WISCONSIN VALUE-ADDED MODEL: A DEMONSTRATION PROJECT FINAL REPORT

Robert H. Meyer
Emin Dokumaci
Elisabeth L. Geraghty
Sean McLaughlin
Ernest A. Morgan

Value-Added Research Center
Wisconsin Center for Education Research
University of Wisconsin-Madison
1025 W. Johnson St., Madison, WI 53706

May 2009

This paper was developed as part of a U.S. Department of Education sub-grant on Longitudinal Data Systems through the Wisconsin Department of Public Instruction, Office of Educational Accountability.

Introduction

This document serves as overview and final report for the *Value-Added and Growth Model Demonstration Project*. This project was developed as part of a U.S. Department of Education sub-grant on Longitudinal Data Systems through the Wisconsin Department of Public Instruction, Office of Educational Accountability. In the initial section the document describes cooperative efforts between the Wisconsin Department of Public Instruction (DPI) and the Value-Added Research Center (VARC) at the University of Wisconsin-Madison; the project's partners, and the project's deliverables. The second section provides a technical analysis of the statewide value-added results. The final section suggests some possible next steps for building a statewide value-added system in Wisconsin.

I. Project Summary

1.1 Project Timeline

The project was initiated in December 2007 via planning meetings between DPI and VARC staff. Along with establishing a framework for the project, these December meetings identified potential partner districts and Cooperative Educational Services Agencies (CESAs).

In January 2008, VARC hosted a meeting of project partners including, DPI, Milwaukee Public Schools (MPS), Madison Metropolitan School District (MMSD), Waunakee Community School District, and the Cooperative Educational Services Agency #2 (CESA #2). At this meeting VARC staff conducted a focused discussion on the possible needs of districts and CESAs for a statewide value-added system.

In February 2008, DPI and VARC established a formal data release agreement for statewide assessment test results data needed to complete the project. The November 2005 data and November 2006 data were provided by DPI to VARC in mid-February. The November 2007 data was provided in March 2008. From February through May, DPI and VARC staff worked to ensure the quality of the data. At the same time, VARC researchers worked on the development of the statewide value-added model.

Over the summer of 2008, VARC staff worked on a paper specifically designed to provide policy makers with an understanding of value-added analysis. DPI provided formal consultations and input regarding the paper. While work on this paper was progressing VARC researchers continued developing and testing a statewide value-added model.

Beginning in September, VARC staff, with formal consultations and input from DPI, began to develop report templates geared toward districts, schools and parents. The aim was to develop reports that provided sufficient and consumable information, without overwhelming consumers of the information. The report development work continued through December 2008.

During October and November, VARC researchers worked on application of the statewide value-added model to the district and school levels. In December, a statewide model was fully developed. Work continued on fine-tuning this statewide model through January 2009.

With the statewide value-added demonstration model complete, VARC staff worked to complete the task of populating sample reports for districts, schools and parents, and to provide an analysis of the results from the statewide value-added demonstration model. This analysis is in Section II of this overview, *Review and Interpretation of Statewide Value-Added Results*. Along with the finalization of the demonstration model, VARC researchers completed the technical paper detailing the specifications of the statewide demonstration model.

1.2 Project Partners

Milwaukee Public Schools (MPS) – MPS is a large urban district with 213 schools and over 85,000 students. MPS has been reporting value-added data at the school and grade level for seven years.

Madison Metropolitan School District (MMSD) – MMSD is a medium sized and very diverse district. The district enrolls over 24,000 students in 55 schools. MMSD began working with value-added data during the 2007-08 school year.

Waunakee Community School District – Waunakee is a smaller, high performing district located outside of Madison. The district enrolls approximately 3,500 students in 6 schools.

Cooperative Educational Services Agency #2 (CESA #2) – CESA #2 has over 70 districts within its seven counties. This includes more than 135,000 students and 7,200 teachers. CESA #2 provides leadership and coordination of services for school districts, including curriculum development assistance and data collection processing and dissemination.

The Wisconsin Department of Public Instruction (DPI) – DPI is the state agency that advances public education and libraries in Wisconsin. Within DPI is the Office of Educational Accountability, the team responsible for meeting the assessment and accountability requirements of state and federal law.

1.3 Project Deliverables

The Value-Added Growth Model Demonstration Project contract outlined three deliverables.

Deliverable 1 – Planning Meetings
VARC provided meeting content, an agenda, and discussion questions to DPI
VARC provided summary report of district discussion

In January 2008 the Wisconsin Center for Education Research (WCER) hosted a planning meeting that included the project partners involved in the demonstration project. Participants included CESA #2, MMSD, MPS, Waunakee Community School District, and DPI. Participants were asked to participate in guided discussions about the use of value-added data, features of a value-added model, value-added reports, and district professional development and dissemination. Detailed notes were taken during the discussion. Following the discussion, WCER summarized the notes for DPI and provided general responses to each focus question as well as district specific responses to each question. Responses to the questions generally fell into the three following categories:

1. What is the value-added productivity of schools within each district?
2. How do district value-added indicator systems best make use of state-level data?

3a. Can state-level value-added data be used for program evaluation?
3b. Can state-level value-added data be used to identify the most effective district and state programs and policies?

Deliverable 2 – Value-Added Paper and Model Specifications
VARC provided sections 2a and 2b

Deliverable 2 is divided into two sections.

Section 2a provides a plain-text explanation of value-added, including a general description of growth models and specifics about the VARC value-added model. This description includes the controlling factors that are often used in a VARC value-added model as well as possible comparisons that can be made and possible applications. We also include a discussion of possible areas of caution, such as the need to be aware of confidence intervals and the need to not use the data to evaluate specific teachers. Finally, some additional resources are provided, including descriptions of variations of value-added models, a step-by-step guide through a simple example of adjusted growth, and a value-added dictionary.

Section 2b includes a discussion of alternative value-added models used by VARC to perform value-added analysis on statewide data.

Deliverable 3 – Results of Demonstration Project and Report Models

VARC provided the following reports:

Statewide Analysis:
- Executive Summary
- Review and Interpretation of Statewide Value-Added Results

School/District Reports
- District Compared to the State 2005-06, 2006-07
- School Compared to the State 2005-06, 2006-07
- Individual Growth 2005-06, 2006-07

This final deliverable includes a presentation of results in the form of a series of interpretive reports. The reports include presentations of an individual student's longitudinal growth that a parent or teacher could use, school level reports showing value-added data for tested grades, and state-level reports showing how districts and schools compare to the state average in value-added. School and district reports are presented in a tabular and graphical format. Each report provides several years of data so that some trend analysis can be done. In addition, we provide a scatter plot comparing value-added and attainment. This report shows, for example, how third graders statewide performed in math or reading compared to other third graders. Specific schools are highlighted so that comparisons and trend analysis can be conducted.

II. Review and Interpretation of Statewide Value-Added Results

2.1 Executive Summary

Purpose and Findings

The purpose of this project is to address questions that cannot be answered in separate district level value-added systems. From a policy perspective, one might want to look at high and low performing schools and districts across the state and try to draw conclusions about what factors into the success of certain schools and districts. A statewide system allows for an analysis of the variability of performance indicators in a given district as it relates to the state. A good start might be to identify districts with low or high performing schools and start asking questions about what these districts might be doing that is contributing to their success or failure.

