

**WASHINGTON** | RESPONSE TO INTERVENTION

**Report of Activities  
and Findings**

**2009-10 School Year**

**June 2011**

**E R S**

---

Evaluation Research Services

**This document is not copyrighted; any or all may be duplicated.**

**The authors would appreciate credit for the materials used and a copy of the reprint.**



# Washington RTI

## Report of Activities and Findings 2009-10

---

Under contract with the Washington Office of Superintendent of Public Instruction (OSPI), Evaluation Research Services (ERS) began implementing a theory-driven, multi-method, stakeholder-oriented evaluation of *Improving Core Subject Instruction for All Students Pilot Project* (House Bill 2136) and *Re-Tooling Instruction through Response to Intervention (RTI)<sup>2</sup>* in early 2009 to “assess the effectiveness of implementing RTI and the impact of RTI implementation on the referral of students in special education programs (HB 2136)” and to “determine the effectiveness of professional development and supports for the implementation of RTI systems as provided by the project (RTI)<sup>2</sup>.” *Improving Core Subject Instruction for All Students Pilot Project* (HB 2136) expanded a program first initiated in two school districts (Vancouver and Walla Walla) in 2003. The more broadly authorized program (authorized in fall of 2007) has been in seven sites, including the two currently participating districts, Vancouver and Walla Walla. Participating districts 1) implement RTI as part of their general education curriculum for all students; 2) use a multi-tiered service delivery system to address student needs in reading, written language, and mathematics; 3) use an assessment system to conduct universal screening, progress monitoring, and to measure outcomes; 4) assure that parents are well-informed of student progress and their right to make a referral for a special education evaluation; and 5) participate in an annual evaluation of the program. Anticipated outcomes include a significantly greater proportion of students meeting state and federal performance standards and a reduction in the number of students inappropriately referred to and placed in special education.

Washington State’s (RTI)<sup>2</sup> project, *Re-Tooling Instruction through Response to Intervention*, targets the professional development needs related to RTI, particularly professional development at Tiers I, II and III. District personnel and staff are provided with ongoing support as they master the principles and practices related to RTI systems. (RTI)<sup>2</sup> provide(s) infrastructure necessary to the preparation of quality teachers and other personnel and to the support of parents of students with or at-risk for disability. The general foci of (RTI)<sup>2</sup> include: 1) development for birth-to-five providers, 2) regional support systems, 3) preparation of highly qualified teachers, and 4) direct support to exemplary demonstration sites. Specific project activities include: 1) reform and improve professional development infrastructure for training parents, school personnel, and early intervention personnel; 2) establish exemplary demonstration sites and identify and

disseminate effective RTI practices; 3) develop a system for ongoing support and sustainability of scientific research-based practices; and 4) enhance systems for the recruitment and retention of highly qualified personnel. The project is coordinated with the State's Performance Plan (Personnel Development System) and with other state-wide professional development efforts.

Evaluation Research Services has worked closely with officials from OSPI, with RTI coordinators throughout the state, and with teachers and administrators in local districts and schools to develop a multifaceted, mixed methods evaluation that addresses three purposes: 1) to provide local district and school officials with summary information from site visits and surveys that can be used to improve practice; 2) to provide OSPI officials with information necessary to state-wide management of the project; and 3) to provide OSPI and the federal Office of Special Education Projects (OSEP) with information on the program's efficacy. Twice-yearly reports are provided summarizing district and school site visits. Summary reports of the survey results are prepared, as well. A comprehensive summative report is also provided. The year-one report is available at <http://www.k12.wa.us/RTI/pubdocs/WA-RTI-FinalReport.pdf>. The present report summarizes *year-two* activities (2009-2010 school year) and findings to date (the projects are jointly referred to as WA State Personnel Development Grant or SPDG throughout this report).

### **Organization Of This Report**

We provide an overview of Year 2 evaluation-related activities (*Overview of Year 2 Activities*), followed by a summary of findings to date (*Summary of Findings to Date*) organized by evaluation question and by grade level (elementary, middle, secondary) within each question. For each question, we indicate: 1) the relevant data sources (*Data Sources*), 2) the rationale underlying the question (*Rationale*), 3) the particular method(s) used to address the question (*Methods*), 4) key findings (*Findings*), and 5) general conclusions (*Conclusions*). Recommendations, for practice and for continued evaluation, are included in a third section of the main report (*Recommended Next Steps*). Information in the *Summary of Findings to Date* and *Recommended Next Steps* sections is presented in bulleted format. Supporting documents, including presentation of school-level and district-level data, are available in the Appendices.

Three-year trends in achievement, behavior, and referral rates to special education are presented by grade level for participating districts and for matched comparison districts (see Question 1, Question 2, and Question 3). Differences in district trends were evaluated according

to overall levels of overall implementation. A mixed-methods approach (see the Proposal for Evaluation, 2008) informed data collection, providing a large qualitative data source that is used to contextualize the other body of findings. *School district* was the primary unit of analysis. Key data sources included:

- site visit data;
- survey data collected from personnel in funded districts and schools in the fall and spring;
- data from the state-administered reading, math, and writing assessments; and
- data on referrals to special education.

### **Overview of Year 2 Activities**

The Program Model (Figure 1), developed over a series of meetings with OSPI officials and a state-wide stakeholder group during the winter of 2008 and early spring of 2009 (see Interim Report, June 2009 for details), represents the norm for RTI implementation in Washington State (i.e., the planned model). It provides the basis for evaluating levels of RTI implementation in funded districts and schools and serves as a vehicle for generating evaluation questions and identifying measures. The Washington stakeholder group identified five general questions as key for the overall evaluation (across multiple years):

- To what extent are the School-wide Activities in Figure 1 implemented in funded districts?
- To what extent do Local Circumstances in Figure 1 inhibit or facilitate implementation of School-wide Activities?
- To what extent are School-wide Activities related to the Change Mechanisms in Figure 1?
- To what extent are School-wide Activities related to Outcomes in Figure 1?
- To what extent are *Change Mechanisms* related to *Outcomes* in Figure 1?

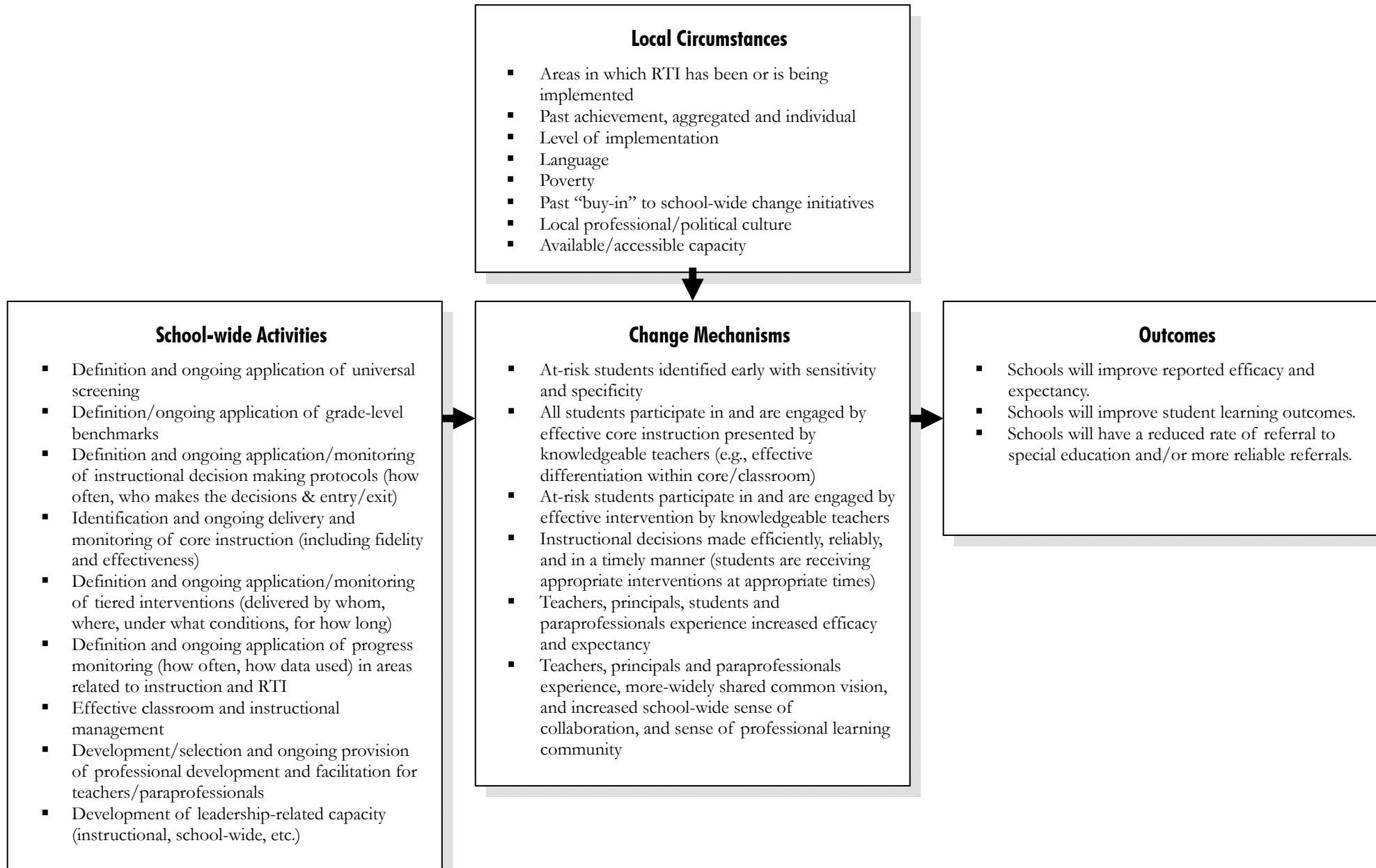
In year one, evaluation focused on the following questions: What are the 2008-2009 levels of student outcomes in the funded and matched districts? To what extent are the *School-wide Activities* in Figure 1 implemented in funded districts? What *Local Circumstances* do districts and schools find challenging?

The year 2 evaluation was centered on changes in the areas addressed in year 1. Specifically, questions included:

- To what extent have student achievement outcomes changed; how do these changes compare to changes in comparison districts?
- To what extent have rates of special education referral changed; how do these changes compare to changes in comparison districts?
- To what extent has implementation of the *School-wide Activities* in Figure 1 changed from 2008-2010; to what extent are differences in the implementation of *School-wide Activities* associated with achievement trends?

For purposes of the evaluation, we are using the implementation framework developed by the Center on State Implementation and Scaling-up of Evidence-based Practices (SISEP) and endorsed by the National Center on Response to Intervention (NCRTI) and by the Center on Instruction (COI). Like similar theories of adoption and use, the SISEP framework is built around stages of implementation that represent increasingly sophisticated levels of program use. While stage theories may oversimplify some features of what is an exceedingly complex process, they can, when well developed, provide a meaningful summary of the more critical aspects of successful implementation. The model includes seven stages: 1) no implementation, 2) exploration, 3) installation, 4) initial implementation, 5) full implementation, 6) innovation, and 7) sustainability ([www.fpg.unc.edu/~nirn/implementation/06/06\\_stagesimple.cfm](http://www.fpg.unc.edu/~nirn/implementation/06/06_stagesimple.cfm)).

**Figure 1: Program Model**



Site visit data were collected by local (i.e., residents of Washington) evaluators trained and supervised by Evaluation Research Services. The same team has been used in years 1 and 2 of the evaluation. Using the same site evaluators across data collection points and allowing them, to the extent possible, to work within the same school district(s) across the evaluation period has provided a degree of continuity. This model, combined with Dr. Jade Wexler’s ongoing support and hands-on direction of the collection effort, has yielded a relatively reliable and very extensive corpus of qualitative data. In the fall of year 2 (i.e., fall of 2009), site visit coordinators conducted initial middle school site visits. They also conducted follow-up visits to the elementary school visited in year 1 of the evaluation. In the spring of the school year (spring 2010), they conducted a second follow up in the identified elementary school, a follow up to the middle school visited in the fall, and an initial high school visit. Summary reports for all visits are in the Appendix to this report. Survey data were collected in the fall and spring of the school year under the supervision of Dr. Sarojani Mohammed. The Appendix also provides summaries of survey results for this year for both survey rounds, organized by question, by district, by school level (i.e., elementary vs. middle vs. high schools) and by (self-reported) implementation status (i.e., whether a school self-identified as “implementing RTT” or not). Achievement data, screening data, information on special education referrals, behavioral referrals, and dropout were provided by officials from OSPI.

**Table 1 – Funded and Matched Districts with Original Matching Criteria**

| Status | District            | Enrollment k-8 | Pct. Special Education (3 to 21) | Pct. free/reduced lunch | Pct. minority (non-white) |
|--------|---------------------|----------------|----------------------------------|-------------------------|---------------------------|
| RTI    | Odessa              | 221            | 14%                              | 45%                     | 5%                        |
| Match  | Curlew              | 233            | 12%                              | 51%                     | 9%                        |
| RTI    | Stevenson-Carson    | 1,127          | 14%                              | 37%                     | 17%                       |
| Match  | Chimacum            | 1,129          | 13%                              | 37%                     | 14%                       |
| RTI    | Kiona-Benton City   | 1,528          | 16%                              | 57%                     | 33%                       |
| Match  | White Salmon Valley | 1,229          | 16%                              | 47%                     | 36%                       |
| RTI    | Hoquiam             | 1,980          | 10%                              | 56%                     | 22%                       |
| Match  | Aberdeen            | 3,459          | 15%                              | 59%                     | 30%                       |
| RTI    | East Valley         | 4,182          | 14%                              | 48%                     | 13%                       |
| Match  | Cheney              | 3,877          | 15%                              | 43%                     | 16%                       |
| RTI    | White River         | 4,329          | 12%                              | 25%                     | 11%                       |
| Match  | Stanwood-Camano     | 5,362          | 13%                              | 22%                     | 11%                       |
| RTI    | Walla Walla         | 6,186          | 12%                              | 52%                     | 37%                       |
| Match  | Eastmont            | 5,482          | 11%                              | 53%                     | 39%                       |
| RTI    | Longview            | 7,271          | 13%                              | 48%                     | 26%                       |
| Match  | Kelso               | 5,185          | 13%                              | 45%                     | 20%                       |
| RTI    | Shoreline           | 9,168          | 12%                              | 22%                     | 38%                       |
| Match  | Central Kitsap      | 11,886         | 14%                              | 25%                     | 32%                       |
| RTI    | Richland            | 10,599         | 12%                              | 30%                     | 18%                       |
| Match  | Peninsula           | 9,443          | 12%                              | 21%                     | 20%                       |
| RTI    | Vancouver           | 22,622         | 11%                              | 48%                     | 28%                       |
| Match  | Evergreen (Clark)   | 26,100         | 12%                              | 36%                     | 27%                       |

As indicated, the general unit of analysis is *school district*. To evaluate progress of the funded districts, 11 non-funded districts were matched in the fall of 2008 on total enrollment, percent served by special education, percent free-reduced lunch, and percent minority. The funded and matched districts are summarized in Table 1. In the year-one report (see <http://www.k12.wa.us/RTI/pubdocs/WA-RTI-FinalReport.pdf>), we evaluated the degree to which the matched and unmatched groups were comparable on key student outcomes. As indicated in that report, the reliability of the match was confirmed based on year 1 achievement data. The several areas where the level of match was less reliable are handled analytically in this report (i.e., by covarying initial differences). In the present report, we focus more directly on differences in outcomes and on the degree to which these differences may correspond with other elements of the Program Model.

