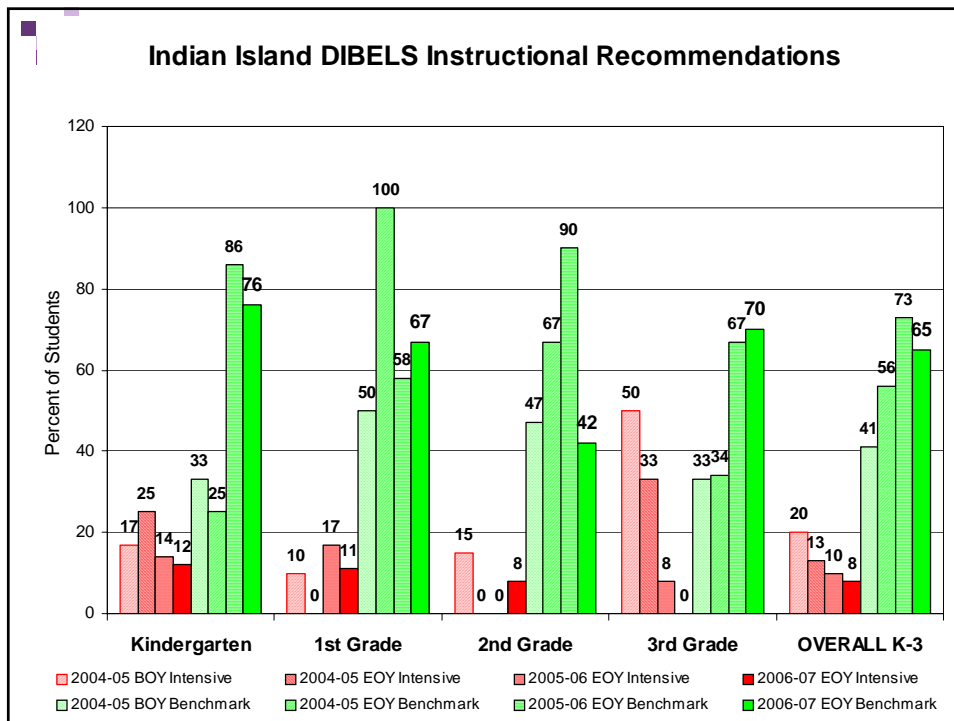


## Indian Island

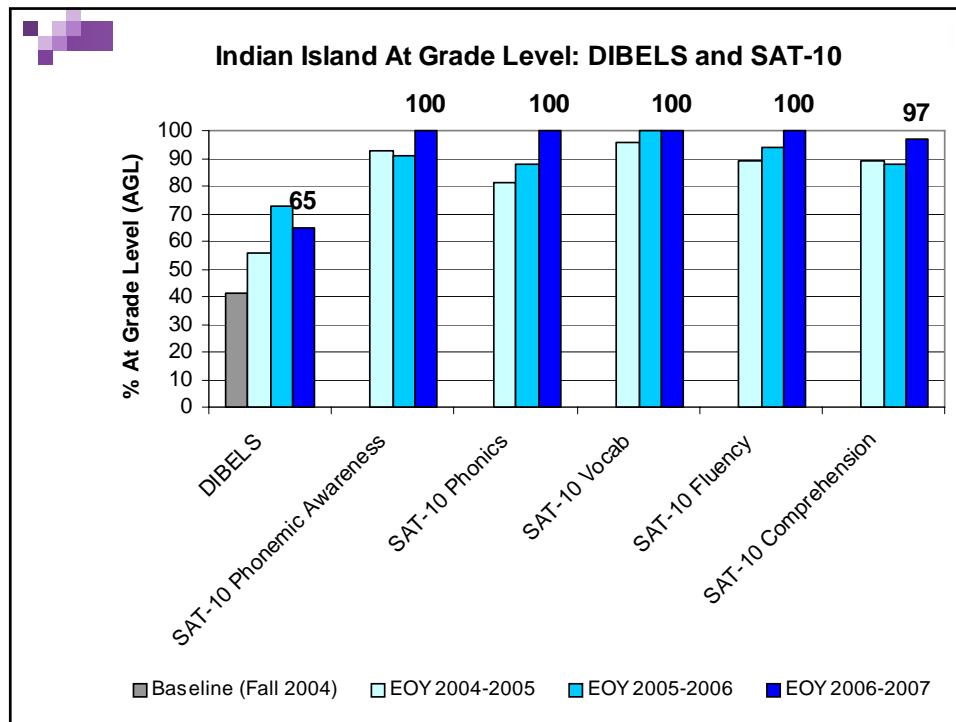
- Mostly stable staff, admin
- High performing school



81





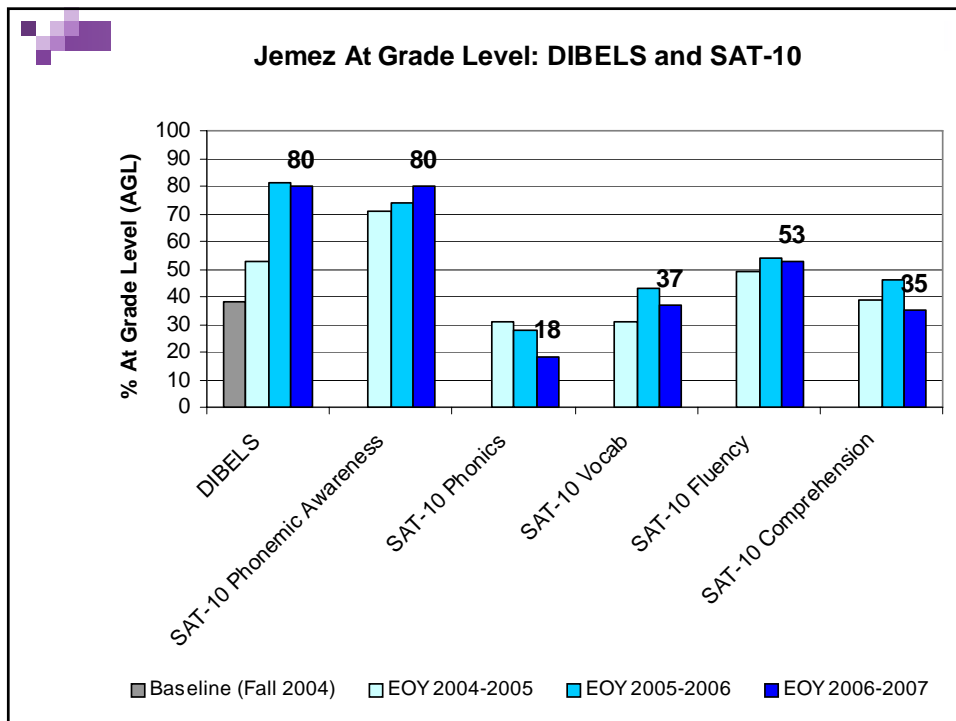
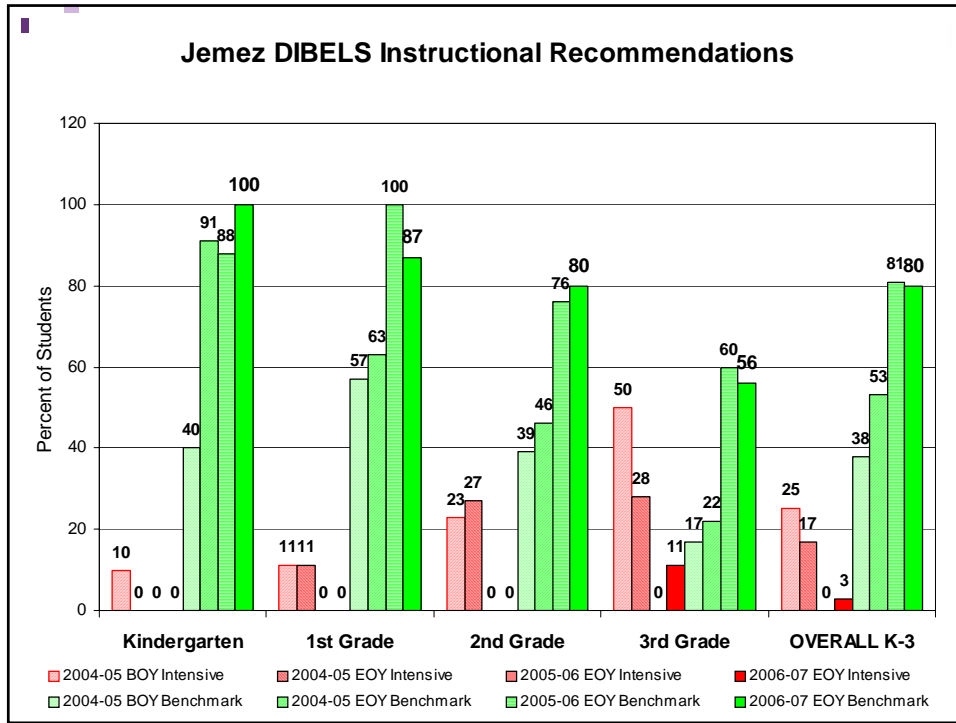


## Jemez

- Success for All established prior to Reading First
- Reading Coach previously used Success for All
- All teachers involved in Success for All selection
- All teachers from community and speak Towa
- Paraprofessionals are degreed (but haven't taken NM test)
- Principal is highly engaged, supportive, and visible in classrooms
- Furthest house from school just over 2 miles
- Most students return daily homework
- Parents and community involved in school activities