Why is a statewide system important? This statewide study has found that it is possible to determine whether there is a difference between the average productivity of districts across the state. Productivity is defined as the contribution of a school or district to a student's growth in scale score points on the WKCE as measured by value-added analysis. This study finds that there is a difference between the average productivity of the three example districts (Madison, Milwaukee, and Waunakee). Madison is more consistent in school productivity than Milwaukee in the sense that the school/district productivity parameters of Madison's schools are more densely populated around the district average than that of Milwaukee's. On the other hand, despite the fact that average productivity is higher in Madison than in Milwaukee, the highest performing Milwaukee schools tend to have higher productivity than the highest performing Madison schools (see Figures 2.9.1 and 2.9.2).

A distinctive characteristic of a statewide value-added model versus a district value-added model is that the statewide model can identify not only the high and low performing schools, but also high and low performing districts. A statewide model also allows for the placement of schools in the context of statewide performance rather than only district performance. Identifying high and low performing districts and schools consists of two parts; (i) district average of school/district productivity and (ii) productivity ranges (variance) of schools within the district. Furthermore, a statewide value-added model can answer whether districts differ systematically (statewide variance of district productivity) in their capacity to create high and low value-added schools. First in Section 2.2, we develop a state value-added model that addresses these characteristics. Later in Section 2.8 and Section 2.9, we present the model estimates for the Milwaukee, Madison and Waunakee school districts and also the statewide variance of district productivity. These estimates highlight the three distinctive characteristics of the statewide value-added model – school/district productivity, variance of schools within a district, and statewide variance of district productivity. Finally, templates for suggested district, school and individual reports are presented in Appendix C.

Next steps

At the end of Section 2.10, we discuss both technical and logistical next steps. While the technical next steps involve possible alternative methods for grouping districts, the logistical next steps focus on a pilot plan jointly developed by CESA#2 and VARC. This plan provides an operational structure for providing value-added metrics and professional development to participating districts in a method appropriate for addressing the three categories of responses identified by participants at the January 2008 planning meeting – see the bottom of page 3.

As the pilot plan is implemented additional questions about the application of a statewide value-added model may arise. Questions might include whether the calculation of adequate yearly progress or the identification of New Wisconsin Promise reward schools is an appropriate application of value-added estimates. In addition, could value-added estimates be used to evaluate the effectiveness of a certain math or reading program used by many districts throughout the state? Value-added analysis provides many exciting opportunities for educators and policymakers to learn more about how students are learning. It will be interesting to see what further collaboration with state, district, and local educators will provide to the discussion.

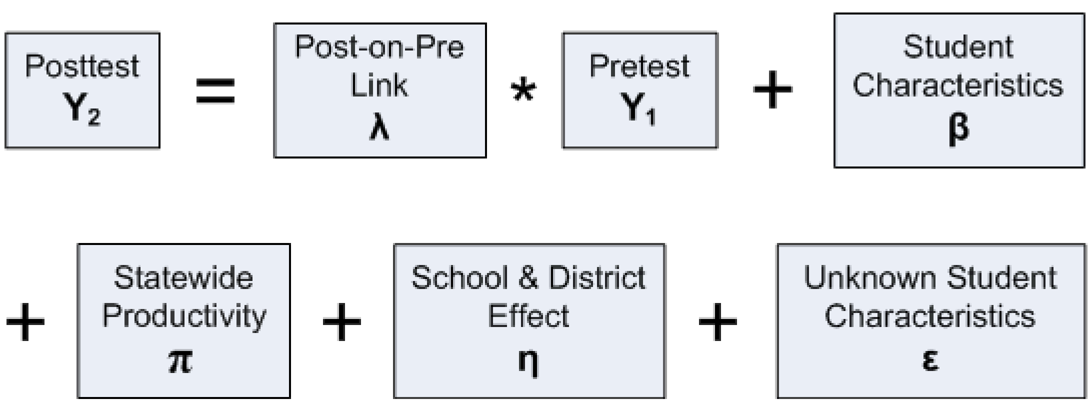
2.2 Simple Presentation of Value-Added Model

In this section we discuss the interpretation and use of results from a statewide value-added indicator system. Separate models are developed for mathematics and reading achievement. We highlight the questions that can uniquely be addressed in a statewide value-added system and that cannot be addressed in separate district value-added systems. Additional technical information on the value-added models developed and used during this project is included in Appendix A. The models developed during this project incorporate one or more of the following features:

- Two or three years of longitudinal test score data. Meyer and Christian (2008) refer to the class of value-added models based on achievement growth data over a year (or school year) as "T2" models, because they require, at a minimum, two test scores, one at the beginning of the period and one at the end of the period. Models based on growth data over two years are similarly referred to as "T3" models. [1]
- A posttest on pretest parameter λ_t. This parameter allows for situations where the variances of the posttest and pretest variables may be atypical.
- Control for measurement in prior achievement.
- Demographic variables (X) to capture differences across students in achievement growth.

Figure 2.2.1 shows a diagram of a state value-added model based on two years of longitudinal test score. (The same model is presented in statistical notation in Appendix A.)

Figure 2.2.1: Diagram of a Core "T2" State Value-Added Model



The model indicates that achievement at the end of a period (posttest Y_2) is the sum of:

1. Student achievement at the beginning of the period (pretest Y_1) times a posttest-on-pretest parameter (λ). As discussed in Appendix A, this parameter may vary across grades and years.

[1] Note that since statewide testing begins in third grade in Wisconsin and most other states, only two years of (up-to-date) attainment data are generally available to estimate value-added models of achievement growth from third to fourth grade.

2. Student growth that is correlated with student characteristics such as income, English Language Learner (ELL) status, and race/ethnicity (β).
3. Statewide productivity (π) (see further explanation below).
4. School and district productivity (η) (see further explanation below).
5. Student growth that is due to unknown student characteristics and random test measurement error (ε).

The value-added productivity parameters produced by this model (and all of the value-added models developed for this project) are defined in greater detail in Table 2.2.1.

Table 2.2.1: Value-Added Parameters

π_{gt}	Statewide productivity for grade *g* and year *t* (typically normalized to equal zero in a baseline year). This parameter can only be interpreted as a genuine statewide productivity effect if test scores are accurately horizontally equated over time so that changes in test score growth do not reflect test form effects.
η_{gklt}	School and district productivity (hereafter called school/district productivity) for school *k* and district *l* in grade *g* and year *t*. This parameter is centered around zero in each year so that the average school in the district has a value-added rating equal to zero. Changes in statewide productivity are thus absorbed by the parameter π_{gt} . Given the fact that this parameter is centered each year, it can be interpreted as measuring the "relative" productivity of a school/district in a given year.
$\eta_{(T)gklt} = \pi_{gt} + \eta_{gklt}$	Total school, district, and state productivity (centered around zero only in a baseline year). This indicator allows for overall changes in statewide productivity, provided (as mentioned above) that test scores are accurately horizontally equated.