## Summary of Findings to Date

---

### Question 1: To what extent have student achievement outcomes changed; how do these changes compare to changes in comparison districts?

#### Data Sources

- Percentage of students scoring at or above “proficient” on the 2008 and 2009 Washington Assessment of Student Learning (WASL) and on the 2010 [Measurements of Student Progress \(MSP\)](#) for grades 3 through 8 and the [High School Proficiency Exam \(HSPE\)](#). We present data by grade levels.

#### Rationale

- The impact of participating in RTI is a key question for the evaluation. Changed practice is supportable to the extent that it supports better outcomes, whether those outcomes are achievement-, behavioral-, or cost-related. The achievement question is addressed in this section of the report.

#### Method

- A key evaluation question is the extent to which participating in RTI is associated with improved student outcomes, sometimes discussed in terms of *program impact*. Findings related to impact should be considered in the context a study design’s internal validity, generally understood as the extent to which one can confidently claim that an intervention “causes” observed differences in groups. Matched-groups designs, typically described as “quasi-experimental,” have moderate levels of internal validity, depending on the extent to which the two groups are similar and on the evaluator’s ability to prevent (or at least to measure) implementation of the program in question (RTI in this case) by comparison sites. In this case, the degree to which RTI has *not* been adopted or has not influenced practice in the matched sites is unknown, though anecdotal reports from local site evaluators suggest that it may have a marginal influence at best. The quality of the match is relatively strong, with notable exceptions in 4<sup>th</sup> grade and 8<sup>th</sup> grade, based on the comparability of average test scores in 2008. Differences in 2008 favor the comparison condition.

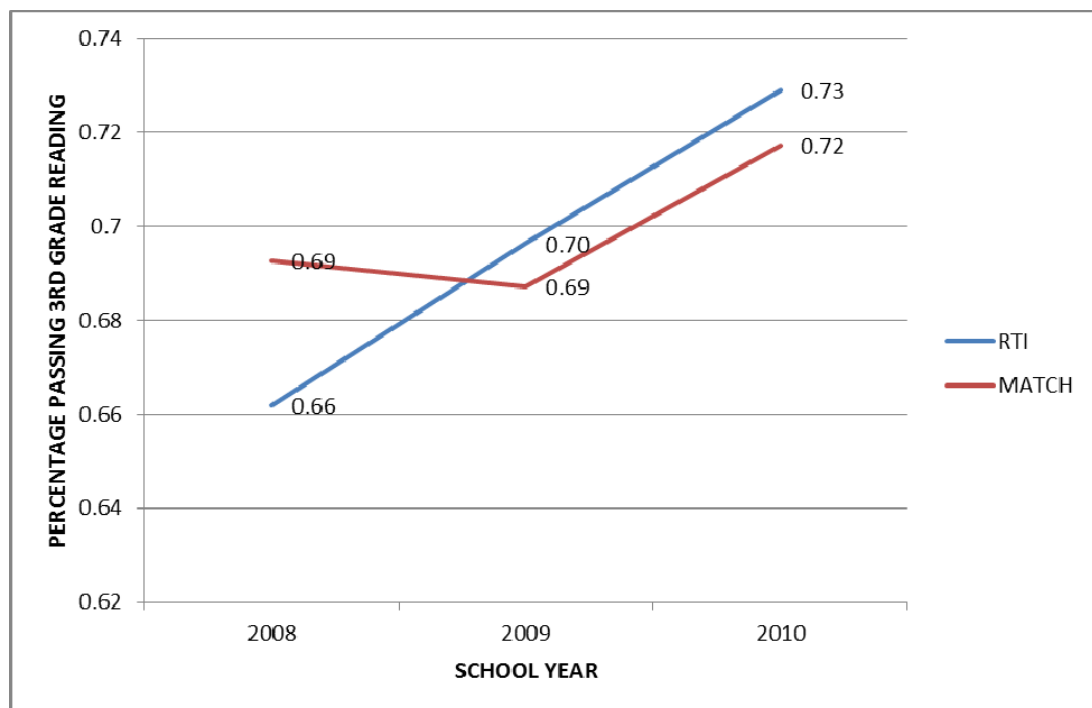
- With only 22 degrees of freedom (11 schools in each group), tests of statistical significance are limited in power and the differences in 2010 rates of passing are not statistically different. However, effect sizes for *adjusted* 2010 values (differences in 2010 adjusted for 2008 differences) offer a meaningful alternative for evaluating change. An effect size is a standardized estimate of change associated with time or treatment. The matched comparison design of this study suggests that differences in 2010 may be due at least in part to implementing an RTI model rather than, or in addition to, other factors such as passing time or the implementation of other initiatives. While the small sample size (number of districts) should temper claims that RTI “caused” these differences, the relatively consistent pattern of findings, particularly in the elementary grades, as discussed in the following sections, are worth noting.

## ■ Elementary School Findings (Grades 3-5)

### Third Grade Results

- On average, RTI-participating districts have outperformed matched districts in 3<sup>rd</sup> grade reading in the percentage of students passing the state-administered achievement measure from 2008 through 2010. The group of funded districts improved from 66% passing in 2008 to 73% passing in 2010, while comparison districts improved by about 3 percentage points during the same time period (69% to 72%). Figure 2 depicts these trends.

**Figure 2 – Comparison of Trends in 3<sup>rd</sup> Grade Reading**

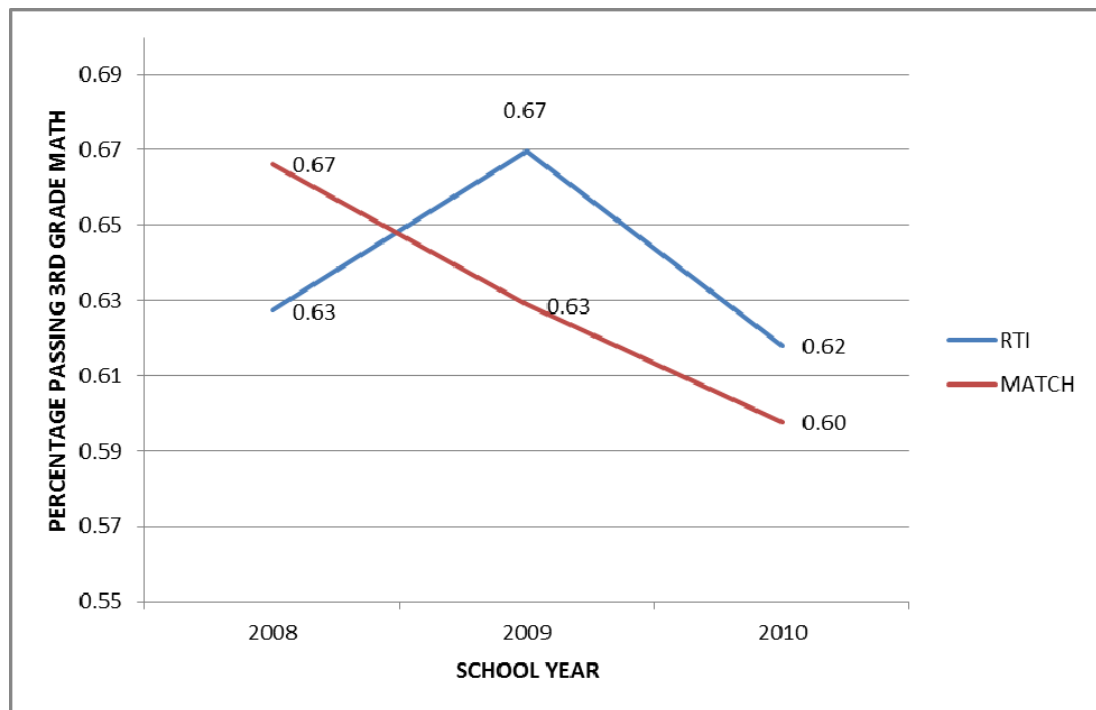


Note: The WASL was administered in 2008 and 2009. The MSP and HSPE were used in 2010.

- Math trends in the RTI group compare favorably, as well, although the pattern differs from reading (Figure 3). In comparison schools, the percent of students passing the state-administered test decreased steadily over the three-year period, from 67% in 2008 to 63% passing in 2009 and 60% in 2010. The trend in the RTI group was more variable, with 63% passing in 2008 and 63% in 2010. In 2009, 67% of 3<sup>rd</sup>-grade students in RTI schools passed the WCAP Math Test.

- As before, the limited power of the school-level analysis complicates tests of statistical significance. However, the effect size associated with the adjusted 2010 group differences was  $\eta^2 = .07$ . (Effect sizes represent the amount of change experienced by one group compared to the change experienced by another group. In this case, the two groups are 1) participating districts and 2) matched comparison districts and the effect size, reported as eta-squared, indicates the size of the difference in post-test means once preprogram differences are controlled.) This is a moderately sized effect and represents a potentially meaningful difference. At the same time, it is worth noting that the effect is due largely to the downward trend in the comparison districts. Average performance differences between the 2008 and 2010 3<sup>rd</sup> grade cohort were minimal in the RTI districts (63% versus 62% passing).)

**Figure 3 – Comparison of Trends in 3<sup>rd</sup> Grade Math**

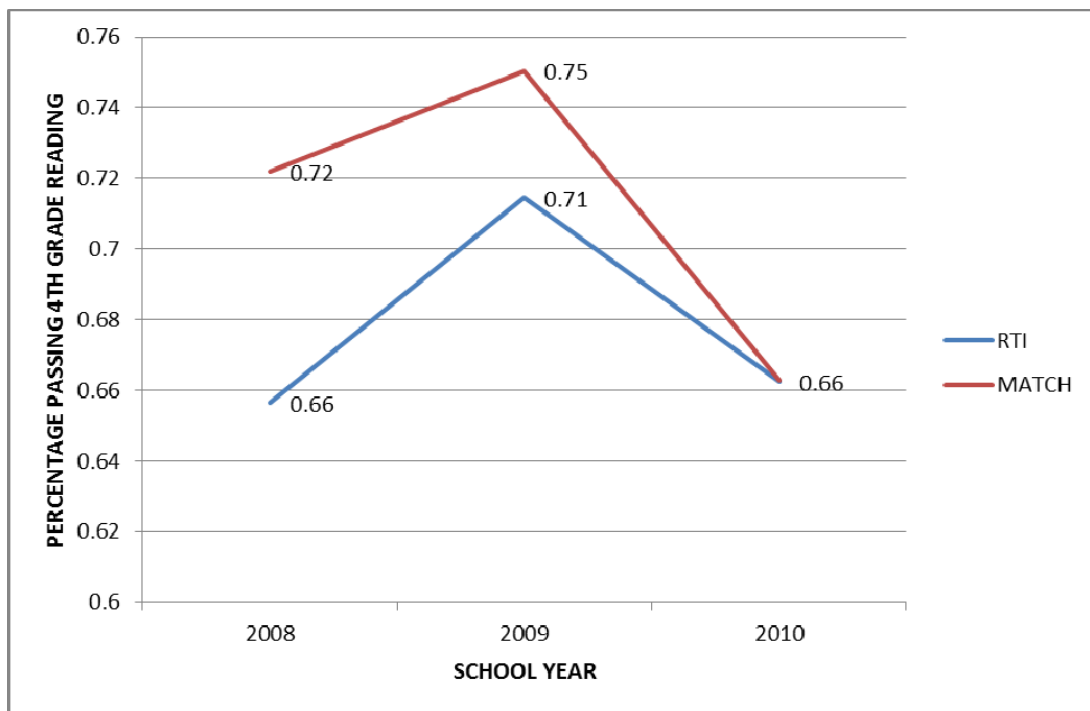


Note: The WASL was administered in 2008. The WCAP was used in 2009 and 2010.

### Fourth Grade Findings

- In 4<sup>th</sup> grade, average reading performance (Figure 4) in RTI-participating districts was relatively stable, with 66% of students passing in 2008 and 66% passing in 2010. Seventy-one percent passed in 2009. In the comparison schools, average performance trended downwards, with average passing rates decreasing from .72 in 2008 to .66 in 2010. While the two groups scored comparably in 2010, this represented a relative decline in the comparison and no change in the participating schools. The adjusted 2010 difference is associated with an effect size of  $\eta^2 = .048$  and is due to decline in the comparison group, generally described as on the low-end of the category of moderately-sized effects.

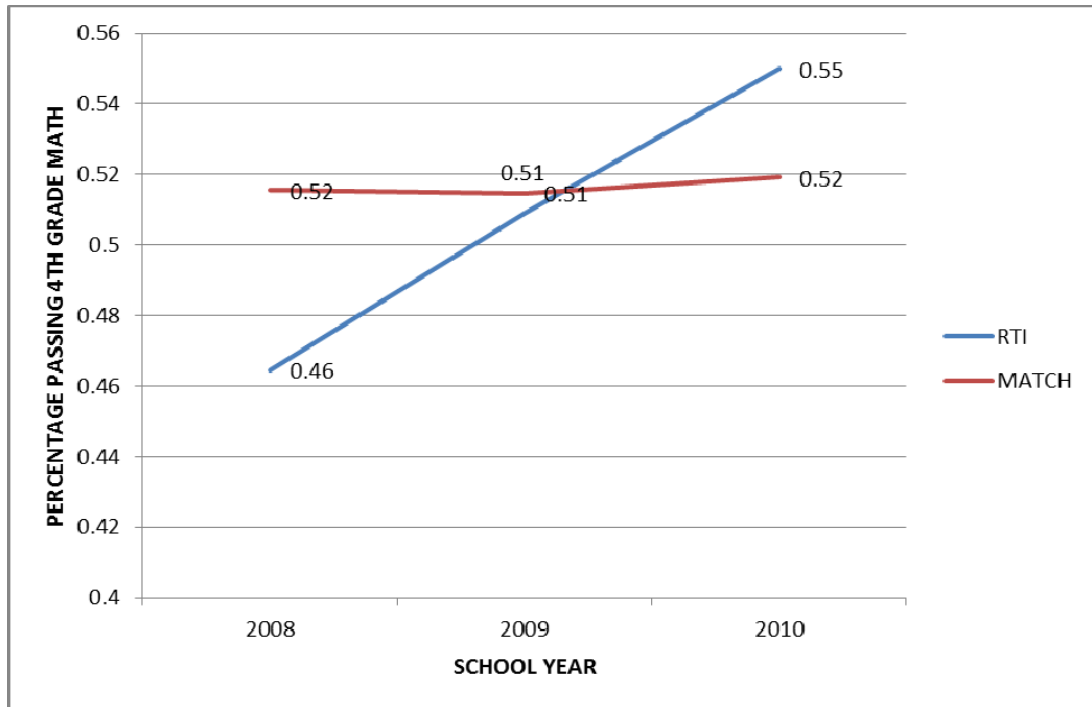
**Figure 4 – Comparison of Trends in 4<sup>th</sup> Grade Reading**



Note: The WASL was administered in 2008 and 2009. The MSP and HSPE were used in 2010.