84





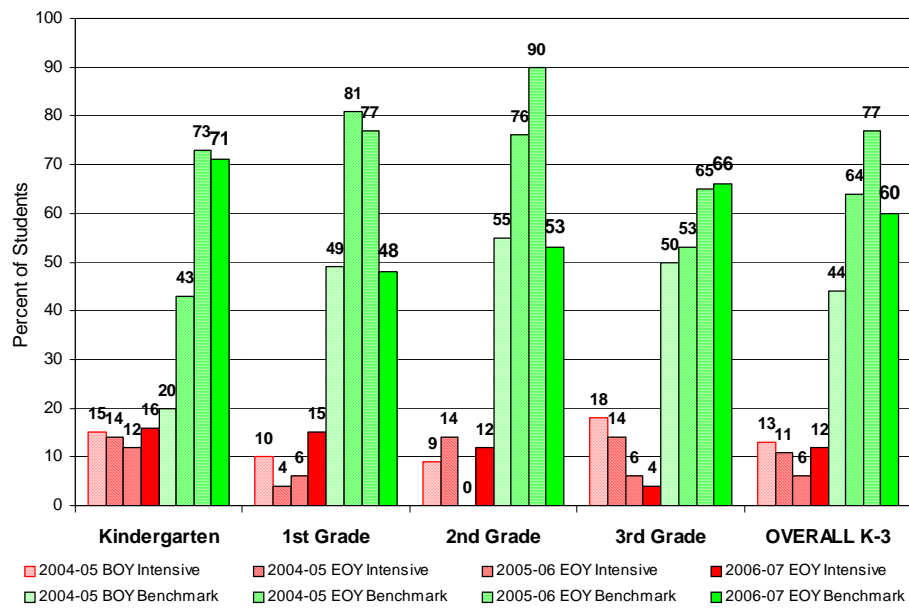
## Lac Courte Oreilles

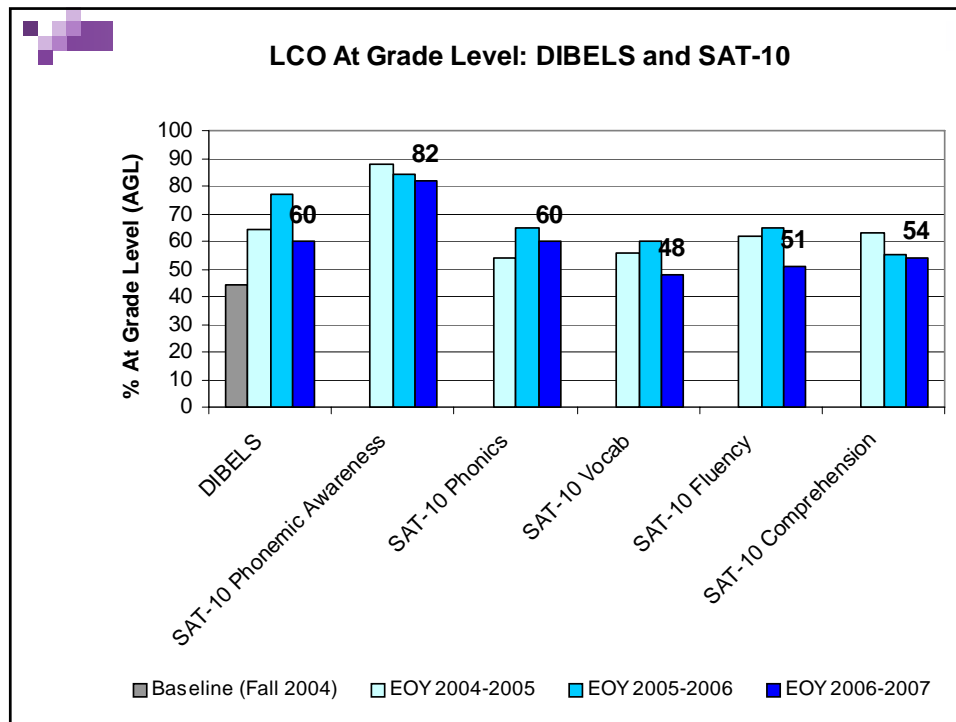
- Two resistant teachers left
- Conflict still apparent among staff
- Little staff collaboration
- Poor relationship between reading coach and several staff
- Fidelity of implementation still problematic



87

### LCO DIBELS Instructional Recommendations

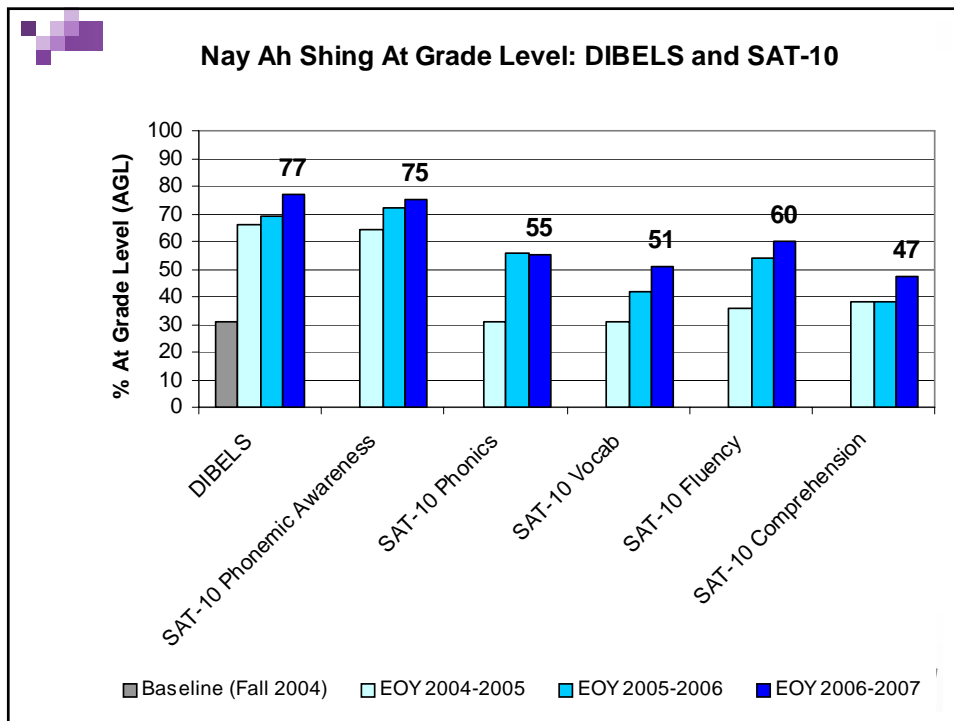
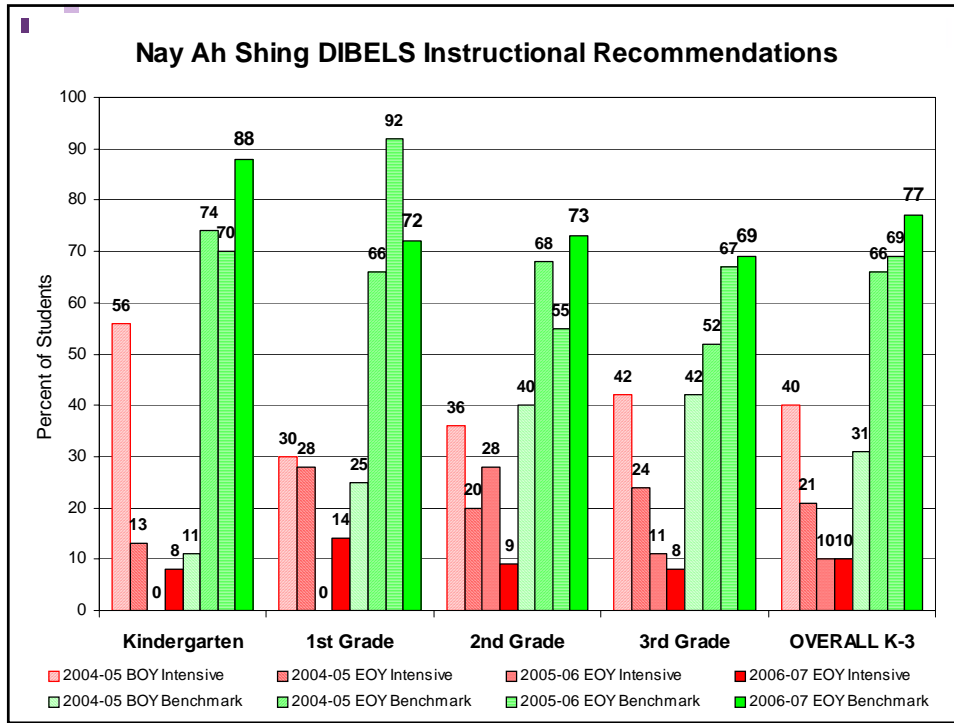




## Nay Ah Shing

- A model school
- Likely to have a sustainable program
- Reading coach is completing administrative degree and gone 4 days per week
- Teacher serving as reading coach
- High fidelity of implementation
- Reading First models used in K-6, math



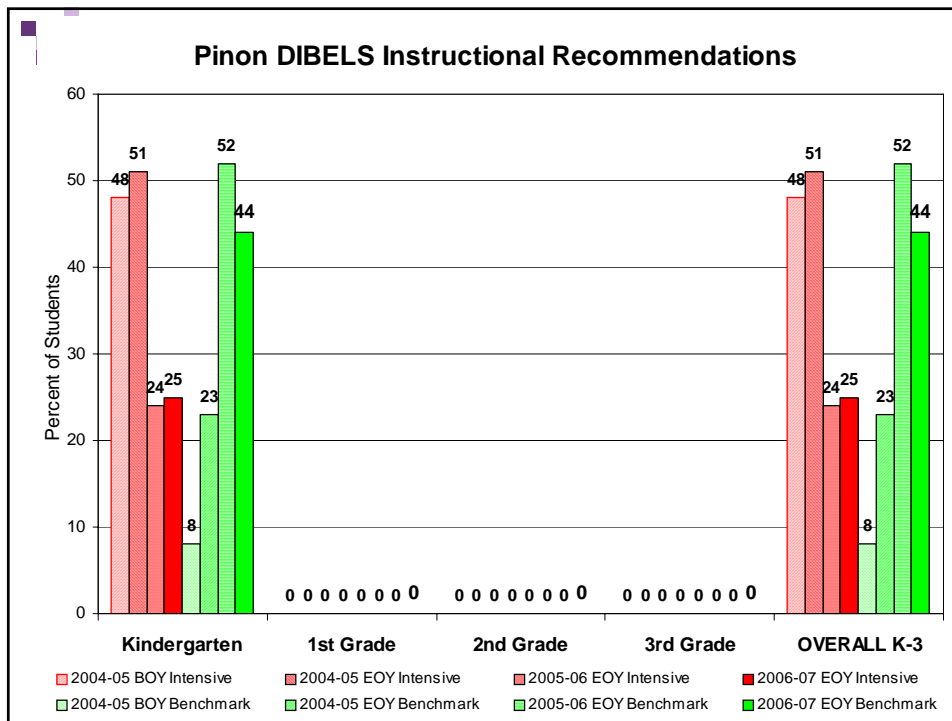


## Pinon

- Kindergarten only
- High turnover in staff
- New reading coach last year
- Has not performed above BIE average for kindergarten