The school/district productivity parameter defined above (either η_{gklt} or $\eta_{(T)gklt}$) is referred to as the "beat the average" (BTA) rating in the Milwaukee value-added system and the "beat the odds" rating in the Minneapolis value-added system because the value of the indicator equals the amount by which it exceeds or falls short of average district productivity in each year (in the case of η_{gklt}) or in the baseline year (in the case of $\eta_{(T)gklt}$). Below, we drop the grade subscript *g* to (somewhat) simplify the notation.

It is important for policy purposes to split school/district productivity η_{klt} into distinct components representing each level of the educational system: district (indexed by *l*), school (indexed by *k*), and classroom (indexed by *j*):

District productivity component = δ_l
School productivity component = ν_{Skl}
Classroom productivity component = ν_{Cjkl}

Given that statewide data does not contain information at the classroom level, our models incorporate information only at the district and school level. Variation in educational productivity at the classroom level cannot be observed and is therefore absorbed into the error term of the model. Differences across schools in average classroom productivity are absorbed into a school/average classroom factor α_{klt} :

$$\alpha_{klt} = \nu_{Sklt} + \bar{\nu}_{C.klt} \tag{1}$$

where $\bar{\nu}_{C.klt}$ = average classroom productivity.[2] The school/district productivity parameter discussed above (η_{klt}) can thus be decomposed into two identifiable components[3]:

$$\eta_{klt} = \alpha_{klt} + \delta_{lt} \tag{2}$$

Both parameters can be viewed as "draws" from probability distributions with means and variances as explained below. The parameter δ_{lt} is equal to the mean of the school productivity distribution for district *l* in year *t*. It can be viewed as the "draw" from a state distribution with mean zero (since the school/district productivity parameters are centered around zero) and variance ω_{δ}^2. This parameter will vary across districts (so that $\omega_{\delta}^2 > 0$) if districts differ systematically in their capacity to create high and low-performing schools. This is an important hypothesis to test in a statewide value-added model. The parameter α_{klt} captures within-district differences in productivity. It can be viewed as a "draw" from a district distribution with mean zero and variance $\omega_{\alpha kl}^2$ (which could vary across districts *l* and schools *k*). The variance parameter $\omega_{\alpha kl}^2$ is a very interesting parameter in a statewide value-added model in that it captures the degree to which districts (and possibly schools) vary in their consistency of performance.[4]

It is important to point out that districts may differ in average productivity even if the district productivity parameters δ_{lt} do not vary across districts (and thus are equal to zero in all districts). Using Equation (2), average district productivity $\bar{\eta}_{.lt}$ is given by

$$\bar{\eta}_{.lt} = \bar{\alpha}_{.lt} + \delta_{lt} \tag{3}$$

[2] The school average is computed across classrooms (represented by the index *j*). To indicate this, the *j* index in the subscript (*jklt*) is replaced by a dot (.), yielding the subscript (*.klt*).
[3] Note that it is not possible to distinguish between the district and school components in districts with only a single school (at a given grade).
[4] This variance parameter also captures differences in the variance of average classroom effects $\bar{\nu}_{C.klt}$ across schools and districts. The variance is a function of the number of classrooms per school (at a given grade level) and thus will tend to be relatively low in schools and districts where the number of classrooms per school (at a given grade level) is relatively high.

that is, the average of within-district school productivity $\bar{\alpha}_{.lt}$ and district productivity.[5] The first term will, in general, not equal zero (even though α_{klt} is drawn from a distribution with mean zero). The distinction between $\bar{\alpha}_{.lt}$ and δ_{lt} is that the former captures the part of school productivity due to chance (that is, good draws from the school productivity distribution) and the latter represents the part of school productivity due to systematic differences in the capacity of districts to produce high vs. low-performing schools.

In summary, in a statewide value-added model, it is possible to investigate three questions that cannot be addressed in separate district value-added systems:

1. Does average productivity (as represented by the mean parameter $\bar{\eta}_{.lt}$) differ across districts?
2. Does the consistency of school productivity (as represented by the variance parameter $\omega^2_{\alpha kl}$) differ across districts?
3. Do districts differ systematically in their capacity to create high and low-performing schools? (Is $\omega^2_{\delta} > 0$?)

In Section 2.8, we present evidence for the State of Wisconsin on the magnitude of the parameters discussed above. In Section 2.9, we present the estimates for Milwaukee, Madison and Waunakee.

Value-Added Tiers: What Constitutes a Large Change in Value-Added Productivity?

The value-added parameters discussed above—for example, state-wide productivity π_t and school/district productivity η_{klt}—are all defined in the units of the Wisconsin Knowledge and Concepts Examination (WKCE) posttest variable since they are obtained from a model of that variable. Although it is often convenient and preferable to measure value-added productivity in units of the state (or district) assessment, one useful and complementary alternative is the tier unit. The tier unit, represented by the parameter ω, is equal to the standard deviation (SD) of school/district productivity η_{klt} in a given year (typically a baseline year).[6] Dividing the school/district "beat the average" rating η_{klt} by the tier/SD unit ω produces an indicator (often referred to as a z-statistic) that has a mean equal to zero, a standard deviation equal to one, and a range of plus or minus 2-3 tier units (-3 to +3). Milwaukee centers its tier scale around 3 so that the tier values generally lie between 0 and 6.

It is important to point out that statewide productivity parameters reflect changes in true average statewide productivity only if test scores are successfully horizontally equated over time.

[5] The district average is computed across the schools (represented by the index *k*). To indicate this, the *k* index in the subscript (*klt*) is replaced by a dot (.), yielding the subscript (*.lt*).

[6] The variance ω^2_t can be estimated directly as a variance component in a model with random school/district effects or by correcting an estimate of the variance of estimated school/district effects (prior to shrinkage) for estimation error.

Errors in horizontal equating of test scores over time would make it hard to interpret a change in estimated state productivity $\hat{\pi}_t$ as a genuine change in productivity. Since test vendors generally do not provide direct evidence on the degree to which test scores are successfully equated, we believe that analysts should exercise caution in accepting unreasonably large increases or decreases in productivity as evidence of true changes in productivity. We propose the following purely subjective rules for detecting possibly erroneous changes in productivity for given schools and for the state as a whole.

- School-level rule: Consider a change in productivity of greater than one tier unit as possible evidence of faulty horizontal equating.
- State-level rule: Consider a change in statewide productivity of greater than 0.5 tier units as possible evidence of faulty horizontal equating.

Based on our experiences, we suspect that horizontal equating errors are common for statewide tests not originally designed to support growth and value-added analysis. In Section 2.5, we examine changes in statewide test scores over time to detect whether the changes are consistent with the suggested statewide rule.

In cases where we suspect that test scores have not been successfully equated, school/district value-added indicators may legitimately be used to compare the productivity of schools *relative* to other schools in the same year. These indicators may also reasonably be compared over time if we believe that the true change in state average value-added is relatively small. In that case a reported change in the relative value added would be approximately equal to the change in absolute value-added.