- In math (Figure 5), the trends were more marked. Participating districts, on average, made consistent and large gains in the percentage of students scoring at or above the passing cut score (from 46% in 2008 to 55% in 2010). Trends in the comparison districts were relatively flat (51% in 2008 and 2010 and 52% in 2009).

**Figure 5 – Comparison of Trends in 4<sup>th</sup> Grade Math**

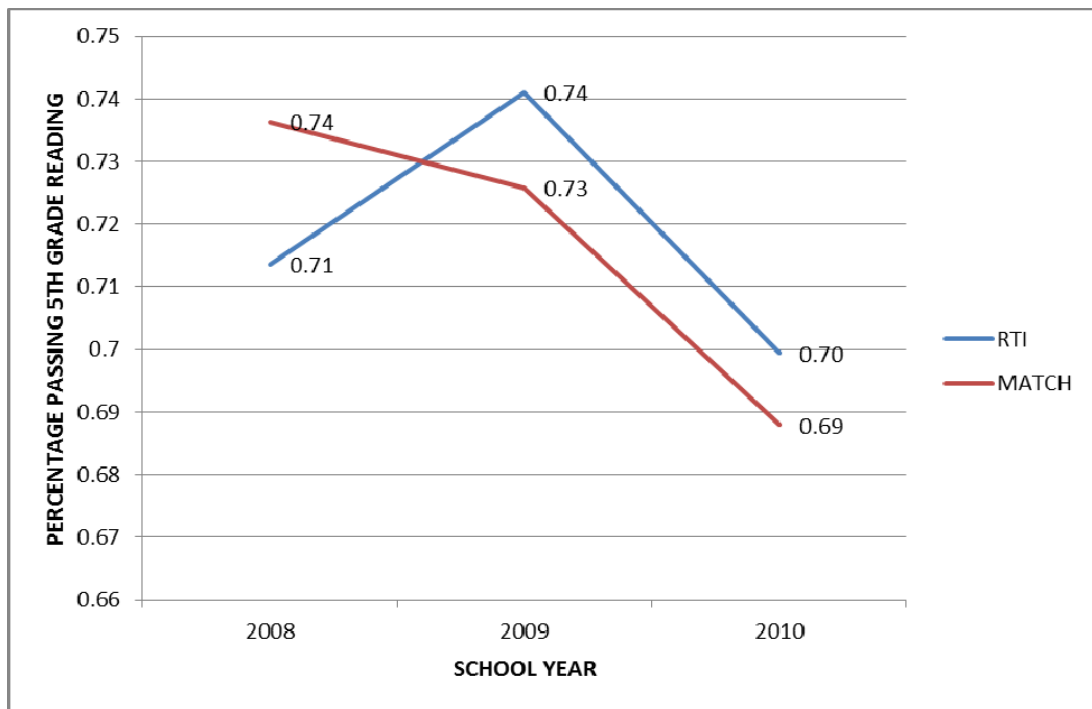


Note: The WASL was administered in 2008 and 2009. The MSP and HSPE were used in 2010.

### Fifth Grade Findings

- The trends in 5<sup>th</sup> grade are similar to those in earlier grades. In reading (Figure 6), participating districts were relatively stable over the three-year period, with 71% passing in 2008 and 70% in 2010 (74% passed on 2009). In the matched-groups comparison, there was a clear downward trend, from 74% to 69% passing over the 3-year span ( $\eta^2 = .04$ ).

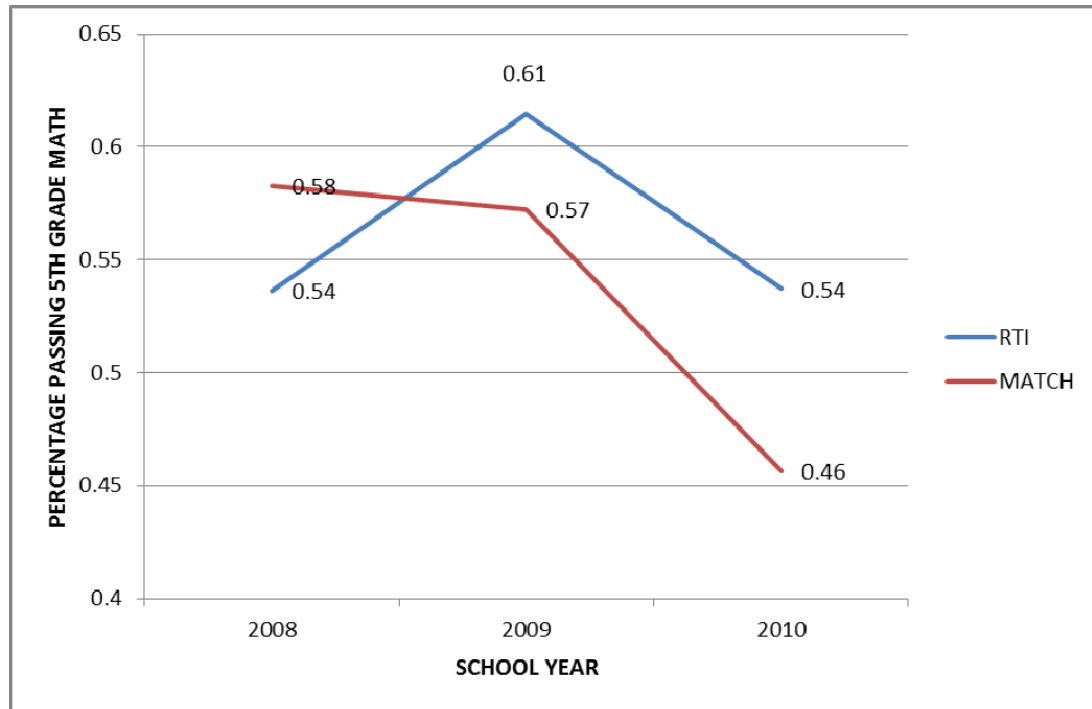
**Figure 6 – Comparison of Trends in 5<sup>th</sup> Grade Reading**



Note: The WASL was administered in 2008 and 2009. The MSP and HSPE were used in 2010.

- In 5<sup>th</sup>-grade math (Figure 7), RTI districts passing rates averaged 54% in 2008 and 54% in 2010. In the matched comparison districts, rates declined from 58% to 46% passing ( $\eta^2 = .07$ ).

**Figure 7 – Comparison of Trends in 5<sup>th</sup> Grade Math**



Note: The WASL was administered in 2008 and 2009. The MSP and HSPE were used in 2010.

### Summary of Elementary School Findings

- The comparison group, on average, was higher at pretest on all measures at all grade levels, while participating districts scored higher (or as high in the case of 4<sup>th</sup> grade reading) than matched comparisons at all elementary grade levels for both reading and math. The standardized mean differences for reading adjusted for 2008 differences (i.e., effect sizes conditioned on pretest differences) ranged from .03 in 3<sup>rd</sup> grade to .05 in 4<sup>th</sup> grade. In math, differences ranged from and .07 in 3<sup>rd</sup> and 5<sup>th</sup> grade to .09 in 4<sup>th</sup> grade.

**Table 2 – Adjusted Effect Sizes**

|         | 3 <sup>rd</sup> Grade | 4 <sup>th</sup> Grade | 5 <sup>th</sup> Grade |
|---------|-----------------------|-----------------------|-----------------------|
| Reading | .03                   | .05                   | .04                   |
| Math    | .07                   | .09                   | .07                   |

Note: Effect sizes reported at eta-squared.

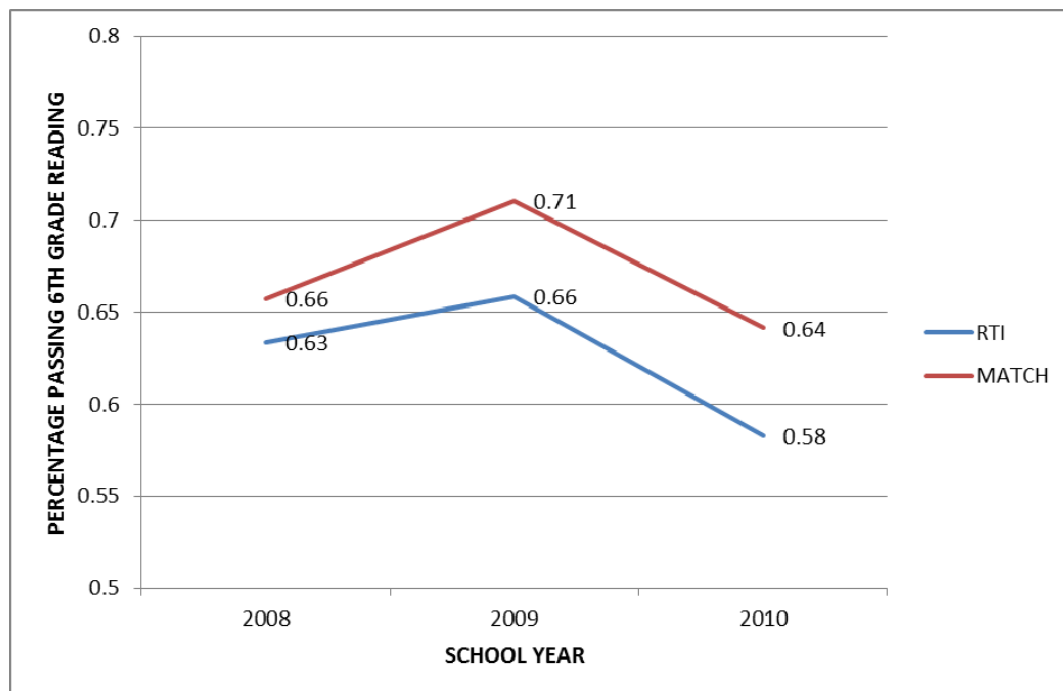
- These results represent district-wide trends. School-level data are not available at this time. The data reported in this section do represent differing levels of RTI implementation. A later section of this report addresses the relationship of program implementation and achievement trends.
- Finally, the spring 2010 change in test, from WASL to the MSP/HSPE (only MSP was used at the elementary level), may explain, in whole or in part, the general decrease in average scores (3<sup>rd</sup>-grade reading and 4<sup>th</sup>-grade math are exceptions) across the two groups. To the extent that changes in measurement do not influence scores *differently* in the participating and matched comparison groups (an unlikely possibility), the interpretation of group comparison results are unlikely to be compromised.

## ■ Middle School Findings (Grades 6-8)

### Sixth Grade Findings

- The middle school findings generally do *not* indicate an effect for RTI in participating districts. As in the elementary school data, RTI participating districts were lower in 2008 than match comparison. However, the positive trends in the lower grades for participating districts (in relation to the comparison districts) are not apparent in the middle schools, (except in 8<sup>th</sup> grade reading). In 6<sup>th</sup>-grade reading (Figure 8), about 63% of students in RTI districts score at or above proficient in 2008, compared to 66% in the matched districts. In 2010, 58% of 6<sup>th</sup> grade students in the participating districts scored at or above the cut score, while 64% of students in the comparison districts met the standard.

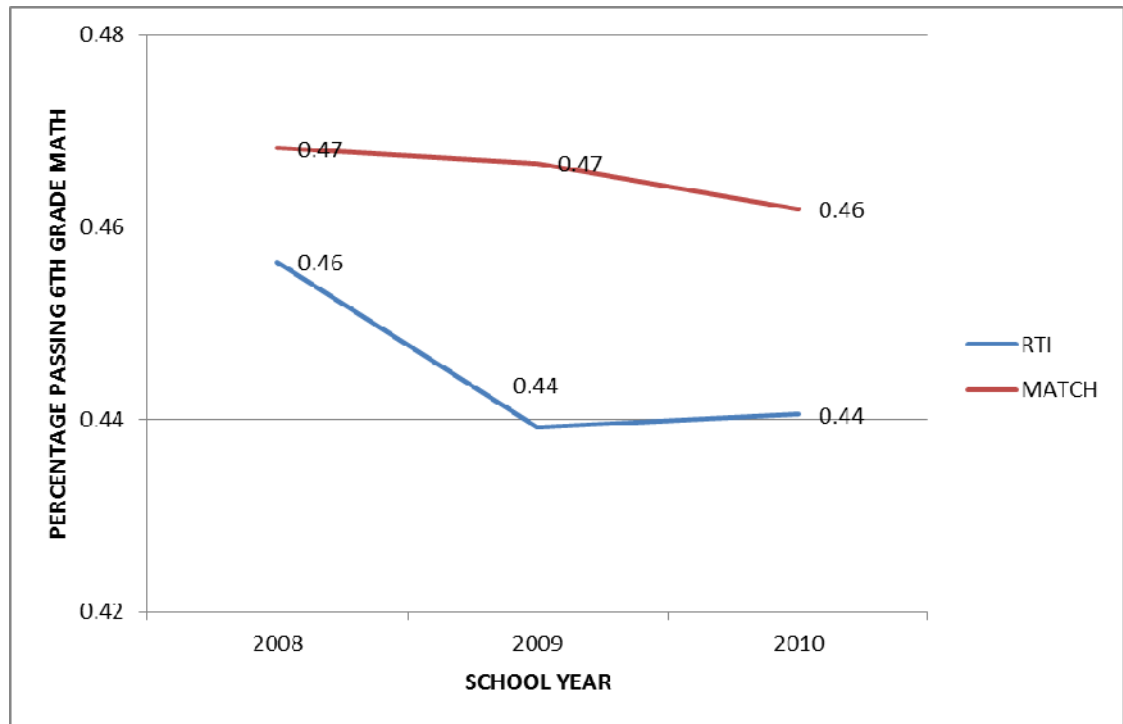
**Figure 8 – Comparison of Trends in 6<sup>th</sup> Grade Reading**



Note: The WASL was administered in 2008 and 2009. The MSP and HSPE were used in 2010.