93



## Pinon SAT-10

- Pinon is a kindergarten only school
- SAT-10 is not required for kindergarten

95

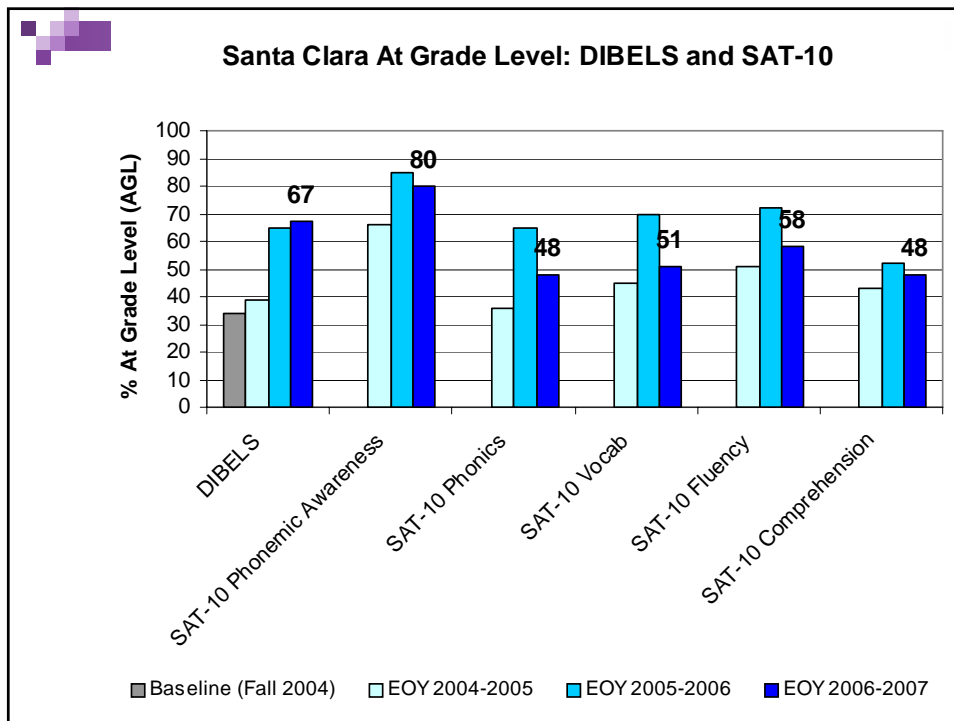
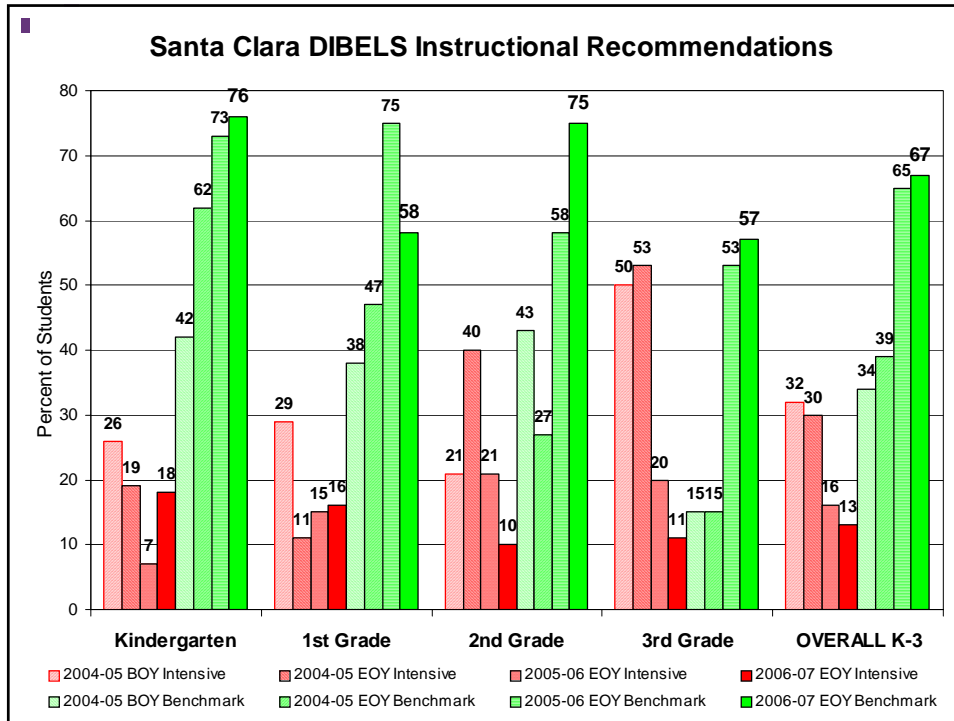


## Santa Clara

- Staff turnover
- High rates of staff crises
- Implementation challenges but improving
- Reading coach requested technical assistance

96





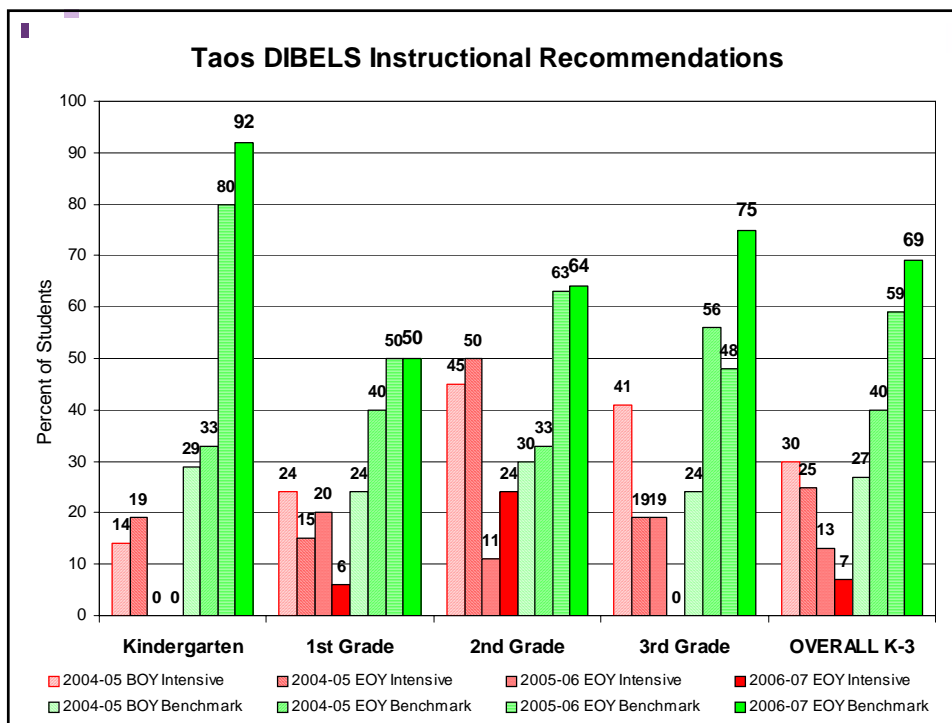


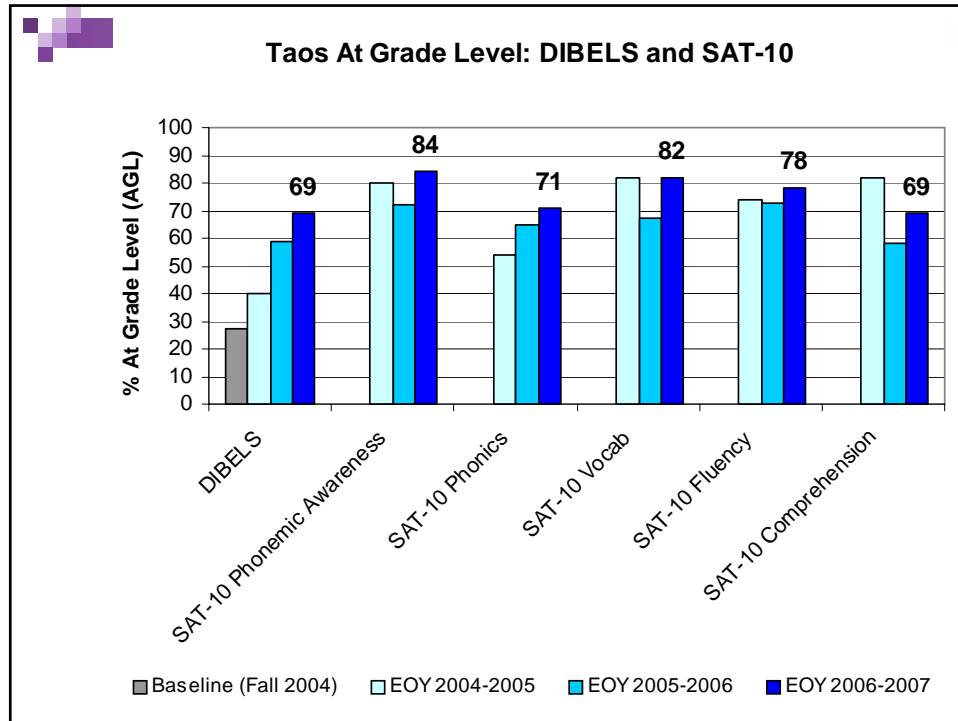
## Taos

- Principal has been on medical leave substantial amounts of time
- Mostly stable staff
- Fidelity of implementation and use of effective teaching practices has improved



99

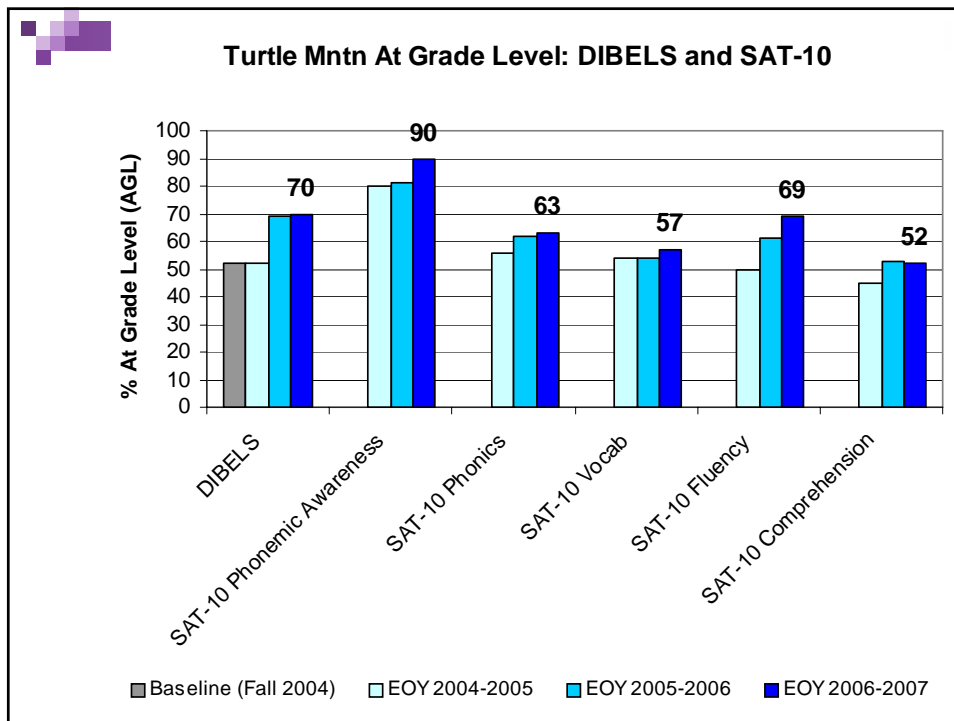
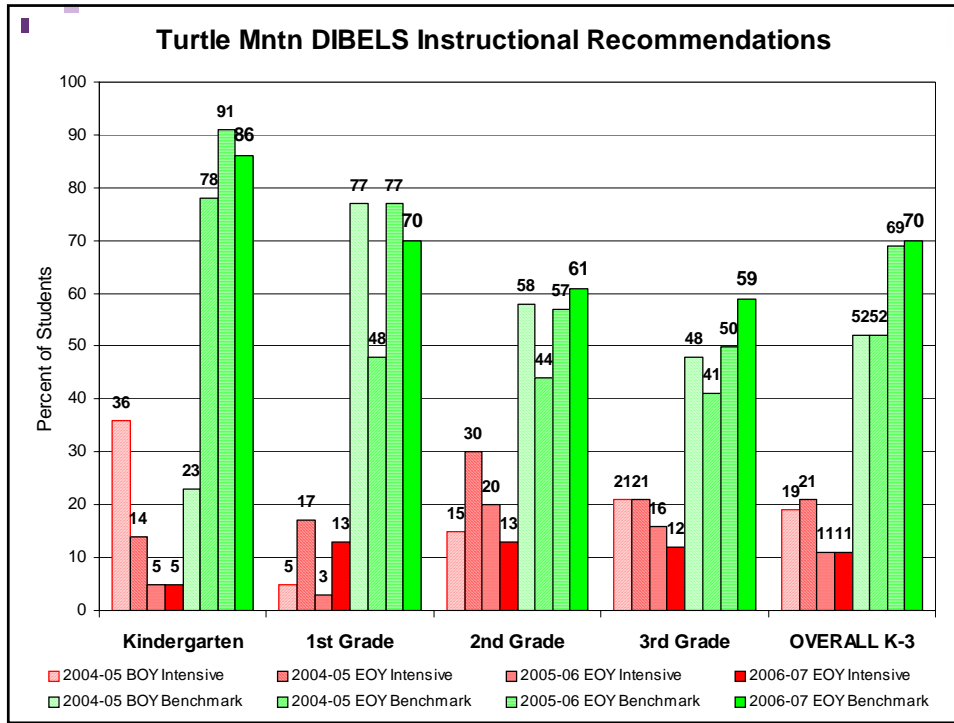