2.3 Number of Districts and Schools

Table 2.3.1 summarizes distributions of the number of schools and students in districts in the merged data set of November 2006 and November 2007 test data. The majority of districts are single-school districts and the percentage of the single-school districts is higher for middle school grades since middle schools tend to be larger than elementary schools. Specifically, about two-thirds of the districts are single-school districts for elementary school grades and about 85% of the districts have only one school for middle school grades. The percentage of the districts which have five or fewer schools is about 90% for third, fourth, and fifth grades and about 97% for sixth and seventh grades. The total number of students in single-school districts is correspondingly high. As seen in the Table 2.3.1, the total number of students who are in one of the single-school districts is about one-fourth of all students in the state for 3rd, 4th, and 5th grades and it is about one-half of all the students in the state for 6th and 7th grades.

Although it is not surprising that the average number of students per district is higher in districts which have high number of schools, it is noteworthy that the largest district in the state (Milwaukee), with more than one hundred elementary schools, enrolls about 9.5% of all students in the state. Five districts, not including Milwaukee, have more than 20 elementary schools and enroll about 12% of the students in the state. The six largest districts thus enroll more than 21% of all elementary students in the state.

The distributions of number of schools and students in districts are very similar in the merged November 2005 and November 2006 test data – see Table 2.3.2.

Table 2.3.1: Distributions of the Number of Schools and Students in Districts in November 2006 - November 2007 test data.

	Number of Districts					Avg Number of Std per District					Number of Students					% of Students				
	Grades					Grades					Grades					Grades				
Number of Schools in District	3	4	5	6	7	3	4	5	6	7	3	4	5	6	7	3	4	5	6	7
1	279	283	301	364	374	49.4	51.1	59.4	80.5	88.7	13779	14475	17892	29317	33188	24.8	26.1	31.7	51.0	55.4
2	43	42	38	28	30	105.0	103.5	122.9	228.1	248.9	4515	4345	4669	6388	7467	8.1	7.8	8.3	11.1	12.5
3	30	28	21	10	10	165.3	168.5	157.9	420.2	522.6	4960	4717	3316	4202	5226	8.9	8.5	5.9	7.3	8.7
4	20	20	16	1	1	202.9	203.0	201.5	86.0	732.0	4057	4060	3224	86	732	7.3	7.3	5.7	0.1	1.2
5	15	12	11	6	1	235.5	244.3	243.4	286.3	497.0	3533	2931	2677	1718	497	6.4	5.3	4.7	3.0	0.8
6	14	14	13	4		322.1	326.6	331.4	403.3		4509	4573	4308	1613		8.1	8.3	7.6	2.8	0.0
7	2	3	3	2	3	293.0	278.7	288.0	1299.0	1096.3	586	836	864	2598	3289	1.1	1.5	1.5	4.5	5.5
8	1	1	1			131.0	155.0	169.0			131	155	169			0.2	0.3	0.3	0.0	0.0
9	4	3	4	3	1	444.8	437.7	438.5	780.7	1555.0	1779	1313	1754	2342	1555	3.2	2.4	3.1	4.1	2.6
10	1	2	1		1	404.0	430.5	411.0		1066.0	404	861	411		1066	0.7	1.6	0.7	0.0	1.8
11	1	1	1	1	1	391.0	427.0	446.0	1609.0	1558.0	391	427	446	1609	1558	0.7	0.8	0.8	2.8	2.6
12	4	4	4	1		543.5	561.5	566.5	568.0		2174	2246	2266	568		3.9	4.1	4.0	1.0	0.0
14	2	2	2			602.5	607.5	614.5			1205	1215	1229			2.2	2.2	2.2	0.0	0.0
17	1	1	1	1		825	799	800	860		825	799	800	860		1.5	1.4	1.4	1.5	0.0
18	1	1	1			656	679	654			656	679	654			1.2	1.2	1.2	0.0	0.0
21	2	2	2	1		1164	1147	1133	982		2328	2294	2265	982		4.2	4.1	4.0	1.7	0.0
25	1	1	1			1505	1440	1476			1505	1440	1476			2.7	2.6	2.6	0.0	0.0
27	2	2	2			1434	1419	1417			2868	2838	2834			5.2	5.1	5.0	0.0	0.0
92					1					5323					5323	0.0	0.0	0.0	0.0	8.9
98				1					5177					5177		0.0	0.0	0.0	9.0	0.0
122	1	1				5333	5175				5333	5175				9.6	9.3	0.0	0.0	0.0
123			1					5143					5143			0.0	0.0	9.1	0.0	0.0
Total	424	423	424	423	423	14806.9	14653.8	14672.5	12780.1	12687.6	55538	55379	56397	57460	59901	100.0	100.0	100.0	100.0	100.0

Source: Merged November 2006 and November 2007 test data from WKCE.

Table 2.3.2: Distributions of Number of Schools and Students in Districts in November 2007 - November 2008 test data.

	Number of Districts					Avg Number of Std per District					Number of Students					% of Students				
	Grades					Grades					Grades					Grades				
Number of Schools in District	3	4	5	6	7	3	4	5	6	7	3	4	5	6	7	3	4	5	6	7
1	273	280	297	362	376	48.1	51.0	57.8	83.2	88.7	13124	14286	17179	30105	33338	24.4	25.6	30.6	51.1	54.9
2	48	47	42	32	33	98.3	104.9	118.6	212.8	275.5	4717	4931	4982	6809	9093	8.8	8.8	8.9	11.5	15.0
3	30	29	22	9	8	152.9	157.0	152.0	409.3	596.1	4587	4554	3343	3684	4769	8.5	8.2	6.0	6.2	7.8
4	22	20	16	1		191.9	207.7	209.1	102.0		4222	4153	3345	102		7.8	7.4	6.0	0.2	0.0
5	14	12	11	5	1	232.5	242.8	230.1	282.6	540.0	3255	2913	2531	1413	540	6.0	5.2	4.5	2.4	0.9
6	14	15	13	6		330.9	322.7	331.5	374.3		4633	4841	4309	2246		8.6	8.7	7.7	3.8	0.0
7	2	2	3	2	3	212.5	212.0	208.3	1298.0	1137.3	425	424	625	2596	3412	0.8	0.8	1.1	4.4	5.6
8	1	1	1	1	2	153.0	170.0	176.0	1522.0	1264.0	153	170	176	1522	2528	0.3	0.3	0.3	2.6	4.2
9	3	3	4	1		452.0	465.0	449.0	498.0		1356	1395	1796	498		2.5	2.5	3.2	0.8	0.0
10	2	2	1	1		395.0	372.5	387.0	366.0		790	745	387	366		1.5	1.3	0.7	0.6	0.0
11	1	1	1	1	1	417.0	459.0	398.0	1447.0	1547.0	417	459	398	1447	1547	0.8	0.8	0.7	2.5	2.5
12	4	4	4	1		533.8	560.5	580.3	596.0		2135	2242	2321	596		4.0	4.0	4.1	1.0	0.0
14	2	2	2			585.5	597.0	618.0			1171	1194	1236			2.2	2.1	2.2	0.0	0.0
17	1	1	1	1		805.0	799.0	850.0	873.0		805	799	850	873		1.5	1.4	1.5	1.5	0.0
18	1	1	1			658.0	656.0	669.0			658	656	669			1.2	1.2	1.2	0.0	0.0
21	2	2	2	1		1125.5	1147.5	1137.0	1046.0		2251	2295	2274	1046		4.2	4.1	4.1	1.8	0.0
25	1	1	1			1340.0	1431.0	1438.0			1340	1431	1438			2.5	2.6	2.6	0.0	0.0
27	2	2	2			1374.5	1380.0	1396.5			2749	2760	2793			5.1	4.9	5.0	0.0	0.0
93					1					5541.0					5541	0.0	0.0	0.0	0.0	9.1
98				1					5652.0					5652		0.0	0.0	0.0	9.6	0.0
124	1					5103.0					5103					9.5	0.0	0.0	0.0	0.0
125		1					5624.0					5624				0.0	10.1	0.0	0.0	0.0
126			1					5436.0					5436			0.0	0.0	9.7	0.0	0.0
Total	424	426	425	425	425	14209.3	14959.6	14842.1	14762.2	10989.7	53891	55872	56088	58955	60768	100.0	100.0	100.0	100.0	100.0