- In math, the same general pattern held, with matched districts performing better in 2008, 2009, and 2010, on average, than 6<sup>th</sup>-grade students in the participating districts (Figure 9).

**Figure 9 – Comparison of Trends in 6<sup>th</sup> Grade Math**

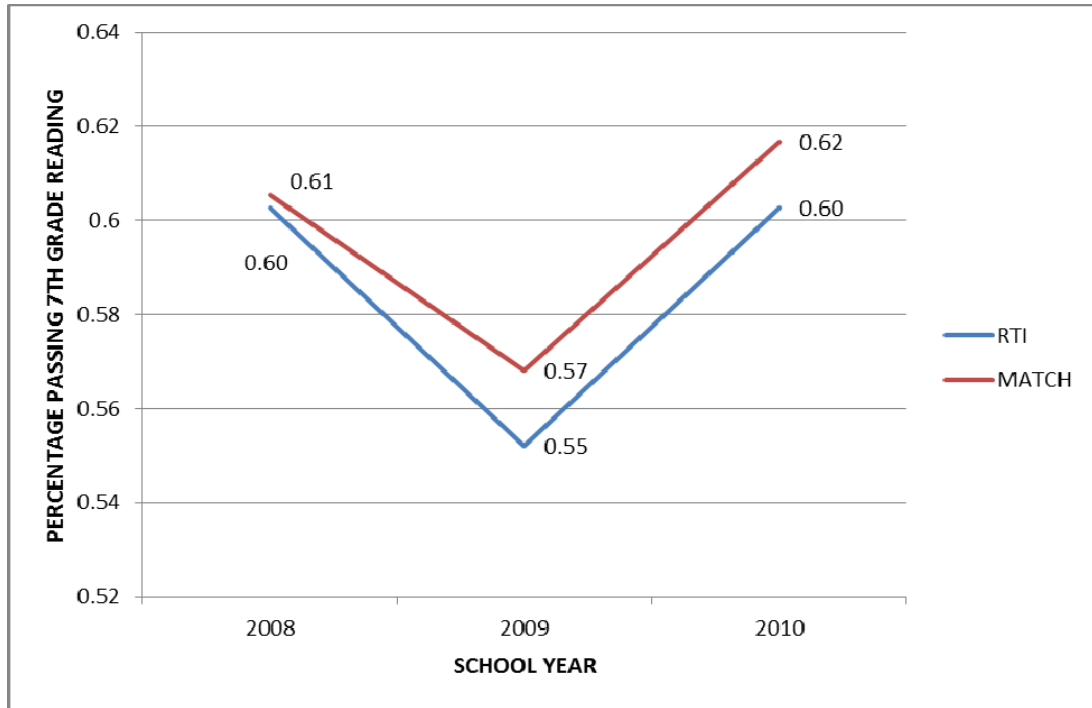


Note: The WASL was administered in 2008 and 2009. The MSP and HSPE were used in 2010.

### Seventh Grade Results

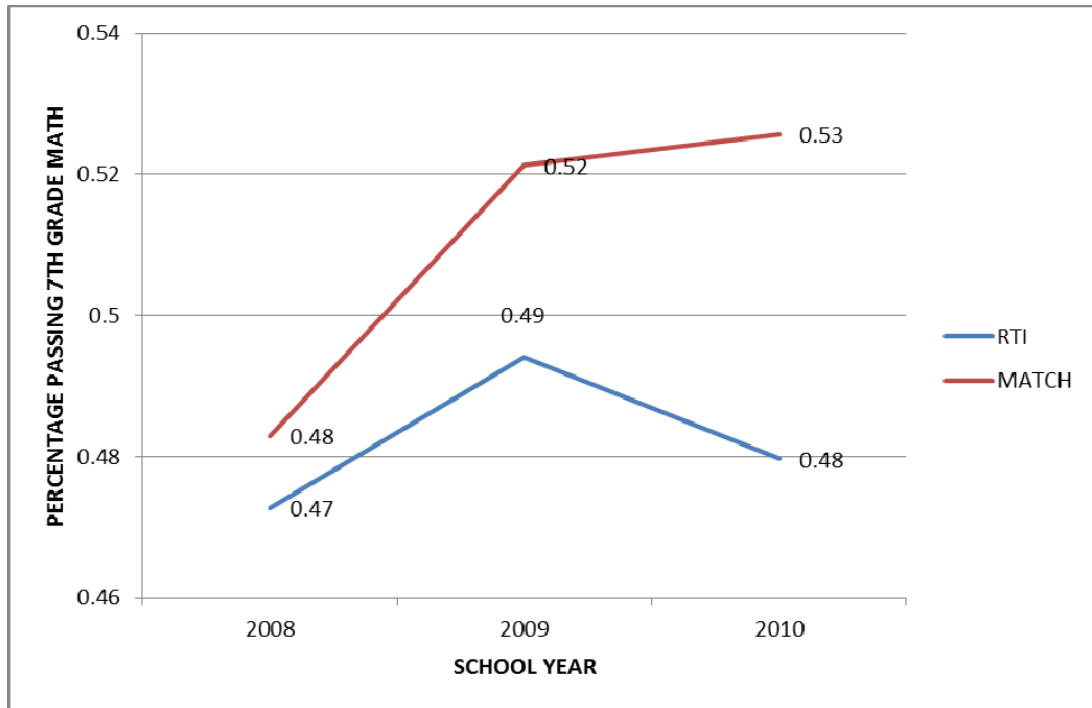
- In 7<sup>th</sup> grade reading (Figure 10), there was very little change in either group. The group of participating districts averaged 60% “passing” in 2008 and 2010. The figures in the matched comparison were 61% and 62% for 2008 and 2010, respectively.
- In math (Figure 11), the trend in the matched comparison group differed distinctly from the average trend in the participating districts. In 2008, 47% of students in participating districts scores at or above proficient, while 48% of 7<sup>th</sup> graders scored comparably in 2010. In the matched comparisons, over one-half (53%) of students scored at or above the standard in 2010; 48% scored at this level in 2008.

**Figure 10 – Comparison of Trends in 7<sup>th</sup> Grade Reading**



Note: The WASL was administered in 2008 and 2009. The MSP was used in 2010.

**Figure 11 – Comparison of Trends in 7<sup>th</sup> Grade Math**

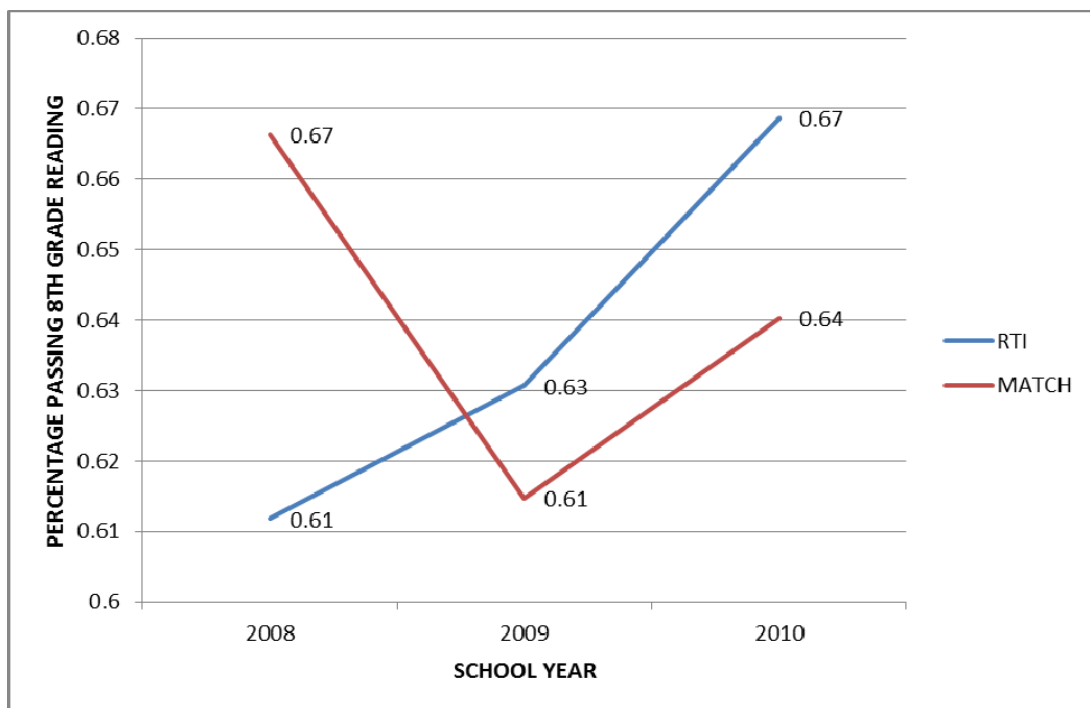


Note: The WASL was administered in 2008 and 2009. The MSP was used in 2010.

## Eighth Grade Results

- On the 8<sup>th</sup>-grade reading measure, RTI-participating districts outperformed the group of matched districts by an overall difference of 9 percentage points. While the percentage of 8<sup>th</sup> graders in the RTI group increased from 61% to 67% over the three-year period, the proportion of students in comparison districts decreased from 67% in 2008 to 64% in 2010, with a dip to 61% in 2009. The effect size in reading was  $\eta^2 = .15$ , which represents a large-sized effect.
- This finding contrasts with results for the other middle school grade levels and with the results for math in 8<sup>th</sup> grade. The reasons are not clear based on the site visit data and results of the survey data.

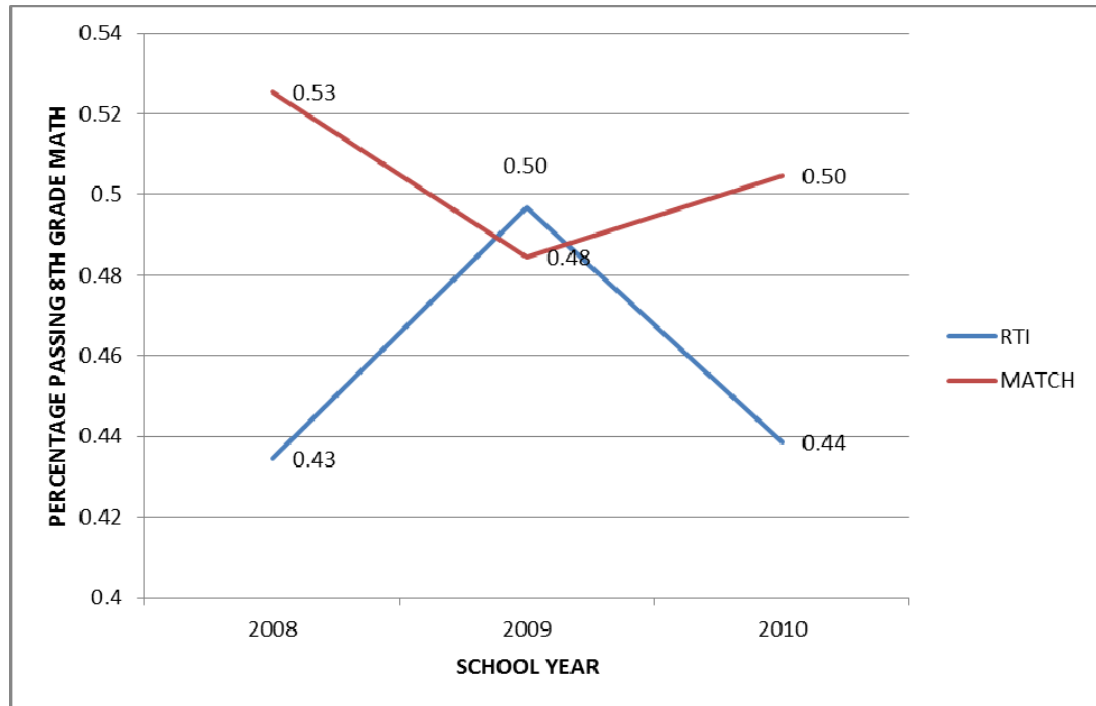
**Figure 12 – Comparison of Trends in 8<sup>th</sup> Grade Reading**



Note: The WASL was administered in 2008 and 2009. The MSP and HSPE were used in 2010.

- In math, the participating districts scored below the matched comparisons in 2008 and in 2010. However, in terms of net change, the participating districts gained a percentage point (from 43% to 44%) while the comparisons lost 3 points (from 53% to 50%).

**Figure 13 – Comparison of Trends in 8<sup>th</sup> Grade Math**



Note: The WASL was administered in 2008 and 2009. The MSP and HSPE were used in 2010.

### Summary of Middle School Findings

- Matching in 8<sup>th</sup> grade may be less successful than at other grade levels, based on 2008 percent-passing averages in reading and math. Assuming that the 2008 estimates are reliable, there is no evidence that the differences compromise results of the comparisons.
- Site visit and survey data indicate low levels of implementation at middle school. Even in early adopting districts, there is uncertainty around what represents best practice at these higher grade levels. Many of these districts are actively developing RTI-aligned models and tools that address local needs and fit local circumstances. This trial-and-error process, even when organized within a relatively well established framework like RTI (even one that is built on the notion of trial and error), is often associated with near-term decreases in average achievement (ref.), as educators at the school, district, and state levels refine basis processes and tools. This argument is hypothetical; confirming data are

not and will not be available until RTI models in early adopting districts have had an opportunity to mature.