## Turtle Mountain

- Largest school
- Conflicting messages from admin, reading coaches to staff
- Reading coaches differ greatly in skills, working styles
- Questionably sustainable



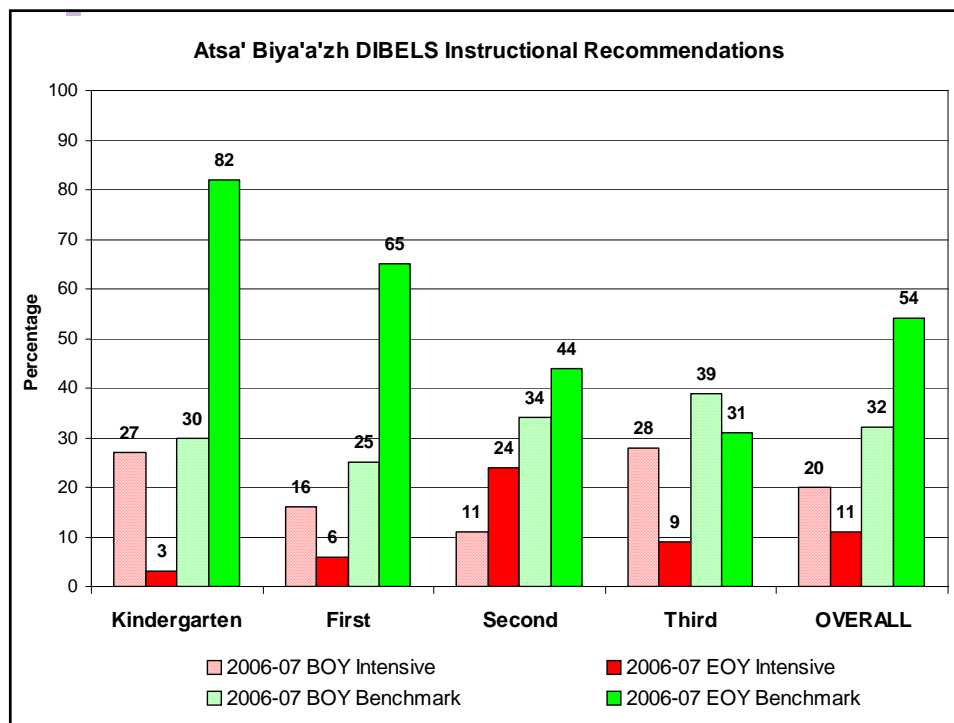


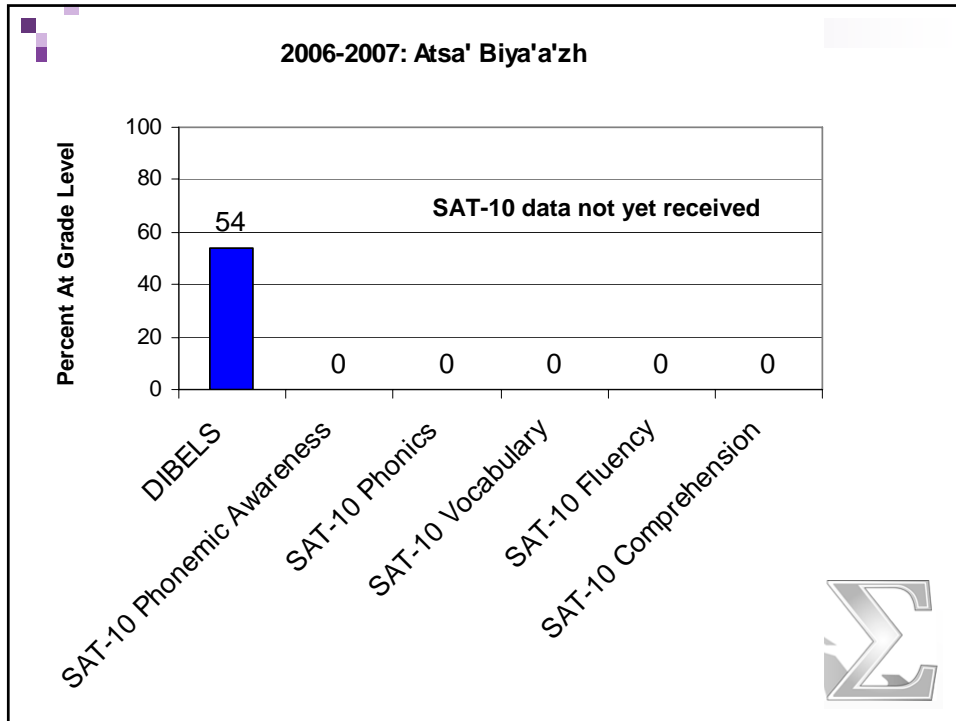
## Atsa' Biya'a'zh

- Highest performing Cohort II school in 2006-2007
- Principal left after start of this school year




105



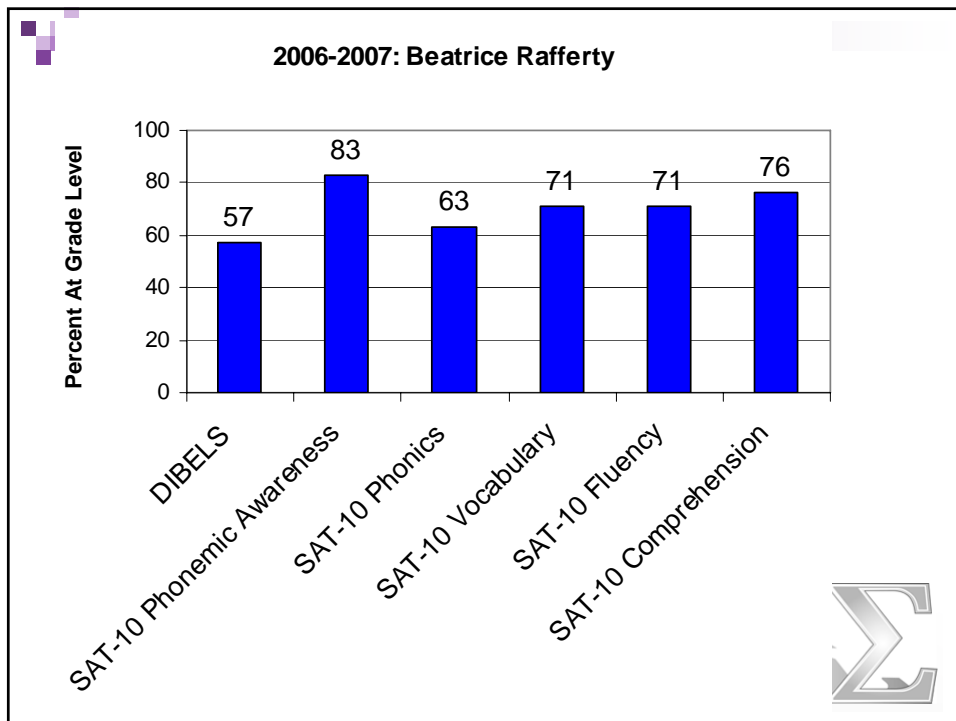
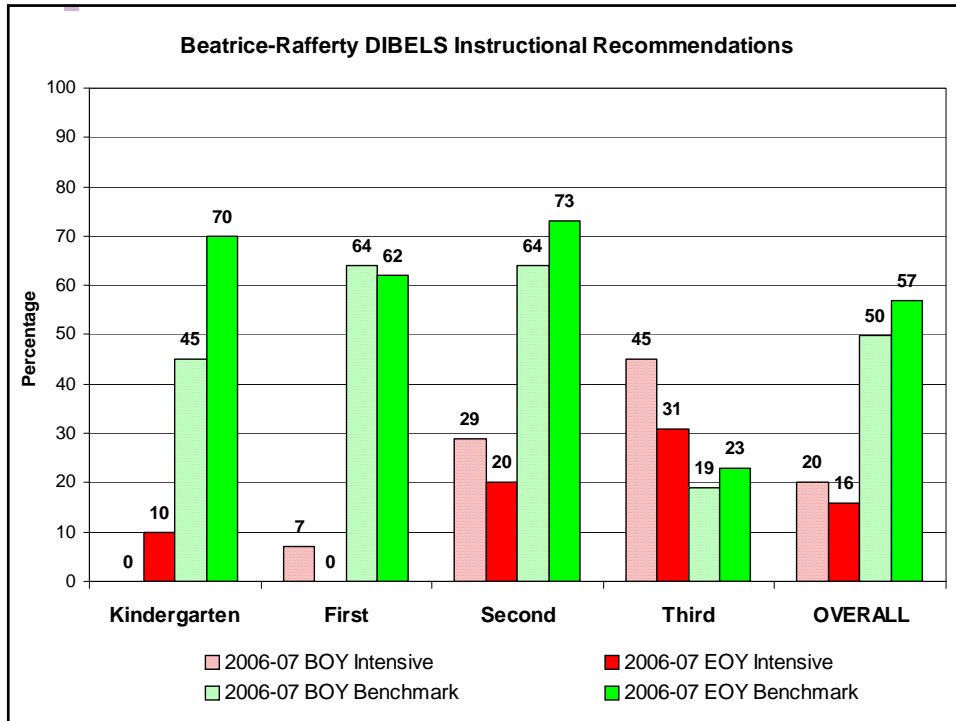