Source: Merged November 2005 and November 2006 test data from WKCE.

2.4 Summary Statistics of Test Scores

WKCE test scale scores are summarized in two tables: Unmatched sample and matched sample. Unmatched samples exclude observations with (*i*) repeated student IDs in a given year and grade, and (*ii*) race and gender variables are unidentified. Matched samples are constructed by merging two years of test data using unique student IDs by grade and only matched observations are retained in the samples. Table 2.4.1 presents the state-wide summary statistics of test scores in unmatched November 2005, November 2006 and November 2007 test data. The variables in the table are the number of students (N) included in the samples, the state means of test scale scores in WKCE units (MEAN), and their standard deviations (STD) for each year, grade and subject.

Table 2.4.1: Summary Statistics of Test Scores in Unmatched Samples

		Nov-05			Nov-06			Nov-07		
Subject	Grade	N	MEAN	STD	N	MEAN	STD	N	MEAN	STD
Mathematics	3	56456	431.59	45.01	58049	434.85	46.53	58441	432.20	44.12
	4	58147	463.19	45.34	57688	466.85	42.84	58613	466.79	45.09
	5	58655	484.25	42.52	58903	489.81	43.83	58275	493.52	48.28
	6	61473	507.66	43.29	59777	513.49	45.28	59569	514.16	46.27
	7	63507	527.95	45.00	62439	535.36	43.26	60655	533.76	44.33
	8	65111	540.20	48.72	63870	542.85	48.14	62626	541.79	49.57
Reading	3	56456	458.08	37.22	58049	458.60	38.84	58441	457.55	40.28
	4	58147	476.98	45.97	57688	477.39	45.13	58613	476.25	47.20
	5	58655	484.97	46.99	58903	484.84	46.87	58275	484.58	46.38
	6	61473	500.95	48.46	59777	503.17	49.38	59569	503.25	48.84
	7	63507	510.91	46.61	62439	513.02	47.59	60655	513.97	48.26
	8	65111	526.14	49.73	63870	526.85	52.24	62626	527.77	51.75

Source: November 2005, November 2006 and November 2007 Test Data

Table 2.4.2 presents the statewide summary statistics of test scores in matched November 2005 and November 2006 test data (Growth Year 1 data), and in matched November 2006 and November 2007 test data (Growth Year 2 data). The state means of test scale scores in WKCE units (MEAN) and their standard deviations (STD), and grades reported in columns named “pre” refer to first year in the growth years and the ones reported in “post” columns refer to second year in the growth years. For example, in Growth Year 1 data the state mean of fourth grade mathematics scale score for the November 2005 test is 463.90 and the fifth grade mathematics scale score for the November 2006 test is 490.85. Their standard deviations are given by 44.91 and 43.28, respectively.

The state mean of test scale scores for the same year and grade may differ across these two growth years’ data since unmatched observations differ (one of the reasons for this difference is that new students may be joining the state school system and current students may be leaving the state school system.) For example, in Growth Year 1 data the state mean of the third grade mathematics scale score for the November 2005 test is 432.72 and the fourth grade mathematics scale score for the November 2006 test is 468.49. Their standard deviations are given by 44.14 and 41.93, respectively. While in Growth Year 2 data, the fourth grade mathematics scale score for the November 2006 test is 467.62 with standard deviation of 42.33. However, as seen in Table 2.4.2, the change in the state means across two growth years of data for the same year and grade is small (the range of difference is about 0 and about 1.)

As seen in Table 2.4.2 and Figures 2.4.1 – 2.4.2, both pretest and posttest average attainment in the same grades between two growth years differs between approximately 0 and 7.5.

Table 2.4.2: Summary Statistics of Tests in Matched Samples

			Nov 2005 - Nov 2006					Nov 2006 - Nov 2007				
Subject	Grade		N	MEAN		STD		N	MEAN		STD	
	Pre	Post		Pre	Post	Pre	Post		Pre	Post	Pre	Post
Math	3	4	53891	432.72	468.49	44.14	41.93	55538	435.97	468.01	45.72	44.18
	4	5	55872	463.90	490.85	44.91	43.28	55379	467.62	494.70	42.33	47.63
	5	6	56088	485.10	514.68	41.61	44.75	56397	490.72	515.17	43.11	45.58
	6	7	58955	508.64	536.28	42.59	42.69	57460	514.56	534.53	44.65	43.86
	7	8	60768	529.06	544.16	44.11	47.31	59901	536.58	543.10	42.39	48.64
Reading	3	4	53891	459.01	479.22	36.28	43.72	55538	459.52	477.52	37.96	46.04
	4	5	55872	477.67	486.23	45.34	45.35	55379	478.08	485.80	44.62	45.19
	5	6	56088	485.81	504.63	46.17	48.00	56397	485.76	504.24	46.05	47.79
	6	7	58955	502.00	514.21	47.65	46.44	57460	504.31	514.93	48.43	47.12
	7	8	60768	511.96	528.25	45.80	50.95	59901	514.25	529.06	46.68	50.57

Source: Merged November 2005 and November 2006 Test Data & Merged November 2006 and November 2007 Test Data from WKCE.

Growth in the average attainment across grades can also be tracked in Table 2.4.2. Growth in the average mathematics scale score is illustrated in Figure 2.4.1 and growth in the average reading scale score is illustrated in Figure 2.4.2. In Figure 2.4.1 and Figure 2.4.2, each line represents a different cohort. For example, in Figure 2.4.1, the line for Grades 3-4 4-5 tracks the average mathematics scale score of students in Grade 3 in November 2005, Grade 4 in November 2006 and Grade 5 in November 2007. As mentioned earlier, the students in Grade 3-4 in the Growth Year 1 sample may not be exactly the same as the students in Grade 4-5 in the Growth Year 2 sample. If they differ greatly, two identically shaped symbols appear on the same grade in the graphs, as the average may differ a lot. As seen in both graphs and in Table 2.4.2, the average score increases with grade. The increase in the average mathematics score is less for higher grades and the increase in the average reading score is not monotonic with grades.