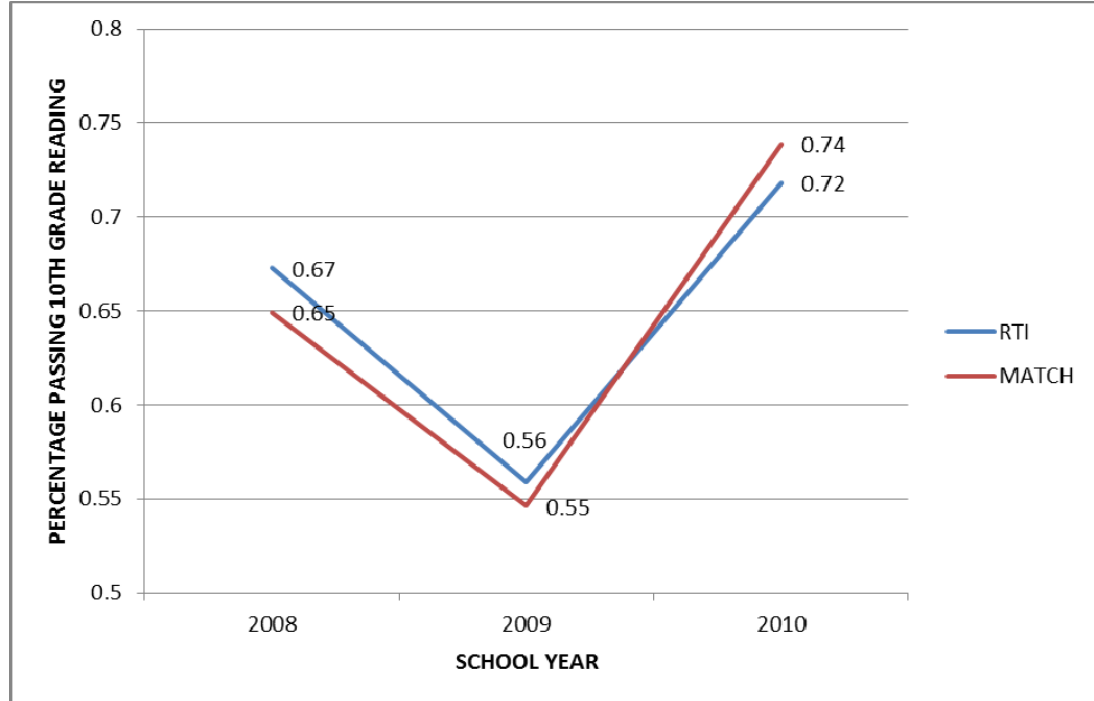
- On a related note, there may be value in highlighting the work of early adopters for purposes of encouraging districts that are taking a “wait and see” attitude and supporting those districts that may be exploring possibilities related to RTI. To the extent that RTI implementation at the higher grades is a work in progress, later adopting districts may benefit from the successes (and failures) of the earlier-implementing group.

## ■ High School Findings (10<sup>th</sup> Grade)

### Tenth Grade Results

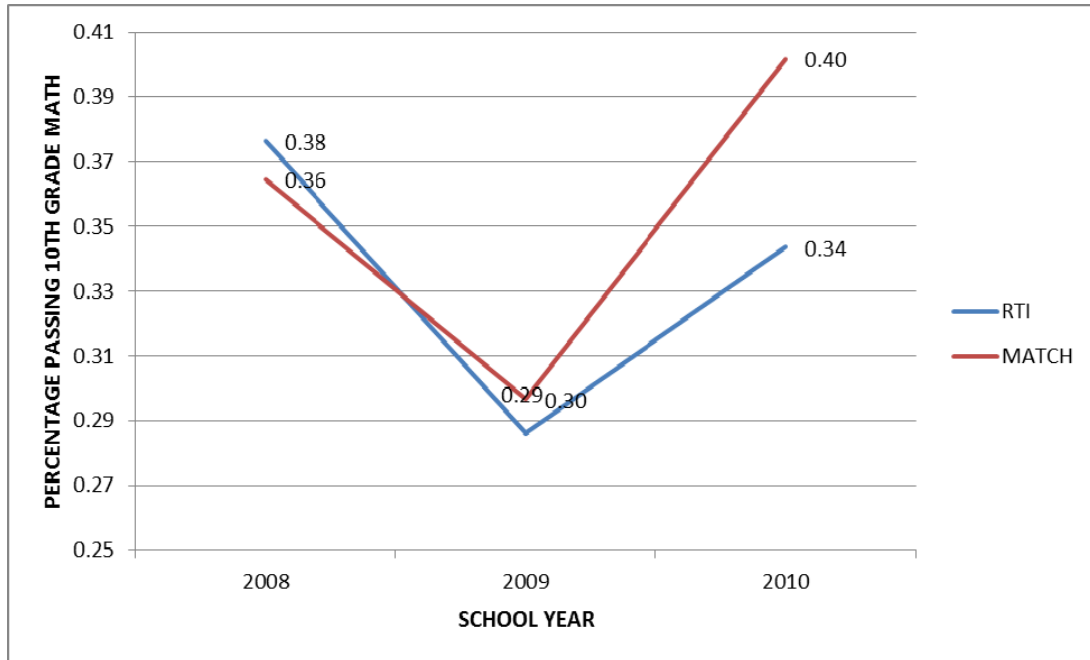
- The trends in 10<sup>th</sup> grade suggest an advantage for comparison districts. In reading (Figure 14), RTI districts improved by about 5 percentage points from 2008 to 2010 compared to an average 9-point gain in the comparison districts. In math, the comparison-group advantage is more evident (see Figure 15).

**Figure 14 – Comparison of Trends in 10<sup>th</sup> Grade Reading**



Note: The WASL was administered in 2008 and 2009. The MSP and HSPE were used in 2010.

**Figure 15 – Comparison of Trends in 10<sup>th</sup> Grade Math**



Note: The WASL was administered in 2008 and 2009. The MSP and HSPE were used in 2010.

### ■ Conclusions For Question 1

- The percent of students scoring at or above proficient in both groups was consistently higher for reading than for math.
- Comparisons across grade levels may be biased to the extent that the WASL or HSPE are “motivated” measures (i.e., scores at some grade levels are used differently than scores at other grade levels to make high-stakes decisions, whether at the student, school, or district levels). Also, as noted, the outcome measure used by Washington OSPI changed in the spring of 2010. The effect of this change is difficult to characterize without evidence of the comparability of the two measures. However, as suggested earlier, the impact on the present analyses is likely to be minimal to the extent that changes in the test are experienced similarly by students in both groups.
- Note also that student achievement is used here as a marker for district-wide performance in the selected grade levels. Different groups of students represent unique cohorts; for example, the group of 2008 3<sup>rd</sup> graders is not the same group of 3<sup>rd</sup> graders tested in 2009. To the extent that any given school, whether RTI or matched, served similar groups of students in 2008, 2009, and 2010, this is not likely to present problems.

However, changes in the communities served by a given school or by a subset of schools would represent a cohort effect, which can confound the interpretation of program effects (i.e., it is a threat to the internal validity of the evaluation design).

- The trends for the RTI schools are *not* disaggregated according to levels of implementation or content areas in which RTI is being implemented. See later sections of this report for these analyses.
- These data are descriptive and should be interpreted in terms of the represented districts only. Attempts to generalize these findings are not encouraged.

## Question 2:

### To what extent have rates of special education referral changed; how do these changes compare to changes in comparison districts?

#### Data Sources

- District-reported numbers of students referred to and receiving special education services in funded and in matched districts.

#### Rationale

- Reduced rates of referral to and enrollment in Special Education are a key outcome in RTI generally and in the evaluation of OSPI's implementation of RTI.

#### Method

- Simple descriptive results are provided. Subsequent years of the evaluation will examine trends and the factors that may moderate the trends.

#### Findings

- Special education referral data were also evaluated by matched pairs. With four exceptions (Chimacum, East Valley, Kelso, Odessa), the number of referrals decreased in all districts, whether participating or matched. There were no apparent patterns related to participation (see Table 3, following page).
- The numbers used in Table 3 include *only* those schools that are implementing RTI (for the RTI districts), and *only* those schools that reported referrals for both years (for RTI and match districts). The SLD Identified column shows the **total number of students** identified as SLD in the implementing districts in November of 2008, 2009, and 2010 (NOT the number of newly identified students in those years.) The SLD Identified numbers are aggregate numbers that mask the magnitude of student mobility, new identifications, and students “exited” from SLD eligibility. It should not be assumed that the numbers from year to year represent the same students, although they do represent the same sample of schools within a given district.

**Table 3 – Number of Special Education Referrals and SLD Students Served in Implementing and Matched Districts in Year 1 and Year 2**

| Status | District                        | Referrals To Special Education |             | SLD Identified* |      |      |
|--------|---------------------------------|--------------------------------|-------------|-----------------|------|------|
|        |                                 | 2008 - 2009                    | 2009 - 2010 | 2008            | 2009 | 2010 |
| RTI    | Odessa                          | 0                              | 3           | 6               | 2    | 2    |
| Match  | Curlew                          | 4                              | 9           |                 |      |      |
| RTI    | Stevenson-Carson                | 20                             | 23          | 32              | 39   | 38   |
| Match  | Chimacum/Aberdeen**             | 27                             | 105         |                 |      |      |
| RTI    | Kiona-Benton City               | 28                             | 14          | 57              | 55   | 39   |
| Match  | White Salmon Valley/Curlew      | 25                             | 9           |                 |      |      |
| RTI    | Hoquiam                         | 45                             | 45          | 41              | 32   | 34   |
| Match  | Aberdeen                        | 125                            | 105         |                 |      |      |
| RTI    | East Valley                     | 37                             | 35          | 32              | 22   | 16   |
| Match  | Cheney/Stanwood-Camano          | NP                             | 82          |                 |      |      |
| RTI    | White River                     | 138                            | 106         | 443             | 470  | 451  |
| Match  | Stanwood-Camano                 | 96                             | 79          |                 |      |      |
| RTI    | Walla Walla                     | 70                             | 73          | 282             | 299  | 290  |
| Match  | Eastmont/Peninsula              | NP                             | 230         |                 |      |      |
| RTI    | Longview                        | 84                             | 1197        | 305             | 320  | 352  |
| Match  | Kelso/Peninsula                 | 114                            | 230         |                 |      |      |
| RTI    | Shoreline                       | 191                            | 171         | 278             | 259  | 256  |
| Match  | Central Kitsap/Peninsula        | 375                            | 230         |                 |      |      |
| RTI    | Richland                        | 103                            | 95          | 253             | 230  | 241  |
| Match  | Peninsula                       | NP                             | 230         |                 |      |      |
| RTI    | Vancouver (Implementing)        | 97                             | 114         | 125             | 129  | 139  |
| Match  | Vancouver (Not Implementing)*** | 181                            | 227         | 595             | 574  | 586  |

NP = not provided

Note: \*SLD Identified numbers were not provided for match districts.

\*\*Where two match districts are listed, the first was the match in Year 1, and the second is the Year 2 match.

\*\*\*Evergreen (Clark) was used as a match for Vancouver in Year 1; however referral data were not provided for Evergreen in Year 1. This table reflects the numbers for implementing schools in Vancouver as the RTI district, and non-implementing schools in Vancouver as the Match district in both Years 1 and 2.

## **Conclusions For Question 2**

Referral to special education is a key outcome in the evaluation. The declining numbers of students being referred is positive news whatever the cause. In this context, the lack of positive effect is not discouraging. In subsequent project years, when multiple data points are available, district-level trends can be estimated and compared for purposes of evaluating RTT's effect on this key outcome.

### **Question 3:**

**To what extent has implementation of the School-wide Activities in Figure 1 changed from 2008-2010; to what extent are differences in the implementation of School-wide Activities associated with achievement trends?**

#### **Data Sources**

- Survey data from years 1 and 2 of the evaluation.
- Site visit data from years 1 and 2 of the evaluation.
- Achievement trend data from 2008, 2009, and 2010.

#### **Rationale**

- The theory is that program implementation is a process that can be described and to an extent measured, and that greater levels of implementation are associated with increases in key outcomes. To the extent that this theory holds, level of implementation represents a reasonable indicator of RTT's ongoing potential to influence student achievement in participating districts. It provides a rationale for continuing to monitor and measure levels of implementation, as a means of interpreting evaluation results and as a basis for intervening in districts where implementation is less sophisticated than desired. However, this is reasonable only to the extent that the implementation/outcomes theory can be demonstrated.

#### **Method**

- The effect of implementation on achievement is ideally evaluated using growth modeling. Growth models represent change over time in terms of a linear function, a *variation* on the familiar  $y = mx + b$  (from high school algebra). In this case, 3-year trends would be estimated for each district at each of the different grade levels. However, the 2010 change in high-stakes tests and the possibility that changed measurement is confounded with changes in the underlying construct (i.e., being measured made growth modeling a less viable analytic option). Instead, an “extreme groups” method was used, where trends over are compared for the highest and lowest implementing districts. This approach also addresses the small sample size and the relatively limited rigor of the measure of implementation. It is a descriptive analysis, intended as a general indication of achievement trends related to implementation.

- Level of implementation level was calculated by standardizing responses (the mean of all responses divided by the standard deviation) to the survey item:

“What level of implementation below would you say applies to **your** use of RTI in the classroom in which you work (select one)?

0 None/I don’t know

1 I am currently learning about RTI and thinking about how it may fit into my teaching

2 I am preparing for the implementation of RTI, working on a plan for implementing, and/or figuring out what I need to put in place to begin implementation

3 I have a few pieces of RTI in place, and a plan for moving forward with the rest of implementation

4 An RTI framework is up and running in my classroom

5 I am looking for ways to improve RTI in my class, all the pieces are in place and being used, but I think I can make it even better by tweaking some aspects to be more relevant to my classroom

6 RTI is operational, innovations are applied in my classroom, and systems are in place to ensure that what we are doing is sustainable and continues to work well”

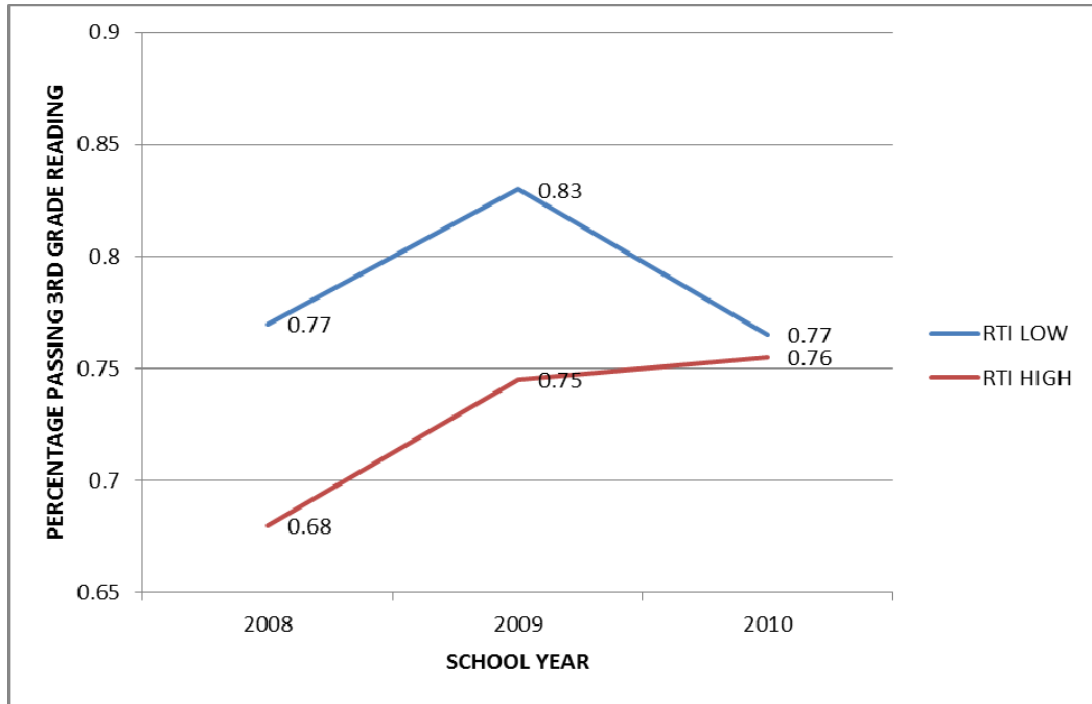
- Detailed summaries of levels of implementation for each district are available in the Appendix.
- Districts were compared on implementation status and the 2 to 3 highest and 2 to 3 lowest districts were identified. In some cases, the degree of separation (between high implementers and low implementers) was considerable. In other instances, the relative differences were less significant.