## Beatrice Rafferty

- Stable staff
- Principal very visible in classrooms



108

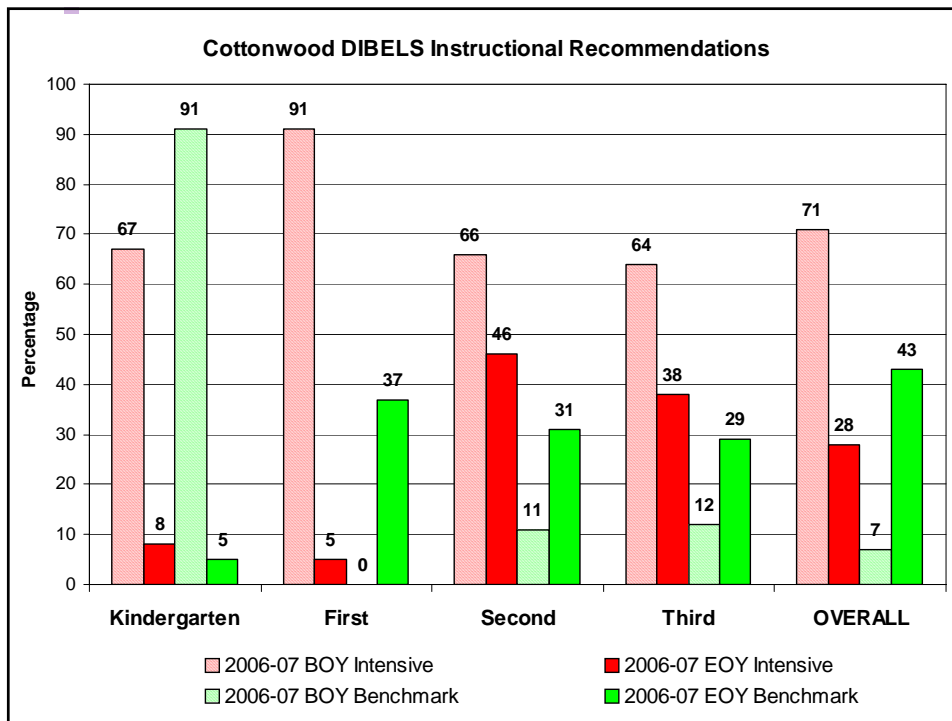


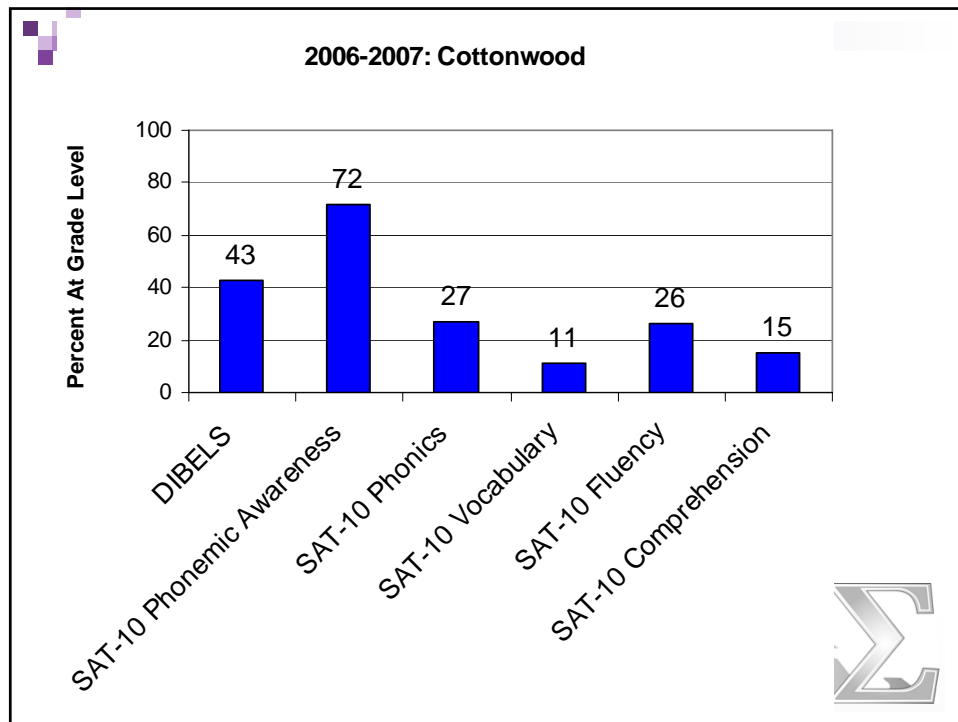
## Cottonwood

- Lowest performing school, BOY 2006-2007
- Reading replaced most subjects
- Ed Tech contracts were ended before school year ended
- Percent progress was substantial



111

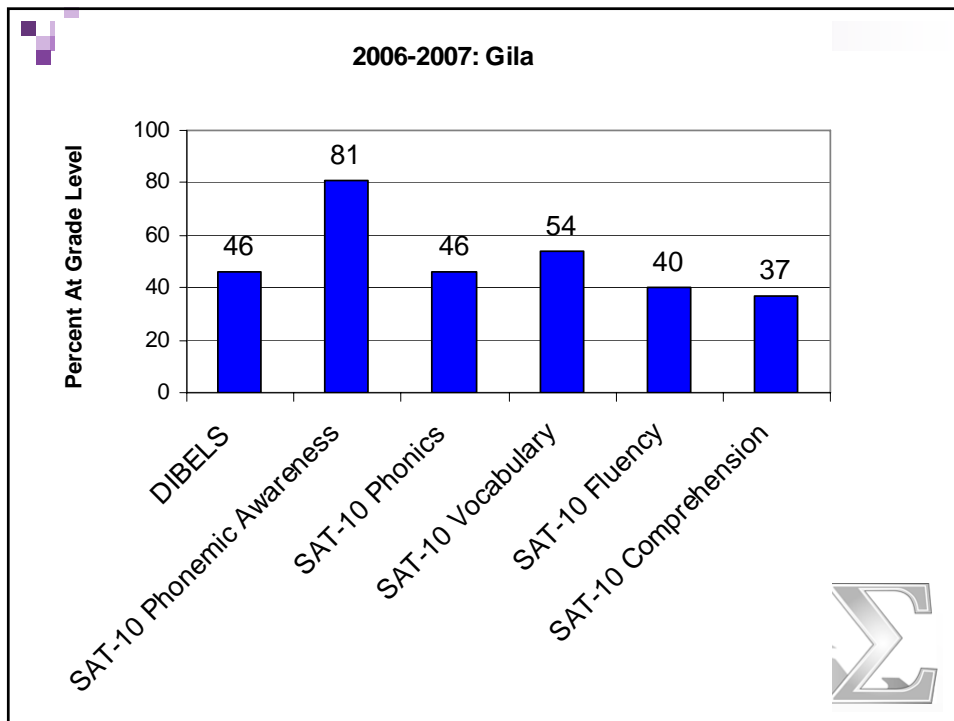
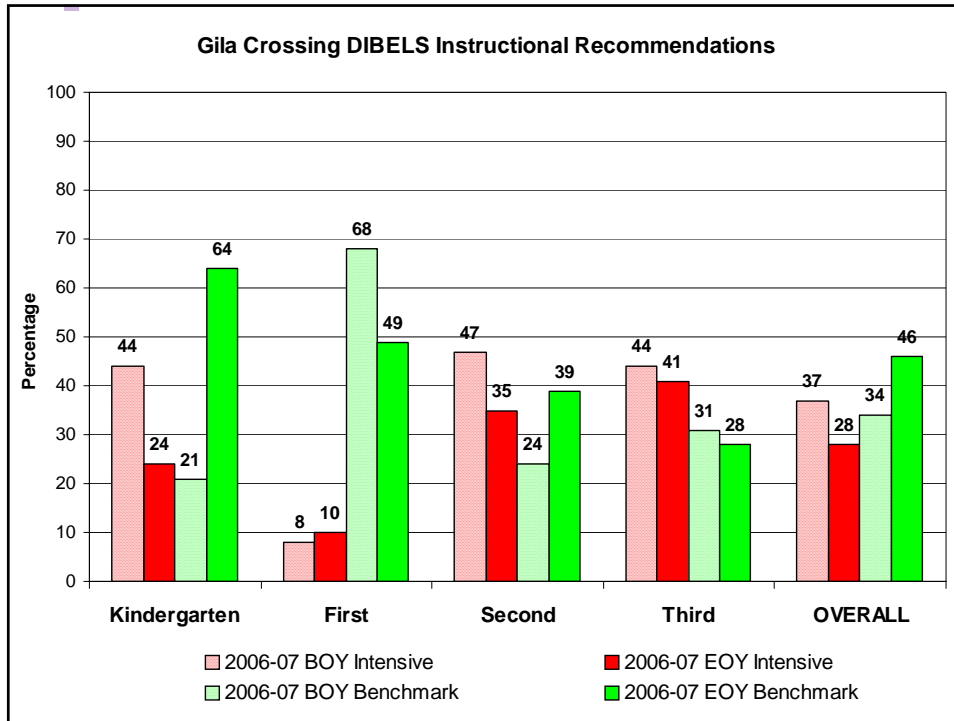




## Gila Crossing

- Fidelity of implementation problematic
  - 3-5 minute template use
- Principal was replaced
- Reading coach not confident



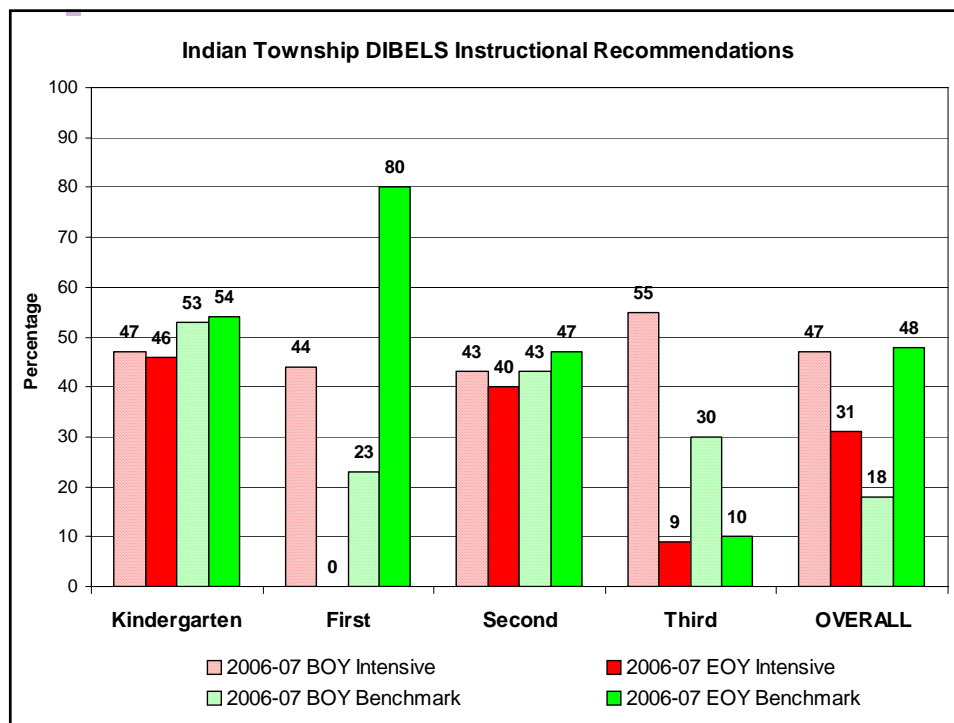


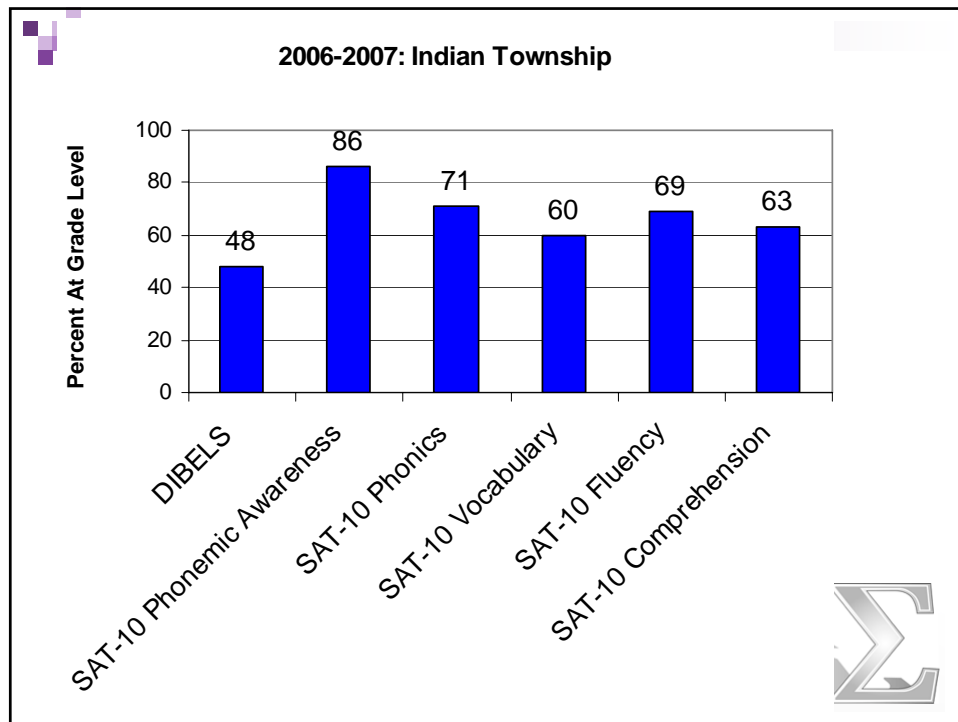
## Indian Township

- Clear need for technical assistance given DIBELS trends
- Need increased use of effective teaching practices
- Coach isolated from classrooms