Figure 2.4.1: Growth in Average Mathematics Scale Score

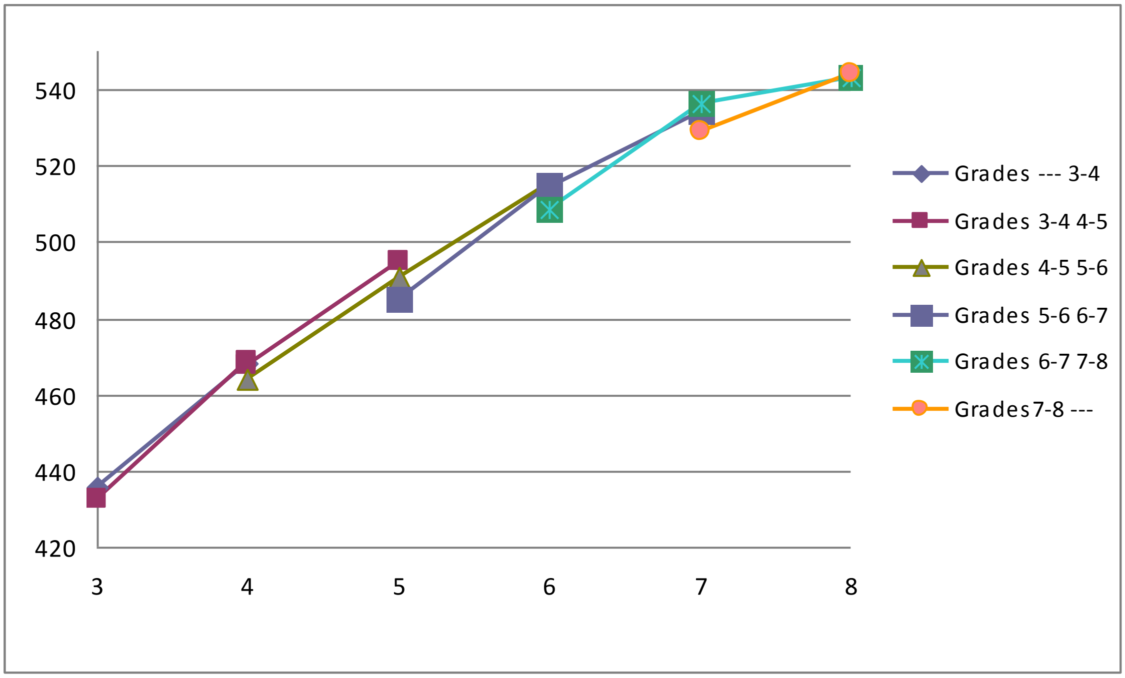


Figure 2.4.2: Growth in Average Reading Scale Score

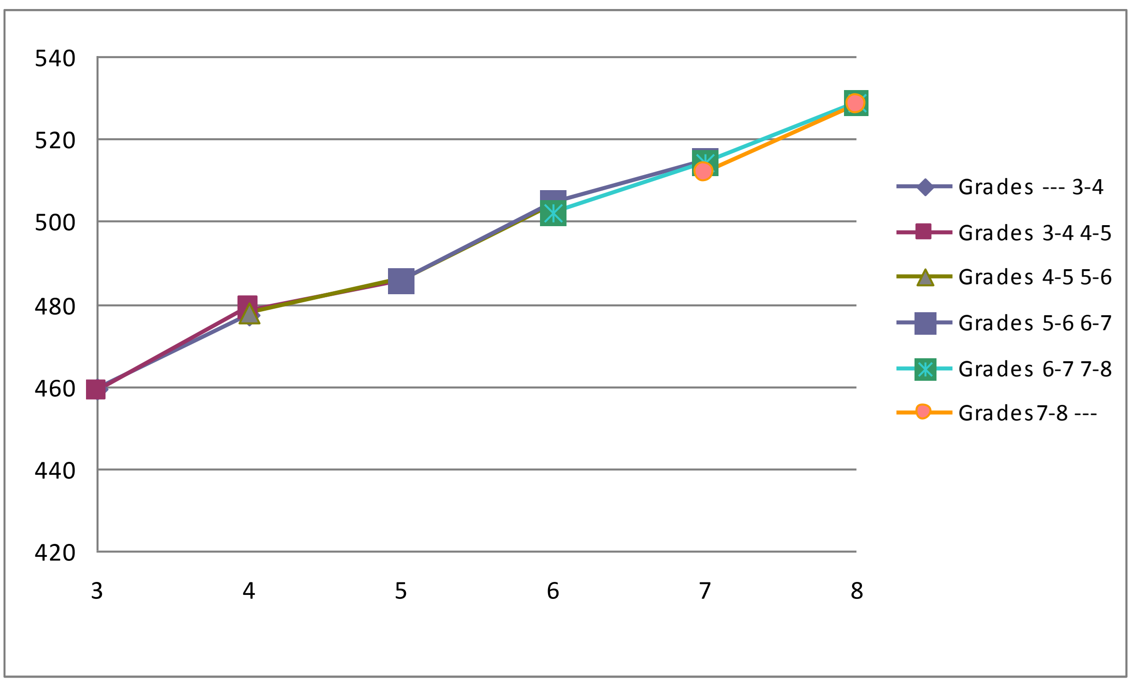
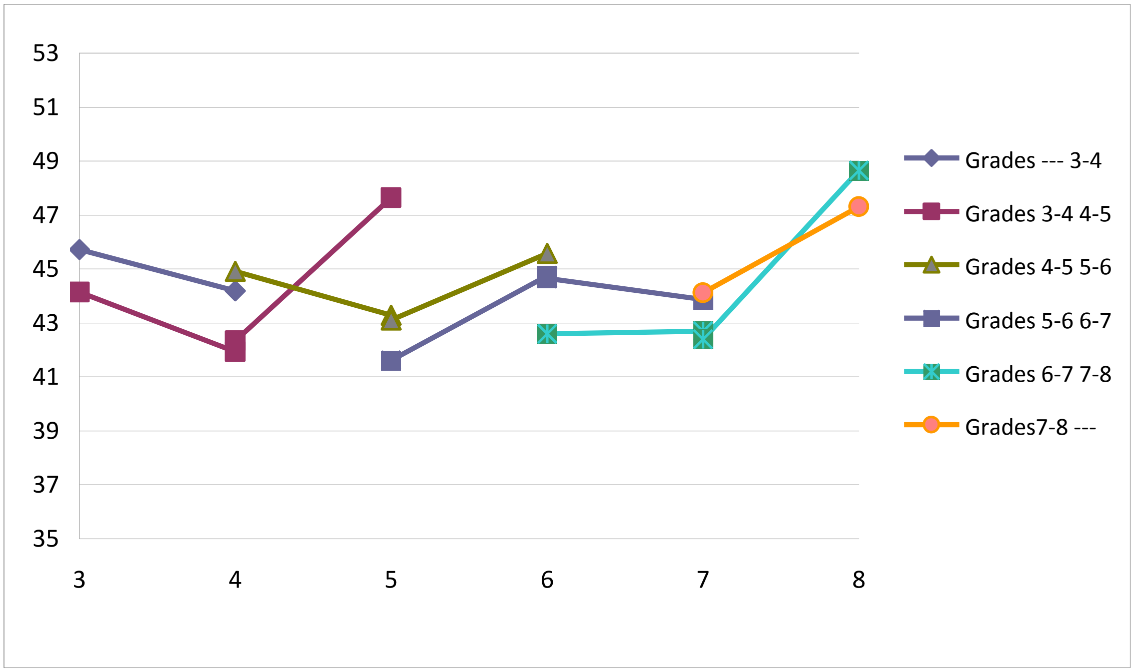