## ■ Elementary School Findings (Grades 3-5)

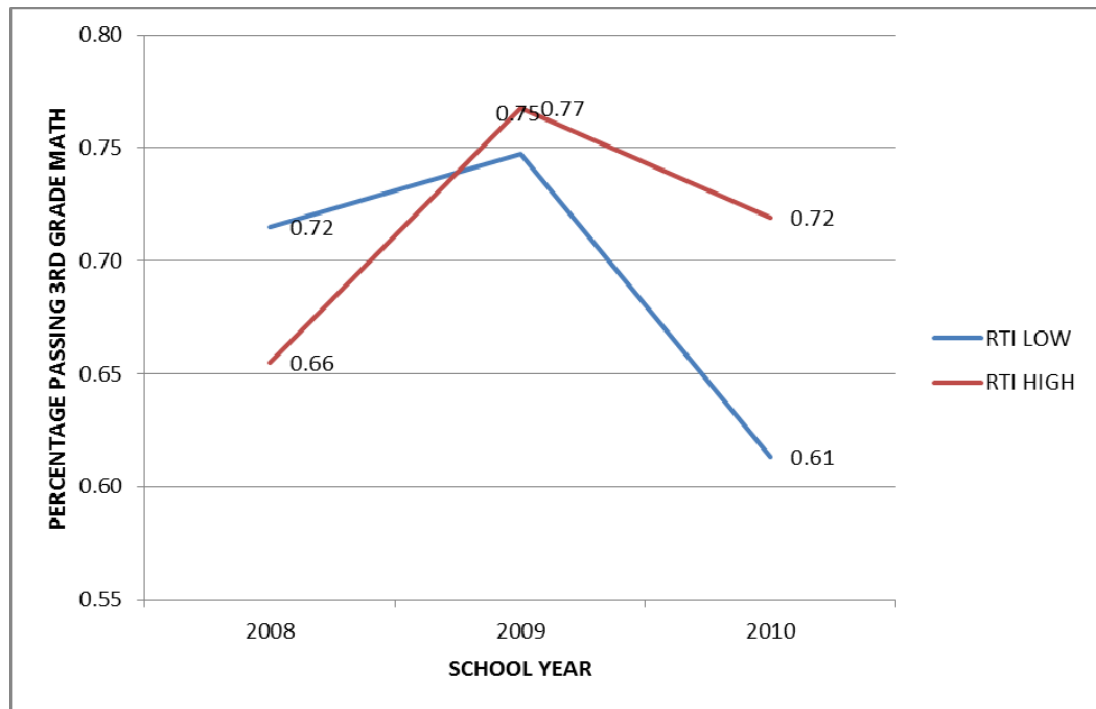
### Third Grade Results

- School 1 (2.0, n = 3) and School 2 (3.2, n = 65) 3<sup>rd</sup> grade educators had the lowest average scores on the implementation item. School 3 (5.7, n = 3) and School 4 (5.1, n = 14) were high scorers on the implementation item.
- In reading, high implementing districts, on average, improved the percentage of proficient students from 68% in 2008 to 76% in 2010. Low implementers reported similar levels of passing in 2008 and 2010 (about 77% proficient readers).

**Figure 15 – 3<sup>rd</sup> Grade Reading Differences and Levels of Implementation**



**Figure 16 – 3<sup>rd</sup> Grade Math Differences and Levels of Implementation**

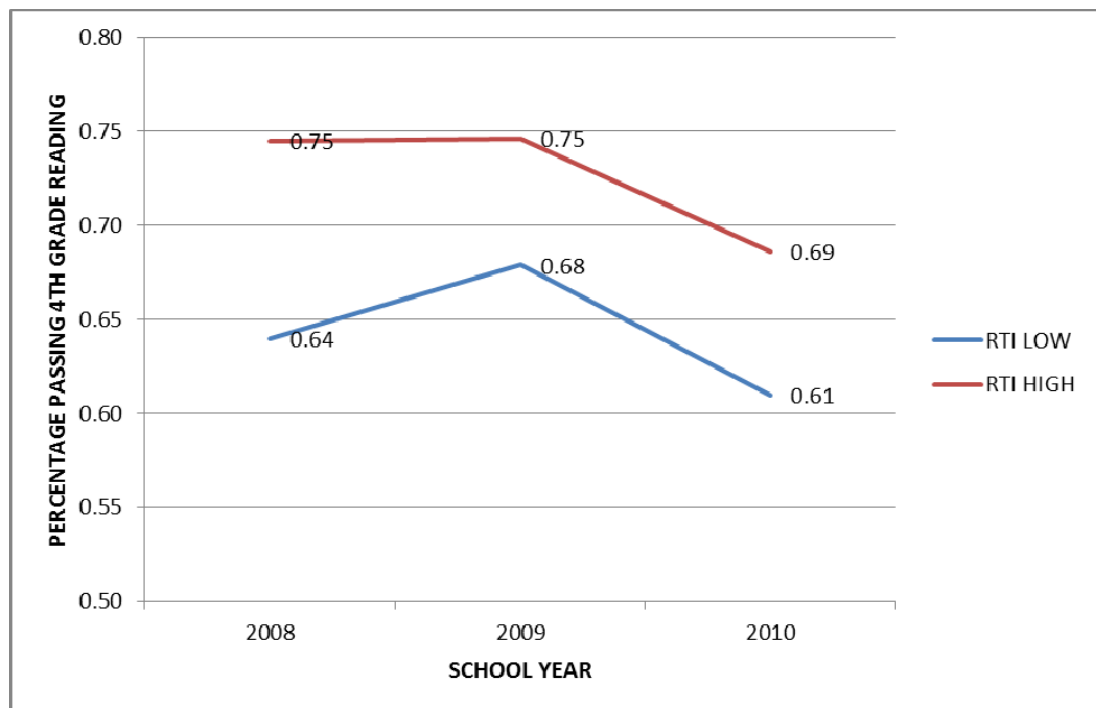


- In 3<sup>rd</sup> grade math, high implementing districts improved from 66% proficient in 2008 to 72% proficient in 2010, on average. Average performance in low implementing districts decreased from 72% passing to 61% passing.

#### Fourth Grade Results

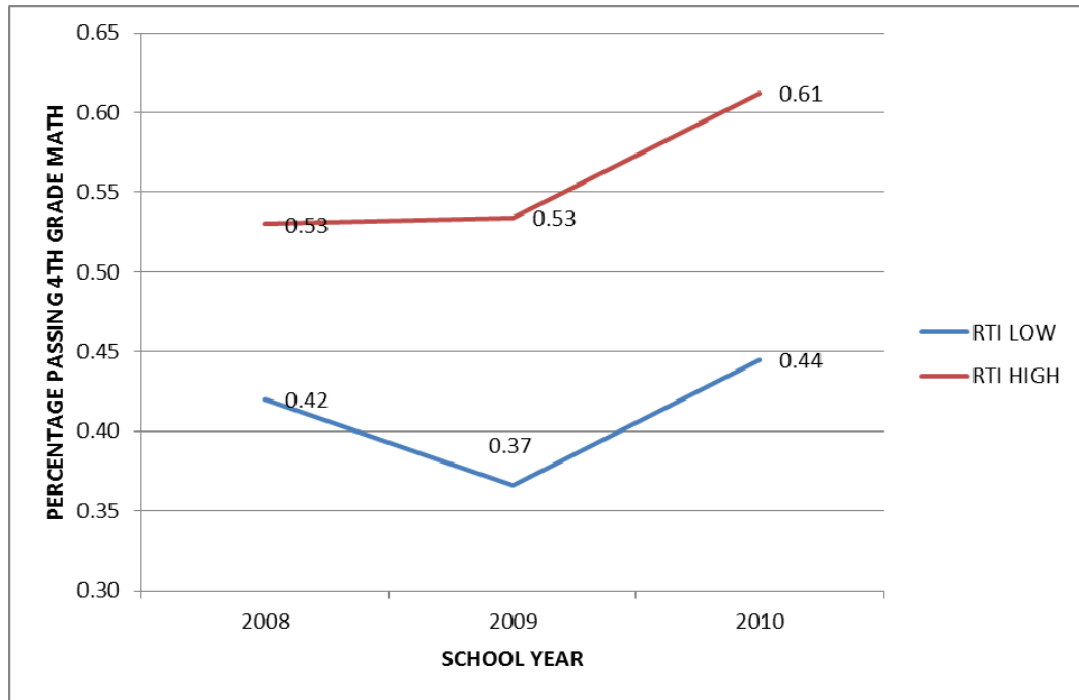
- School 4 (5.3, n = 14) and School 3 (5.7, n = 3) were higher implementers in 4<sup>th</sup> grade. School 5 (3.2, n = 13) and School 2 (3.4, n = 54) reported lower levels of implementation, on average.
- Both groups experienced a decrease in the percentage of students scoring at or above proficient across the 3 years. The group of lower implementing districts experienced less of a loss in this respect.

**Figure 17 – 4<sup>th</sup> Grade Reading Differences and Levels of Implementation**



- In 4<sup>th</sup> grade math, the higher implementing districts outpaced the lower implementing districts, improving by 6 percentage points over the 3 year period, compared to 2% in the lower implementing group.

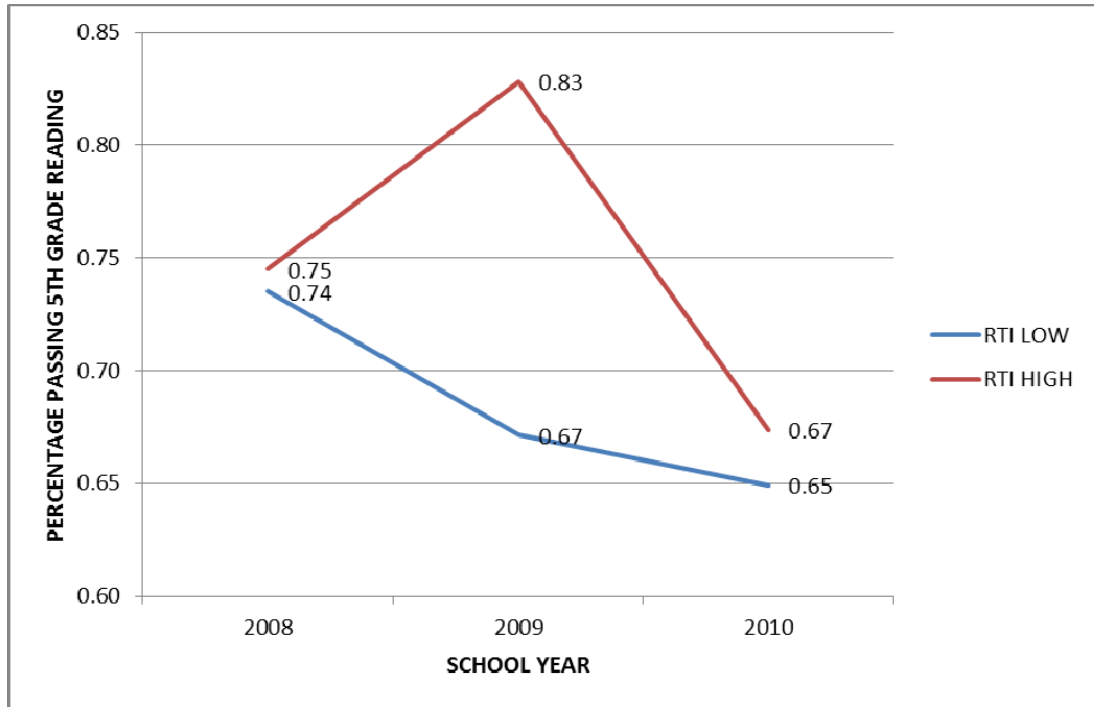
**Figure 18 – 4<sup>th</sup> Grade Math Differences and Levels of Implementation**



### **Fifth Grade Results**

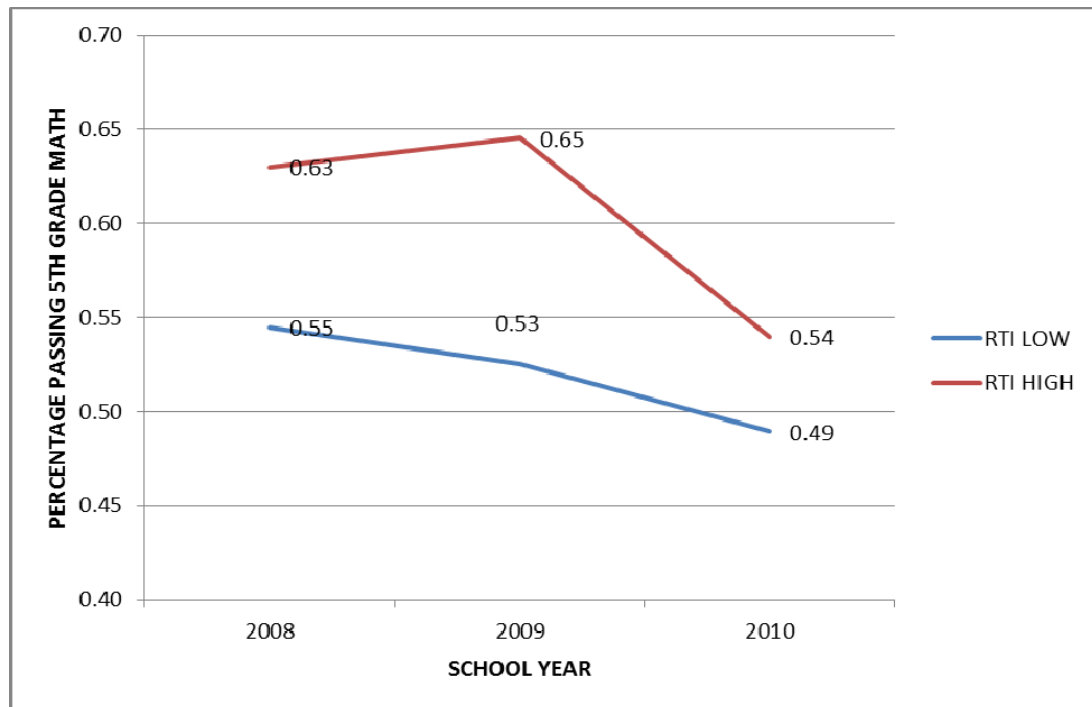
- School 4 (5.4, n = 14) and School 3 (5.7, n = 3) reported higher levels of implementation. School 5 (3.3, n = 18) and School 2 (3.2, n = 65) reported lower implementation, on average.
- In reading, both groups decreased in the percentage of proficient readers, though the higher implementing group experienced a peak in 2009. The overall loss in the group of higher implementing districts was slightly smaller (8 percentage points compared to 9 points in the lower implementing districts).