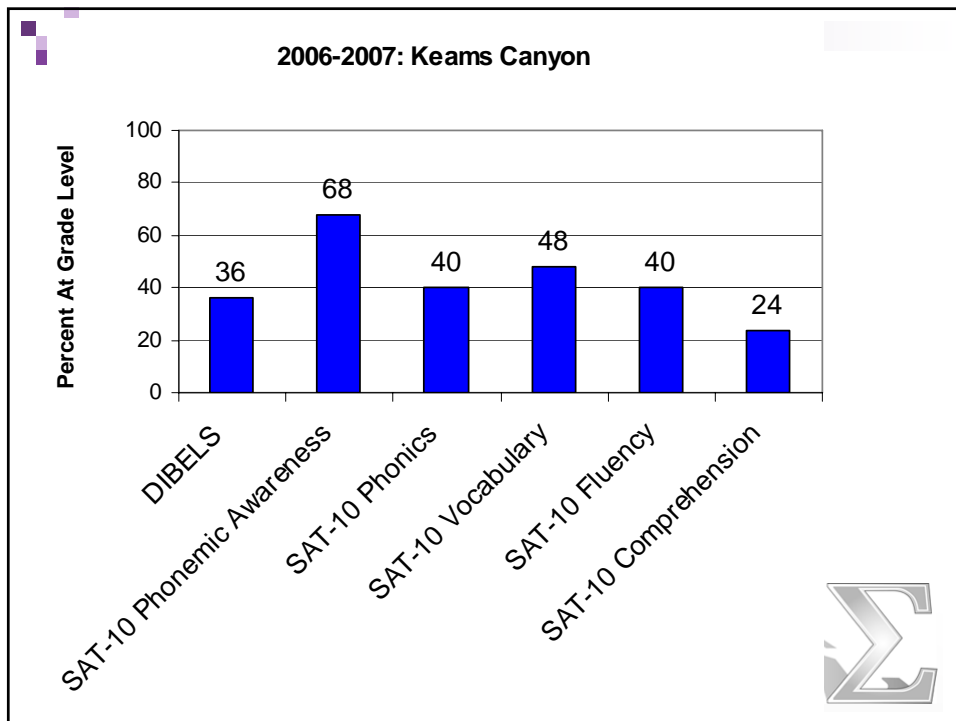
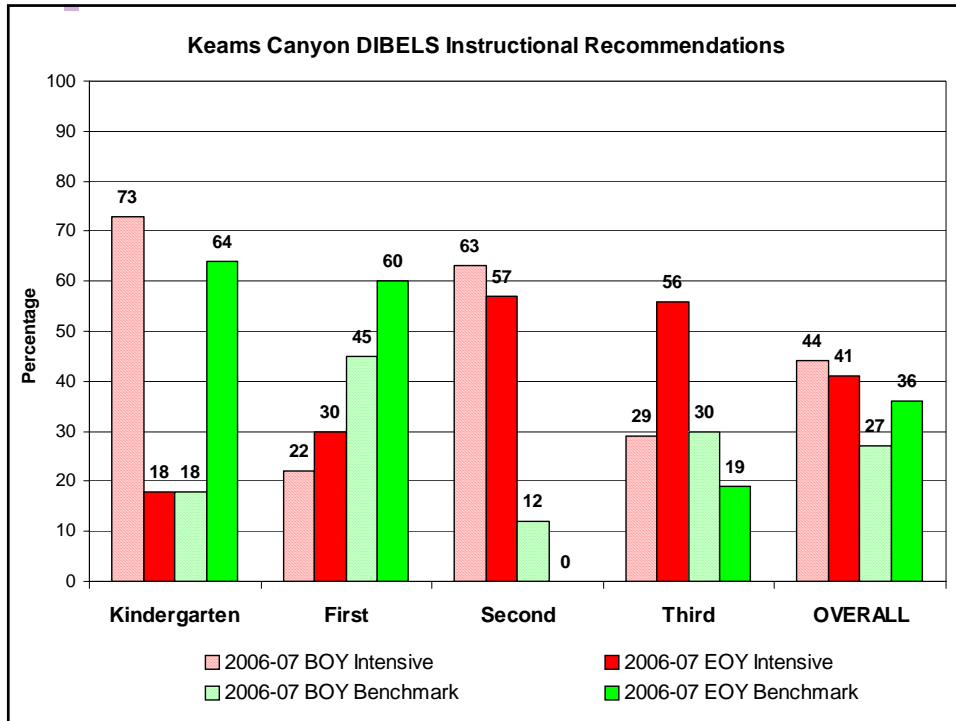
117





## Keams Canyon

- Conflict between admin and staff
- Need increased use of effective teaching practices
- Third grade DIBELS scores include 4<sup>th</sup> grade students, but percentages change little with 4<sup>th</sup> grade removed

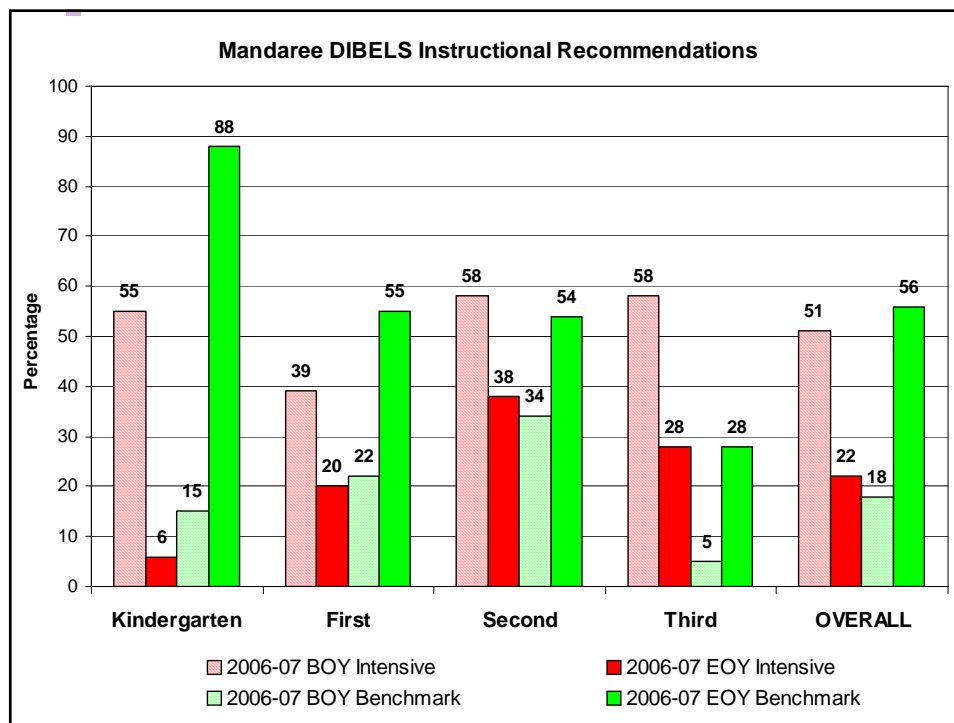


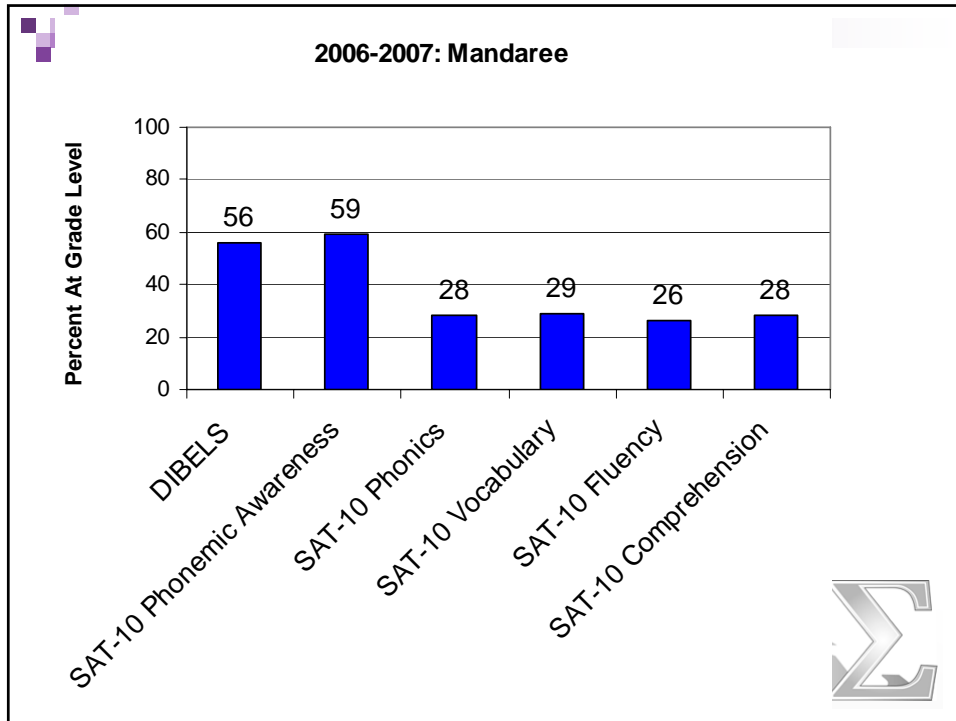
## Mandaree

- Highest performing Cohort II school
- Increased use of effective teaching practices
- New admin
  - Paraphrased from RC: “We worry about keep moral high and supporting each other—we have new administrators all the time” (something like 11 in 18 years)