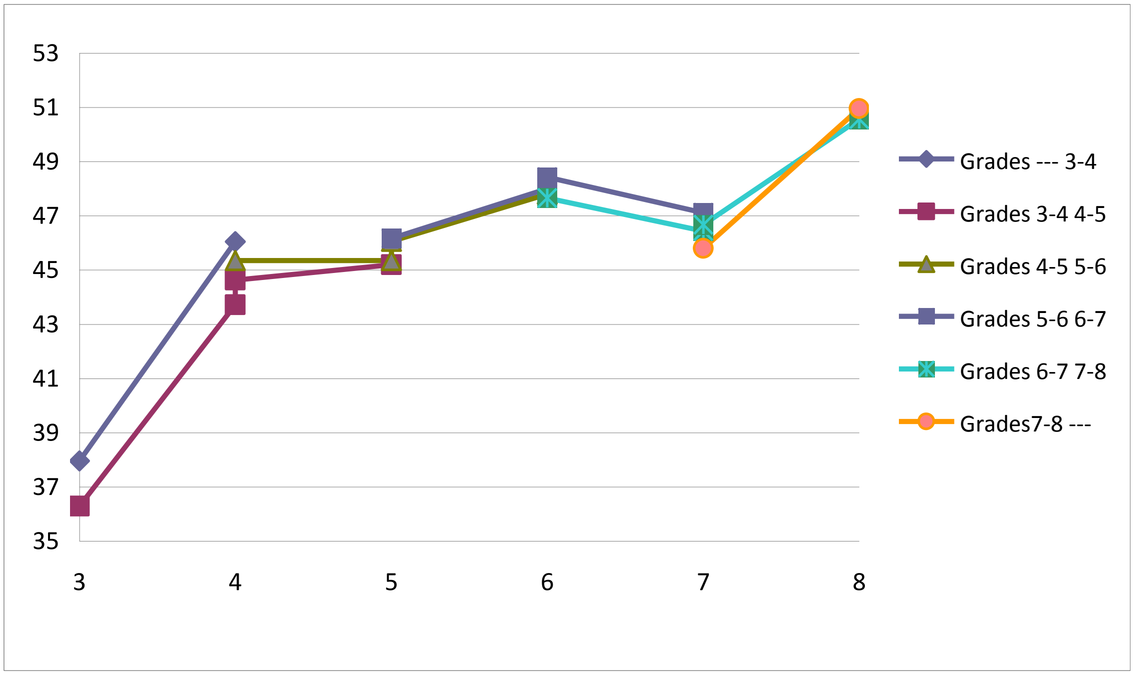
Figure 2.4.3 and Figure 2.4.4 illustrate the standard deviations of test scores across grades. As in the previous two graphs, each line represents a different cohort. Standard deviation of mathematics scores vary across grades (about 0 to 6 scale score units) and cohorts (about 2 to 6 scale score units), while standard deviation of reading scores increases by grade (about 1 to 8 scale score units) and does not vary much by cohorts (about 1 to 2 scale score units).[7]

Figure 2.4.3: Standard Deviations of Mathematics Scores across Grades



[7] The differences in the standard deviations between two grades might be absorbed in the estimates of pre test coefficients. This is discussed in Appendix A.

Figure 2.4.4: Standard Deviations of Reading Scores across Grades



2.5 Comparability of Value-Added Estimates over Time

In cases where there are two or more years of growth data (three or more years of test score data), it is important to assess whether this data supports valid comparisons of growth over time. Unreasonably large increases or decreases in average statewide productivity might indicate problems in maintaining the comparability of test scores over time (that is, the accuracy of horizontal equating across test forms). Errors in horizontal equating of test scores over time would make it illegitimate to interpret a change in estimated state productivity $\hat{\pi}_t$ as a genuine change in productivity. In such a case, school/district value-added indicators can only be used to compare the productivity of schools *relative* to other schools in the same year.

In Section 2.2, we suggested that an increase in average statewide productivity of 0.5 tier units in a given year would constitute an impressively large change. We use this cutoff to determine whether it might be invalid to interpret a change in estimated state productivity $\hat{\pi}_t$ as a genuine change in productivity. As indicated in Appendix B, an approximate estimate of the difference in productivity between two growth years is given by the change in average gain between the two years at a given grade level. These numbers can be computed quite readily and are reported in Table 2.5.1.

The average gain in WKCE units for each growth year is reported in the first two columns of Table 2.5.1. These numbers are taken directly from the attainment statistics presented earlier in Section 2.4, Table 2.4.2. Change in growth in column 3 is simply the difference between column 2 and column 1, columns 4 and 5 report the value-added tier unit (the standard deviation in school/district effects) for the two growth years. These numbers are taken directly from Wisconsin row in Tables 2.8.1 (for mathematics) and 2.8.2 (for reading) in Section 2.8. The tier units at a given grade level are roughly similar in both growth years. The change in average gain given the two tier units are reported in columns 6 and 7, respectively. These numbers in columns 6 and 7 are simply the ratio of change in growth to VA tier unit for Growth Year 1 and 2, respectively. Refer to Appendix B for justification of these calculations. As indicated in the top panel of Table 2.5.1, the change in average gain in mathematics achievement exceeds the specified cutoff of 0.5 tier units in Grades 3, 5, 6, and 7. The change in average mathematics gain is especially large in the latter two grades, substantially in excess of a full tier unit. In fact, in seventh grade, the average gain in achievement declined from approximately 15.1 WKCE scale score points in Growth Year 1 to 6.5 WKCE scale score points in Growth Year 2. This implies that it would be unwise to view the mathematics productivity estimates as comparable between the two growth years in Grade 3 and Grades 5 to 7. Instead, the value-added ratings at these grades in mathematics should only be used to compare schools relative to other schools in the same year.

The average gain numbers in reading are much closer in magnitude between the two growth years at all grade levels. It may be reasonable to view these numbers as being comparable between the two growth years.