**Figure 19 – 5<sup>th</sup> Grade Reading Differences and Levels of Implementation**



- Like reading, both groups experienced a loss in the percentage of proficient students from 2008 to 2010. Additionally, the higher implementing districts increased from 2008 to 2009 before declining from 2009 to 2010, though the 2008 spike in math was not as dramatic as in reading.

**Figure 20 – 5<sup>th</sup> Grade Math Differences and Levels of Implementation**



### Summary of Elementary School Findings

- Perhaps the most notable finding is that high implementing districts have a considerably higher percentage of proficient students in 2008. To the extent that implementation is a process, it seems reasonable that higher implementers were earlier implementers, and the relatively consistent advantages in 2008 reflect their greater use of RTI in 2008 and prior.
- The exception, which is notable, is 3<sup>rd</sup> grade reading and math, where the low implementers have an advantage in 2008, although they are overtaken by the group of high implementing districts over the 3-year period ending in 2010.
- The earlier comments about the 2010 changes in measurement apply to these analyses, as well.
- It is also worth noting that district size is not controlled for in these comparisons. For example, in 5<sup>th</sup> grade School 4 and School 3 are relatively small districts compared to

School 2 and School 5. Implementation may differ in distinct ways depending on district size (and probably does given the fewer number of schools in the small districts).

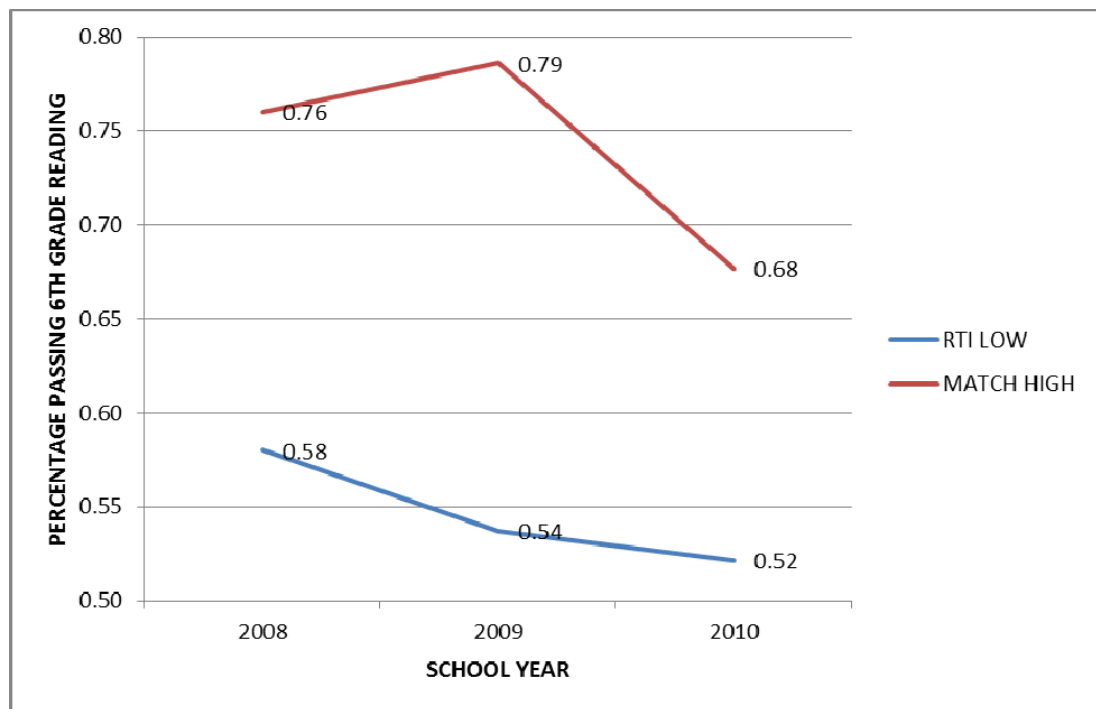
- Measurement of implementation is based on the single self-report survey item, although school site visit data were used as a supporting evidence base. This is not ideal, though the relatively predictable pattern of findings is encouraging. Future evaluation in OSPI should consider more comprehensive, rigorous, and multi-faceted measures of implementation.

## ■ Middle School Findings (Grades 6-8)

### Sixth Grade Results

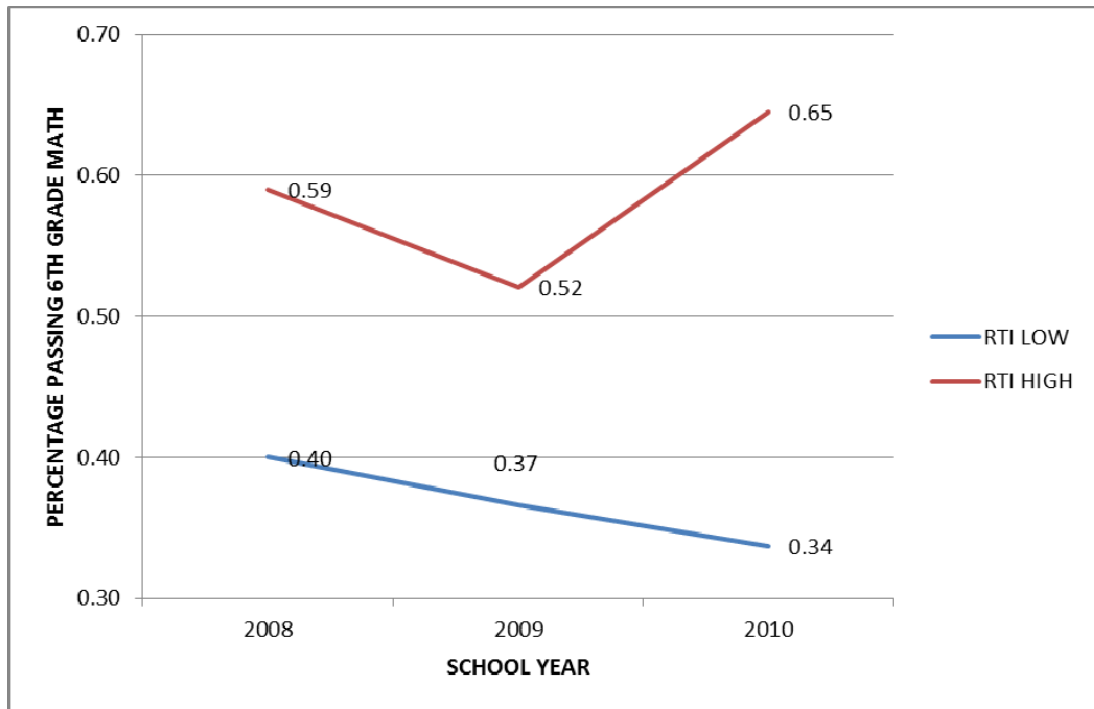
- The higher implementing districts in 6<sup>th</sup> grade were School 3 (6.0, n = 4) and School 6 (4.5, n = 26). School 7 (2.0, n = 5) and School 8 (2.5, n = 4) reported lower levels of implementation.

**Figure 21 – 6<sup>th</sup> Grade Reading Differences and Levels of Implementation**



- Both groups experienced a decline overall. However, as before, the group of higher implementers experienced a small upward bump prior to the 2010 change in OSPI's high stakes test.

**Figure 22 – 6<sup>th</sup> Grade Math Differences and Levels of Implementation**

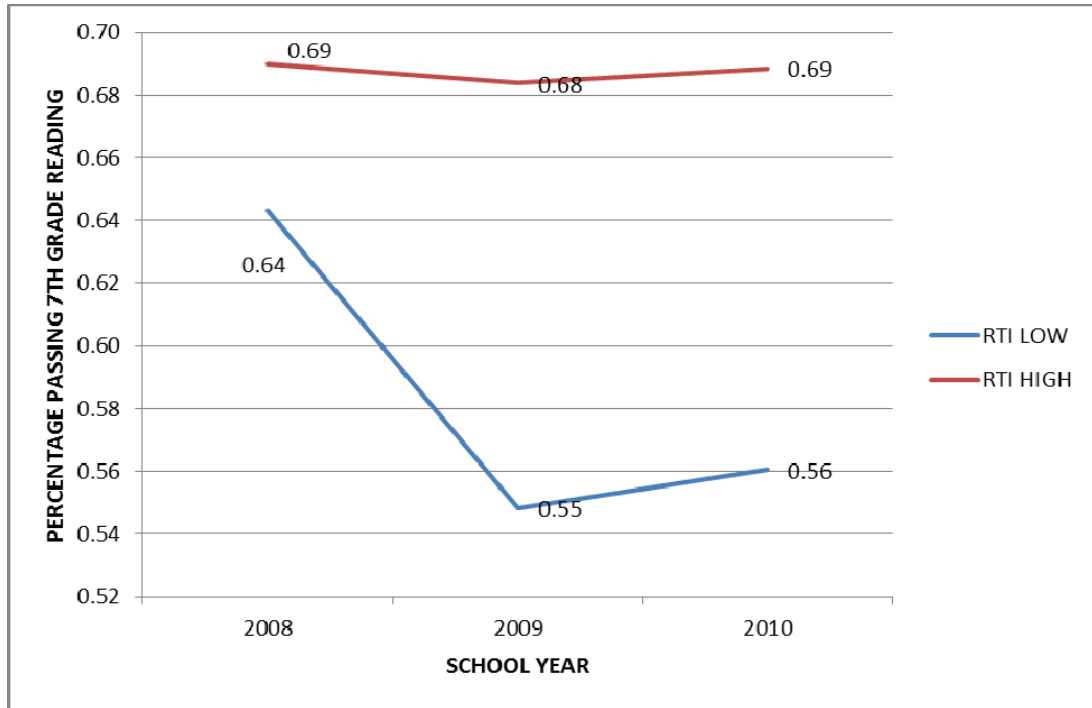


- The 6<sup>th</sup> grade math results differ from reading for the high implementing group, which increased over time after dipping slightly in 2009. The trend in the low implementing math group was similar to the reading group. Both experienced a consistent downward trend across the 3 years.

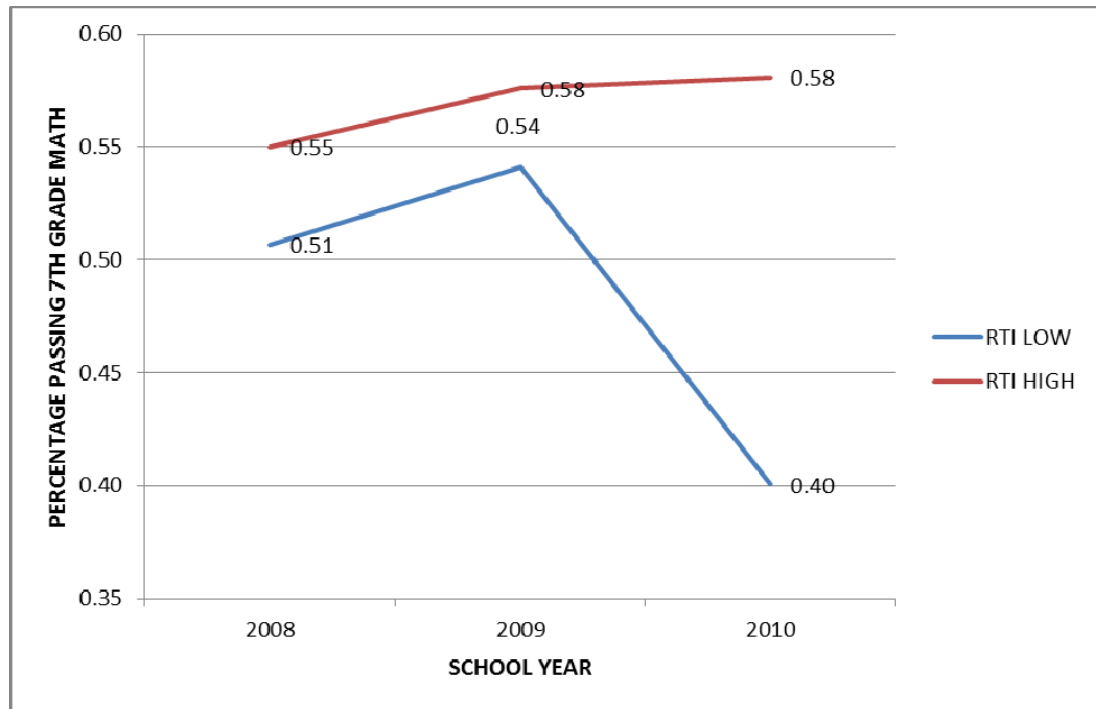
### **Seventh Grade Results**

- High implementers in 7<sup>th</sup> grade were School 4 (4.5, n = 25) and School 6 (4.2, n = 22). School 3 (2.0, n = 9), School 7 (2.0, n = 5), and School 9 (1.5, n = 8) reported lower levels of implementation.
- In reading, the high implementing group started higher, on average, than the lower implementing group and remained relatively consistent in the percentage of students scoring at or above proficient. The lower implementing group experienced a considerable decline over the 3-year period, from 64% proficient in 2008 to 56% proficient in 2010.