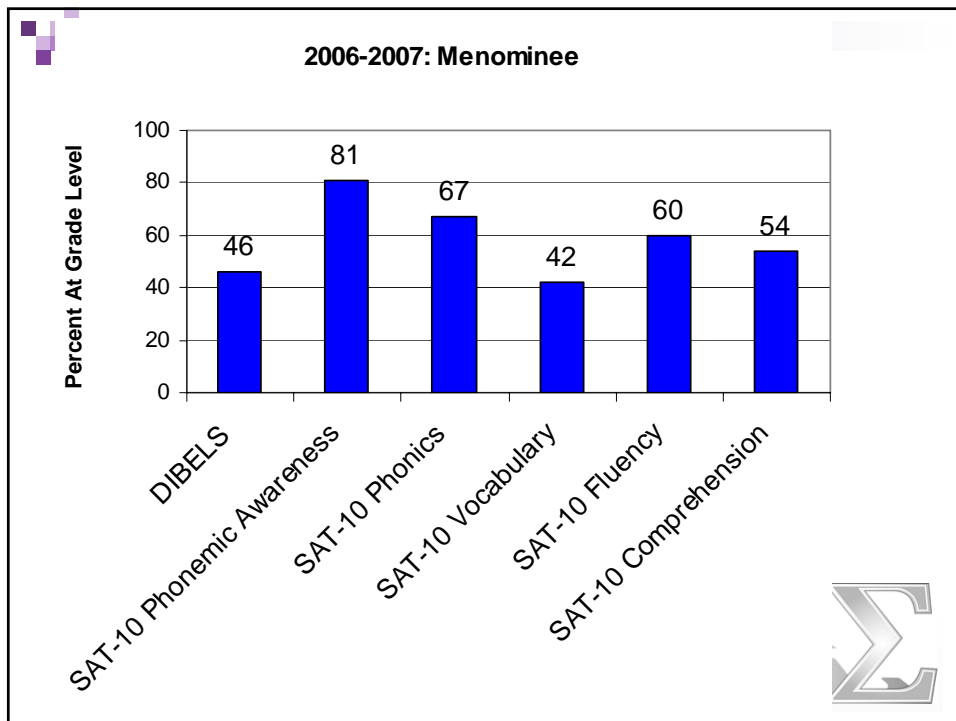
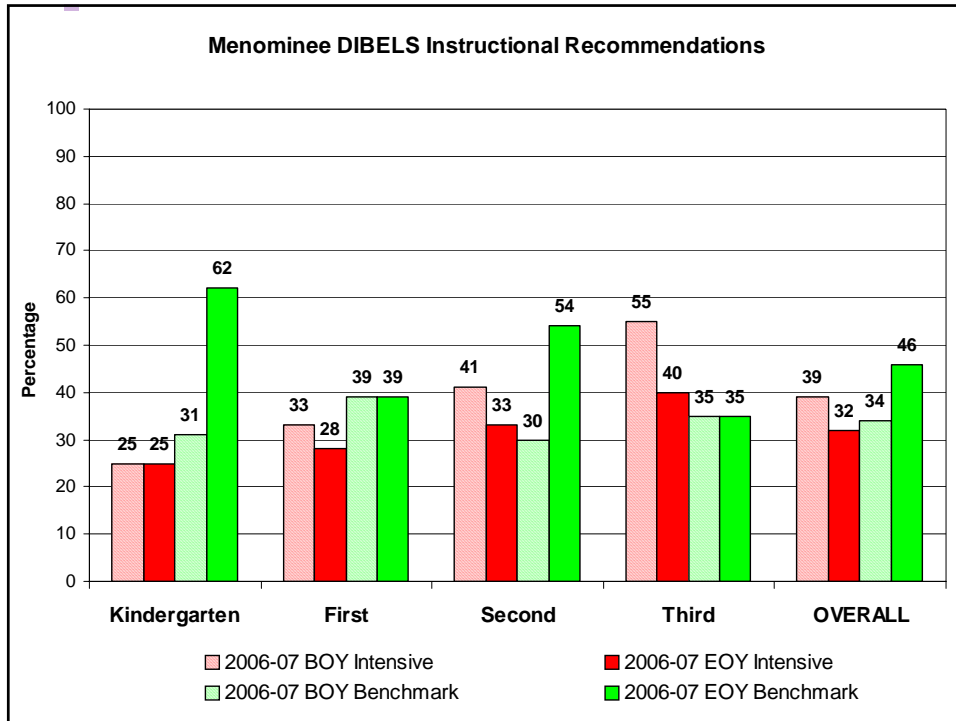
123





## Menominee

- Need increased used of effective teaching practices

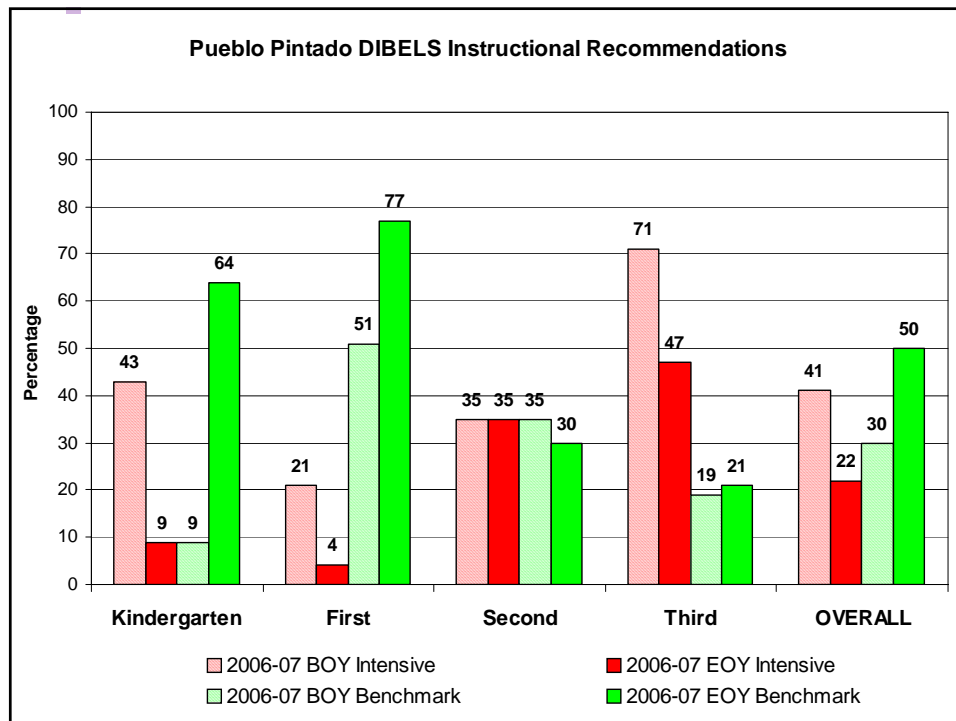


## Pueblo Pintado

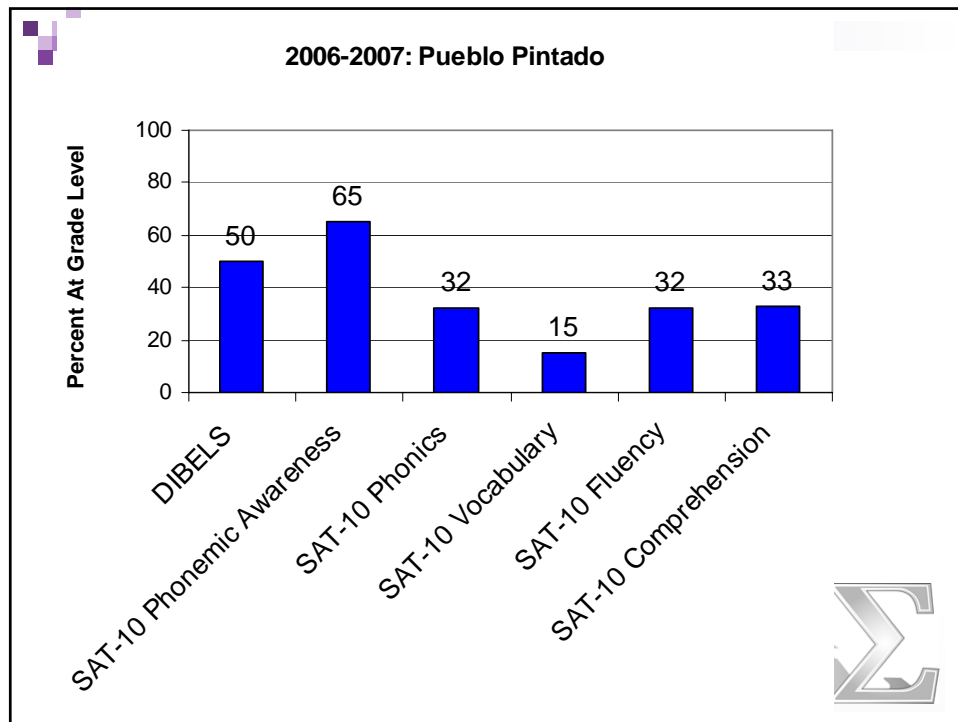
- Need to increase use of positive behavior management
- Turtle Mountain's reading coach has provided strong technical assistance
- Isolated administration



129

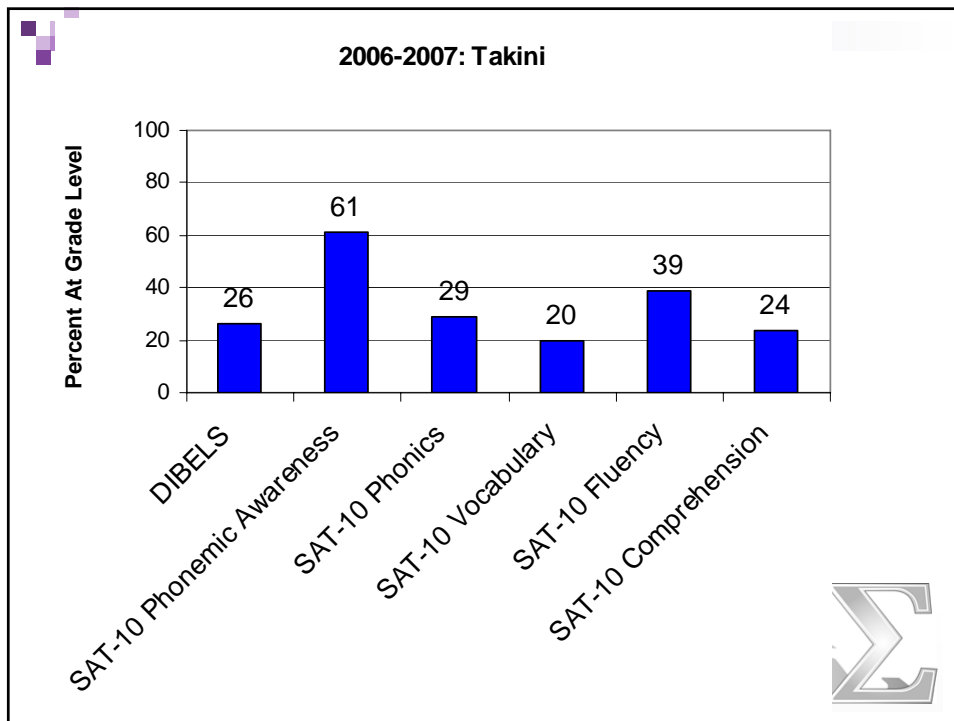
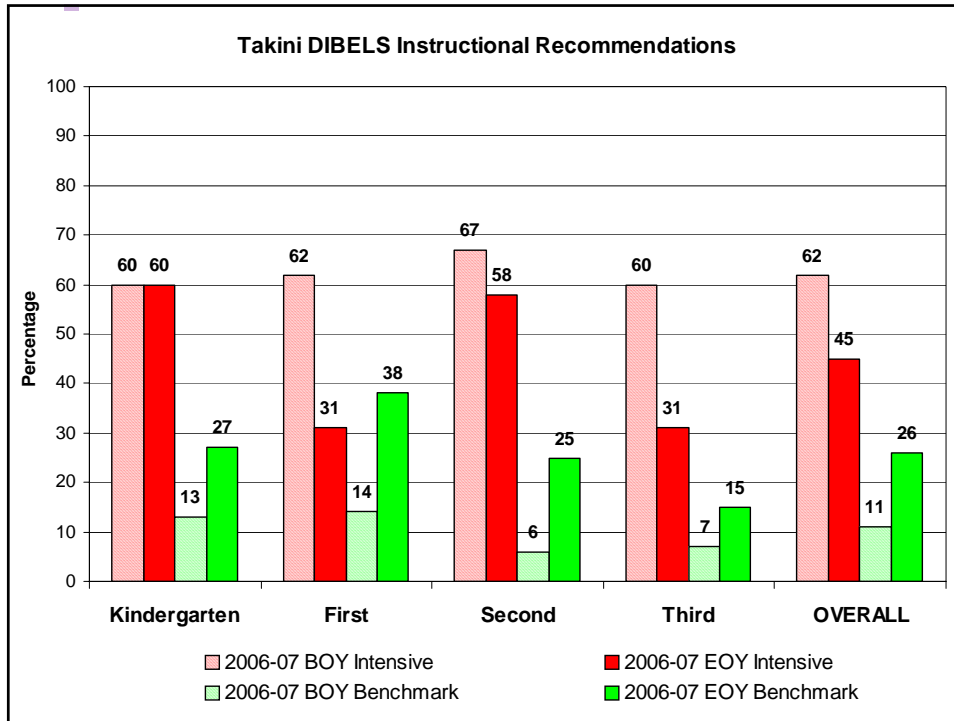