**Figure 23 – 7<sup>th</sup> Grade Reading Differences and Levels of Implementation**



**Figure 24 – 7<sup>th</sup> Grade Math Differences and Levels of Implementation**

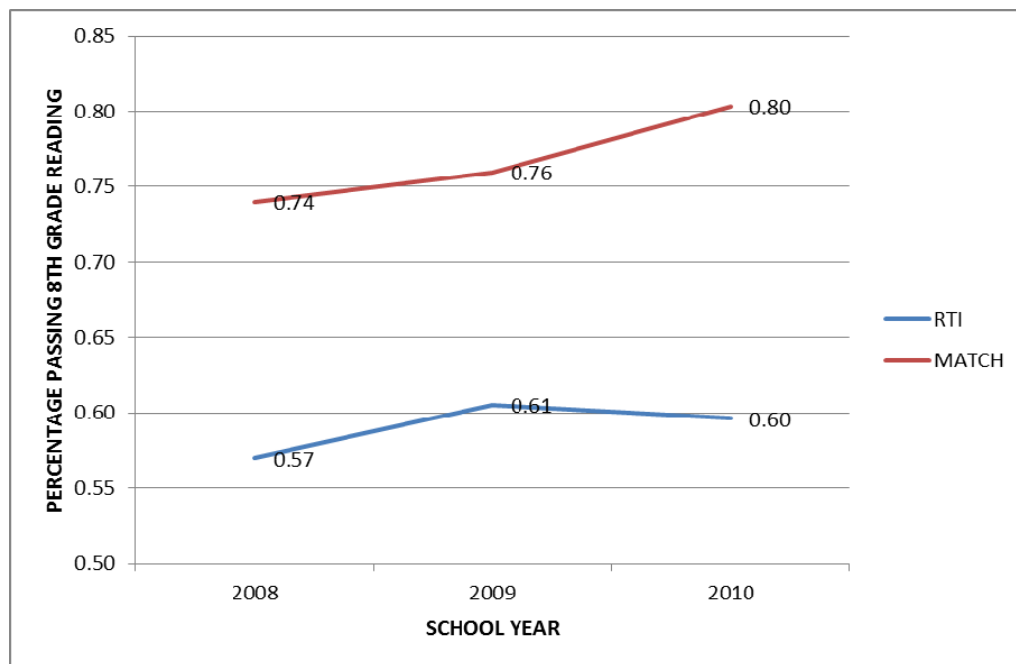


- Seventh-grade math trends were similar to reading, though more favorable to the RTI high implementing group. In this case, the higher implementing districts improved the percentage of proficient students over the 3-year period, while low implementers had a significant decrease, from 51% in 2008 to only 40% proficient in 2010.

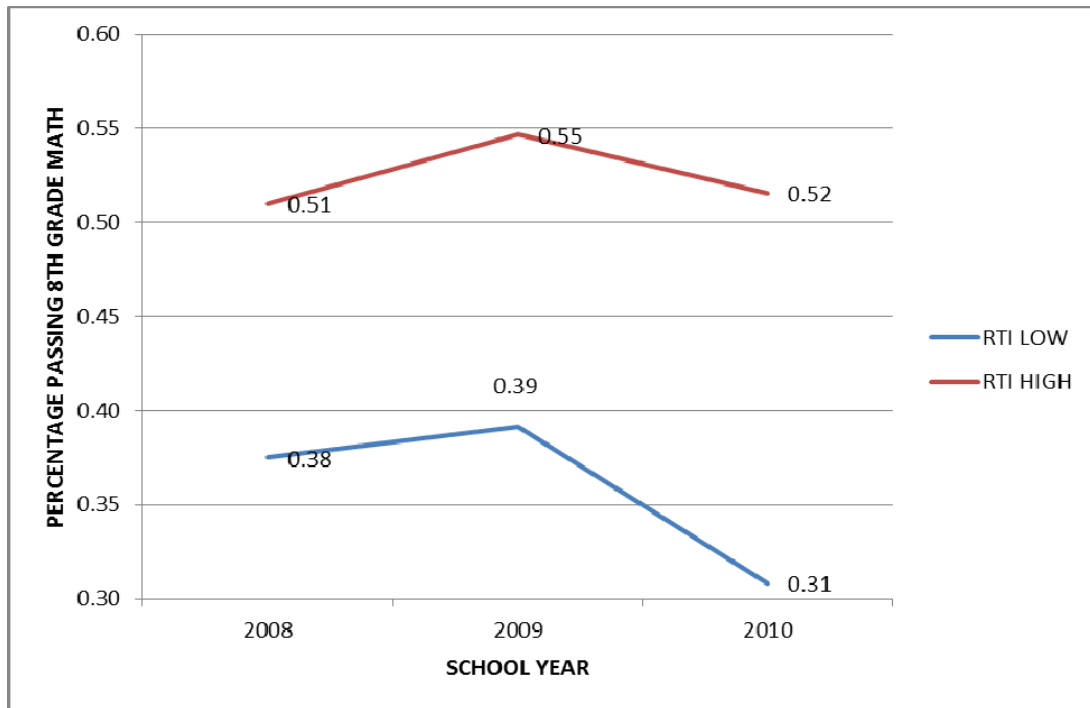
### Eighth Grade Results

- The high implementing districts in 7<sup>th</sup> grade were the same in 8<sup>th</sup> grade. School 4 (4.2, n = 25) and School 6 (4.2, n = 21). Similarly, 2 of the 3 lower implementers in 7<sup>th</sup> grade were also low in 8<sup>th</sup> grade – School 3 (2.1, n = 8) and School 9 (1.7, n = 7).
- Both groups made progress in reading over the 3-year period. High implementers performed better in 2008 (74%) than lower implementers (57%) and made greater gains over the 3-year interval (6% increase compared to a 3% increase).
- In math, the high implementing group also outperformed the group of lower implementing districts in 2008 (51% to 38%). However, unlike 8<sup>th</sup> grade reading, the high implementing group was the only one of the two groups to experience improved (though marginal) average performance over time (from 51% to 52%). In the lower implementing districts, only 31% of 8<sup>th</sup> grade students were proficient in math in 2010.

**Figure 25 – 8<sup>th</sup> Grade Reading Differences and Levels of Implementation**



**Figure 26 – 8<sup>th</sup> Grade Math Differences and Levels of Implementation**



### **Summary of Middle School Findings**

- The middle school trends are similar to the elementary grade results in the 2008 differences in the large to very large differences in percent of proficient students in 2008, due perhaps to differences in the timing of adoption and first implementation of RTI.
- School 3 was a high implementing district in the elementary grades and a relatively low implementer at the higher grade levels. This may be by design or it may be a result of earlier funding (i.e., Reading First). Either way, it is interesting and may be worth following up in year 3.

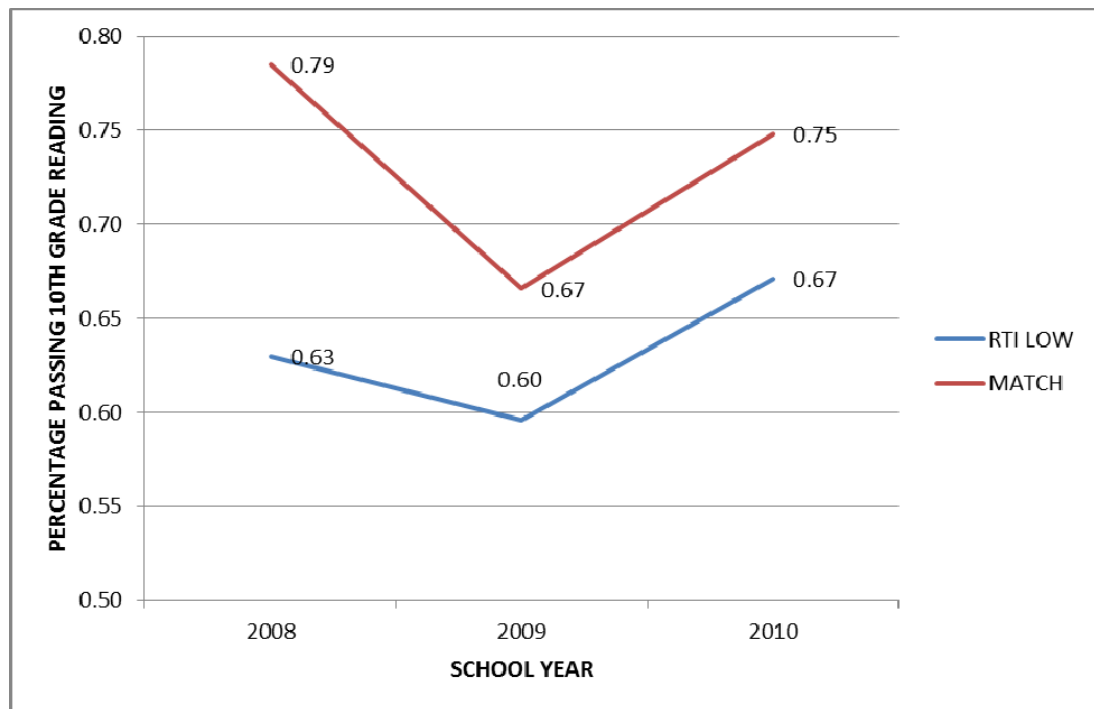
### **High School Findings (10<sup>th</sup> Grade)**

#### **Tenth Grade Results**

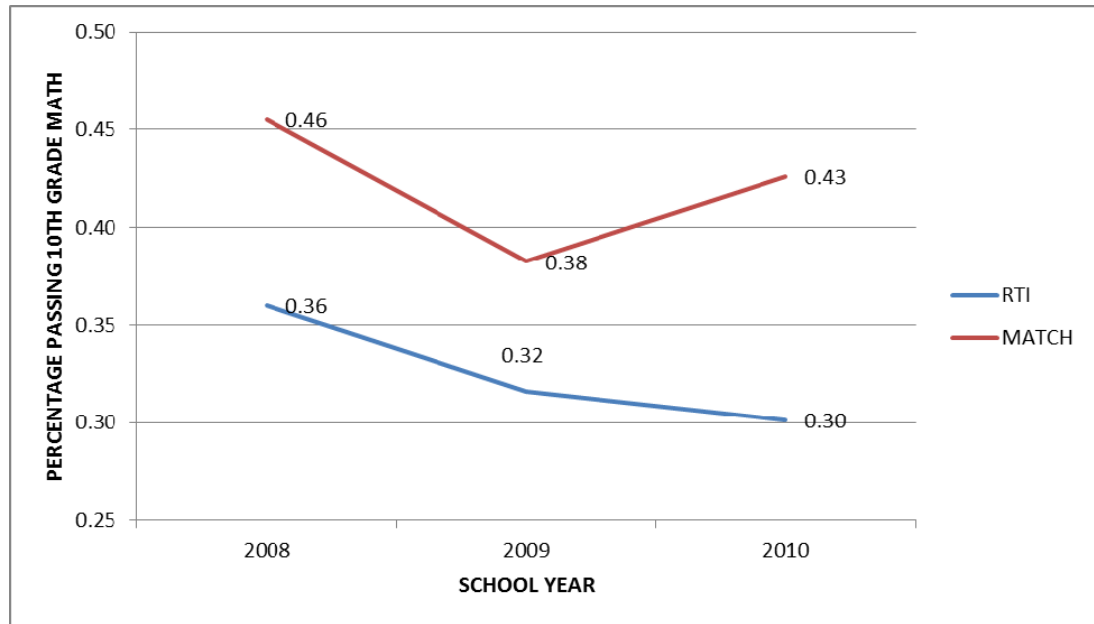
- School 4 (3.5, n = 24) was again a high implementing district in 10<sup>th</sup> grade. School 10 (3.6, n = 34) also reported high levels of implementation in 10<sup>th</sup> grade. School 2 (2.2, n = 160), School 3 (2.1, n = 13), School 9 (1.8, n = 20), and School 7 (1.8, n = 8) reported lower levels of implementation.

- Reading and math trends are represented in Figures 27 and 28. In both cases, 10<sup>th</sup> grade students attending the RTI high implementing districts perform higher in 2008 than students in the lower implementing districts. However, in reading students in the lower implementing districts scored higher, on average, in 2010 than in 2008 (63% to 67%) while 10<sup>th</sup> graders in the higher implementing districts dipped from 79% to 75% over the 3-year period. In math, both groups experienced a 3-year decrease in the percentage of proficient students, with the lower-implementing districts (from 36% to 30%) experiencing a somewhat larger drop than 10<sup>th</sup> graders in the higher implementing districts (46% to 43%).

**Figure 27 – 10<sup>th</sup> Grade Reading Differences and Levels of Implementation**



**Figure 28 – 10<sup>th</sup> Grade Math Differences and Levels of Implementation**



■ **Conclusions For Question 3**

- Taken together, these data suggest that implementation, as measured by the single survey item and augmented by relevant site visit data, covaries with achievement across the three-year period from 2008-2010 and perhaps prior to 2008, particularly in the higher grades where more variation in implementation is to be expected. This supports the notion that implementation relates to improvement and that implementation can be indicated using relatively simple methodologies.
- There are also several counterintuitive results, notably the 3<sup>rd</sup> grade reading and math results. Future evaluation may help to clarify these trends.
- A more comprehensive and thorough-going measure of implementation may also be useful in this respect. The framework being used is aligned with the SISEP and National RTI Centers. More sensitive measurement of the framework's elements may be possible.

## Recommendations for Ongoing Practice and Evaluation

---

- State-level officials at OSPI are establishing a national standard for using evidence and data to make program-related decisions, as indicated by their ongoing collaboration with the National RTI Center and the Center on State Implementation and Scaling-up of Evidence-based Practices and by their continuing commitment to program evaluation, formative and summative. Their appreciation for the “science of implementation” and their increasingly sophisticated use of related data in planning and monitoring state-wide efforts are worth noting. OSPI officials and members of the several state-wide systems of support may also represent a useful model for education departments in states of similar size that may be considering adoption or implementation of tiered instructional frameworks.
- The data at middle and secondary schools suggest that students in high implementing districts outperform students in lower implementing districts. While these results should be considered preliminary (and not a basis for claims that RTI *causes* higher achievement), their remarkable consistency across grade levels and content areas provides a degree of support for RTI’s potential and for the use of *level of implementation* as a proximal indicator of progress. Also, to the extent that implementation of a sophisticated model like RTI is multiyear process, it is likely that the high implementing districts have been implementing RTI or RTI-aligned practices for a period of time that predates this evaluation, accounting for the consistently higher 2008 performance by students in the high implementing districts.
- Given the potential utility of the *level of implementation* construct, there may be value in adopting a more rigorous and comprehensive approach to its measurement, using tools and data elements that are embedded in ongoing state- and district-level systems and operations. State-level monitoring and technical assistance will be increasingly efficient to the extent that already existing or easily implemented data elements can be validated as indicators of implementation.
- The change in state-wide measurement (from the Washington Assessment of Student Learning to the [Measurements of Student Progress and](#) the [High School Proficiency Exam](#)) hinders the estimation of reliable achievement trends which limits analytic possibilities. The measures may be perfectly equated (though the descriptive trends seem not to support this possibility); however, comparison across the old and new measures may be misleading to the

extent that the two sets of measures vary in difficulty, metric, or in the constructs being measured. Equating the measures (or documenting that they are indeed equivalent) will increase the reliability of trend estimates and allow consideration of the effect of student-, school-, and district-level variables on progress over time.

- We continue to be concerned with the reliability of the behavior-related data. Our proposal is to identify a handful of meaningful data elements in the behavioral domain that can be reliably collected as part of the program evaluation protocols (the survey or the site visits). While this approach has its own set of limitations, the externally-collected data will provide a basis for evaluating the relative reliability of the district-reported data (i.e., if the evaluator-collected data are comparably variable and unsystematic).