## Takini

- Reading coach fired mid-year 2006-2007
- New reading coach previously teaching staff
- New principal
- Ed Techs are well-trained, invested in students

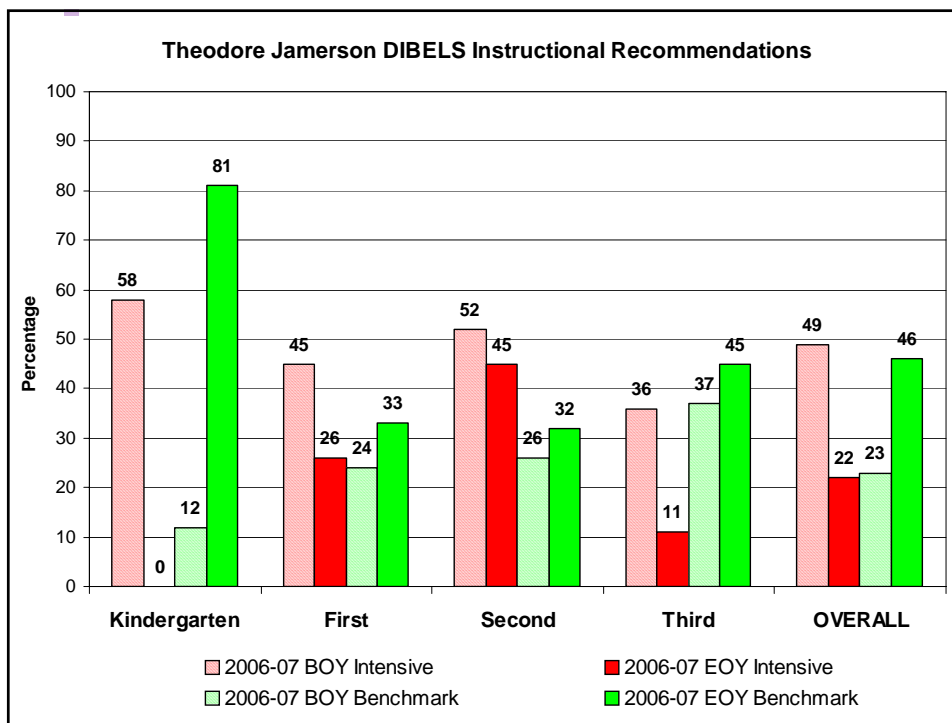


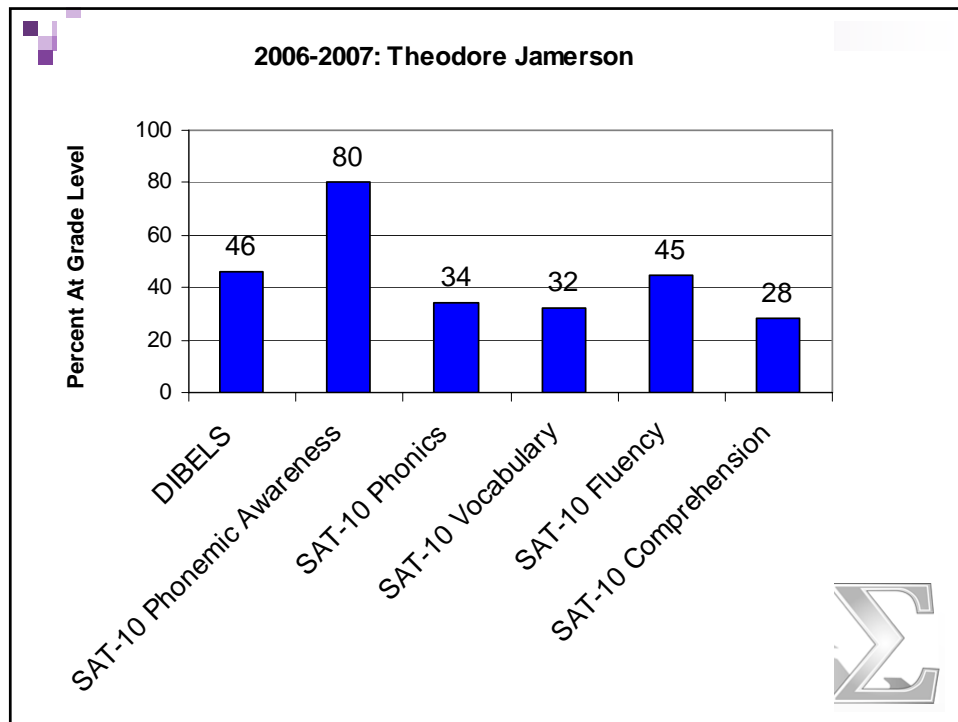
## Theodore Jamerson

- Highly transient school population
- Very crowded school
- High fidelity of implementation




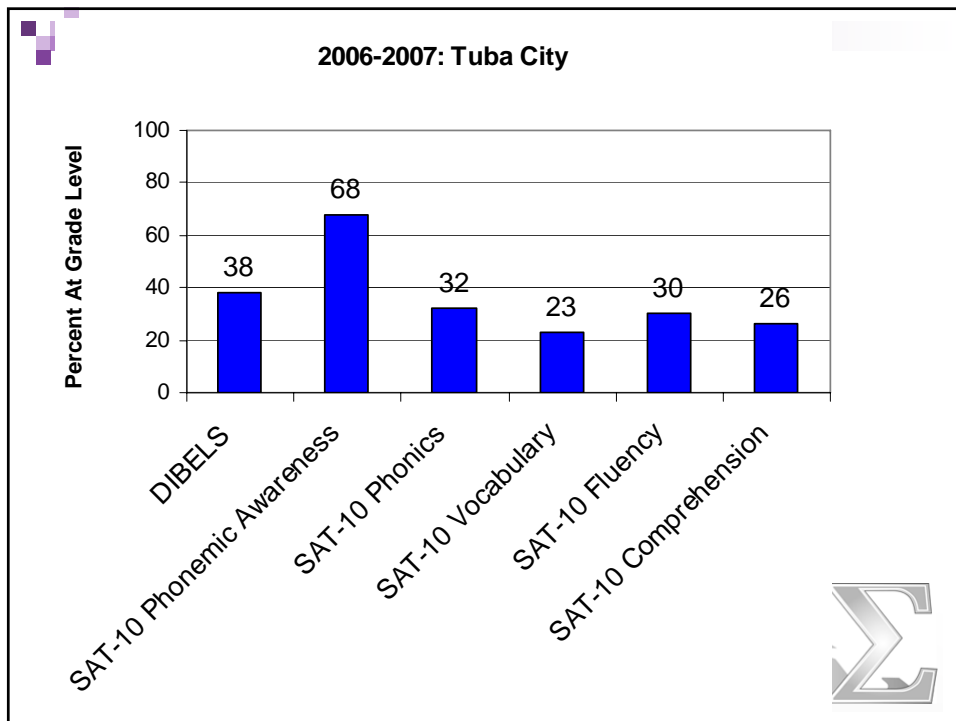
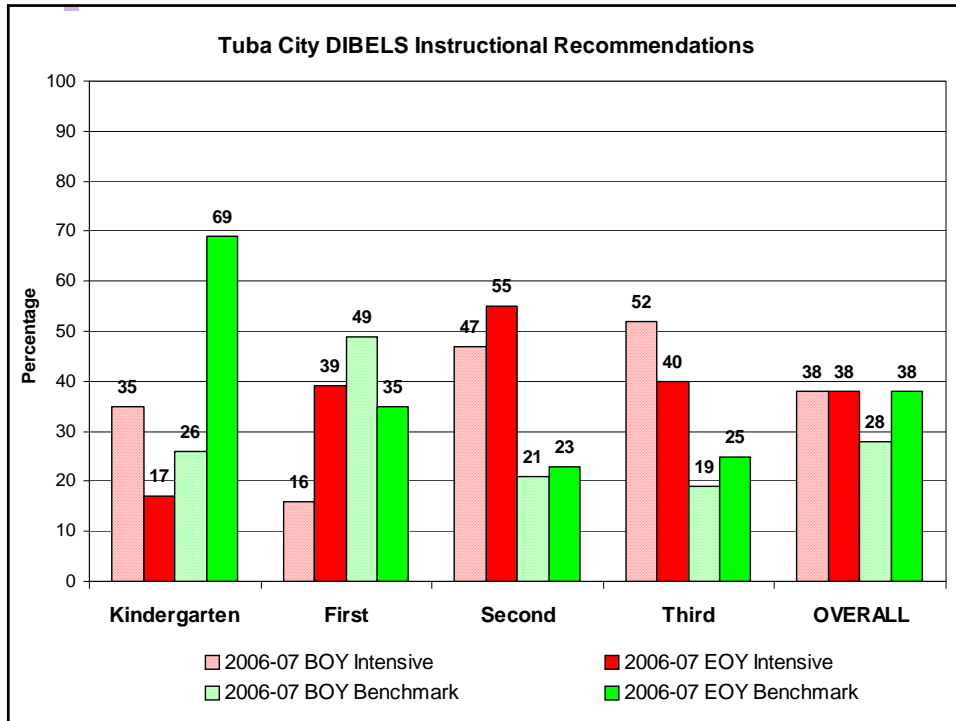
135





## Tuba City

- Large school
  - Reading coach not confident
  - Need increased use of effective teaching practices
  - Some well-trained Ed Techs
- 



## Twin Buttes

- Small school
- Changing administrative roles
- Need increased use of effective teaching practices



141