Table 2.5.1: Comparison of Average Gain in Achievement in Growth Year 1 (Nov 2005 - 2006) and in Growth Year 2 (Nov 2006 - 2007)

	Mathematics						
Grade	Average Gain Growth Year 1 (WKCE units)	Average Gain Growth Year 2 (WKCE units)	Change in Growth (WKCE units)	VA Tier Unit (SD) Growth Year 1	VA Tier Unit (SD) Growth Year 2	Change in Growth (Growth Year 1 Tier units)	Change in Growth (Growth Year 2 Tier units)
3 to 4	35.77	32.04	-3.72	6.79	7.10	-0.55	-0.52
4 to 5	26.95	27.08	0.13	6.92	7.65	0.02	0.02
5 to 6	29.58	24.45	-5.13	7.20	7.07	-0.71	-0.73
6 to 7	27.64	19.97	-7.66	5.41	4.83	-1.42	-1.59
7 to 8	15.10	6.53	-8.58	5.86	5.06	-1.46	-1.69

	Reading						
Grade	Average Gain Growth Year 1 (WKCE units)	Average Gain Growth Year 2 (WKCE units)	Change in Growth (WKCE units)	VA Tier Unit (SD) Growth Year 1	VA Tier Unit (SD) Growth Year 2	Change in Growth (Growth Year 1 Tier units)	Change in Growth (Growth Year 2 Tier units)
3 to 4	20.20	18.00	-2.20	4.73	5.18	-0.47	-0.42
4 to 5	8.56	7.72	-0.84	4.72	4.72	-0.18	-0.18
5 to 6	18.82	18.48	-0.33	5.24	5.35	-0.06	-0.06
6 to 7	12.21	10.62	-1.59	4.33	3.96	-0.37	-0.40
7 to 8	16.29	14.81	-1.48	4.37	3.96	-0.34	-0.37

Note: As a rough rule of thumb, year-to-year changes in average gain that exceed 0.5 tier (VA standard deviation) units indicate possible test form effects and thus may not represent genuine changes in average state productivity. Grades in which average gain exceeds this threshold are shaded in the tables.

2.6 Estimates of Control Parameters

This section reports estimates of the "core" value-added models in mathematics and reading for all grade levels. See Appendix A for a technical description of the model. The first section discusses the coefficients on prior achievement. The second section discusses the coefficients on the demographic variables included in the model.

Coefficients on Prior Achievement (Corrected for Measurement Error)[8]

One of the important features of the core value-added model used in this project is that it allows for the possibility that the coefficients on prior achievement (λ_{gt}) may differ across grades and years and may not equal one, a parameter restriction that is imposed in some value-added models. In particular, the coefficients on prior achievement may vary in response to irregularities or nonlinearities in scaling that affect the variability of test scores across grades and years. In Appendix A we show that the coefficient on prior achievement can be written as the product of two terms:

$$\lambda_{gt} = M_{gt}\lambda_{gt}^{**} \tag{4}$$

where M_{gt} = the ratio of the standard deviations of measured achievement:

$$M_{gt} = \frac{s_{gt}}{s_{g-1,t-1}} \tag{5}$$

and λ_{gt}^{**} is given by:

$$\lambda_{gt}^{**} = \left(\frac{\sigma_{g-1,t-1}}{\sigma_{gt}}\right)\lambda_{gt}^{*} \tag{6}$$

Note that λ_{gt}^{**} is equal to the parameter that would be estimated if all posttest and pretest scores were transformed to scales with identical standard deviation. We conjecture that λ_{gt}^{**} may vary only slightly across grades and years. If so, plots of estimates of λ_{gt}^{**} with respect to grade or year should be relatively "smooth." Estimates of λ_{gt}. on the other hand may be quite unstable due to scaling irregularities. The ratio of test score standard deviations M_{gt} provides a direct measure of scaling irregularities as they affect the variability of test scores. Empirical evidence on the issue of scaling stability is presented below.

---

[8] All estimates were obtained using an estimation procedure that controls for measurement error in prior achievement. Estimates of λ_{gt} that do not control for measurement error are substantially biased downward.

Table 2.6.1 presents estimates of the three parameters discussed above:

- Ratio of the standard deviations of measured achievement M_{gt}
- Coefficient on prior achievement in units of the standard deviation of post achievement on prior achievement λ_{gt}^{**} (Adjusted pre test coefficient)
- Coefficient on prior achievement λ_{gt} (Unadjusted pre test coefficient)

Figures 2.6.1 and 2.6.2 present estimates of λ_{gt} (solid lines) and λ_{gt}^{**} (dashed lines) graphically for mathematics and reading, respectively.

As indicated in the table, the standard deviation ratio varies widely, from a low of 0.95 to a high of 1.21.[9] Estimates of the coefficient on prior achievement λ_{gt} correspondingly vary widely, from a low of 0.81 to a high of 1.08. As conjectured, estimates of λ_{gt}^{**} are much more stable. In mathematics the λ_{gt}^{**} estimates lie within a narrow band ranging from 0.85 to 0.91. The estimates tend to be smaller in the early grades (approximately 0.86) and larger in the later grades (approximately 0.90). In reading the λ_{gt}^{**} estimates lie in a similar band, although there is less evidence that the estimates differ across grades.

It is very important in the statewide value-added model to allow for the possibility that the coefficients on prior achievement (λ_{gt}) may differ across grades and years. The evidence presented above suggests that most of the variation in estimates of the coefficients on prior achievement is due to instability in the variance of test scores across grades and years.

[9] See also Figures 2.4.3 and 2.4.4 for evidence that the standard deviations of student achievement in mathematics and reading are somewhat unstable across grades and years.

Table 2.6.1: Coefficients on Prior Achievement (Corrected for Measurement Error)

Nov 2005 - Nov 2006						
Grades	Variable	Math		Reading		N
		Estimate	Std Err	Estimate	Std Err	
3 to 4	Post/Pre Test SD Ratio	0.95		1.20		53891
	Pre Test in SD Units	0.86	0.003	0.90	0.003	
	Pre Test	0.81	0.003	1.08	0.004	
4 to 5	Post/Pre Test SD Ratio	0.96		1.00		55872
	Pre Test in SD Units	0.85	0.003	0.87	0.003	
	Pre Test	0.82	0.003	0.87	0.003	
5 to 6	Post/Pre Test SD Ratio	1.08		1.04		56088
	Pre Test in SD Units	0.86	0.003	0.84	0.003	
	Pre Test	0.92	0.003	0.87	0.003	
6 to 7	Post/Pre Test SD Ratio	1.00		0.97		58955
	Pre Test in SD Units	0.89	0.003	0.90	0.003	
	Pre Test	0.89	0.003	0.88	0.003	
7 to 8	Post/Pre Test SD Ratio	1.07		1.11		60768
	Pre Test in SD Units	0.90	0.003	0.90	0.003	
	Pre Test	0.97	0.003	1.00	0.003	
Nov 2006 - Nov 2007						
Grades	Variable	Math		Reading		N
		Estimate	Std Err	Estimate	Std Err	
3 to 4	Post/Pre Test SD Ratio	0.97		1.21		55538
	Pre Test in SD Units	0.87	0.003	0.88	0.003	
	Pre Test	0.84	0.003	1.07	0.004	
4 to 5	Post/Pre Test SD Ratio	1.13		1.01		55379
	Pre Test in SD Units	0.90	0.003	0.87	0.003	
	Pre Test	1.01	0.004	0.88	0.003	
5 to 6	Post/Pre Test SD Ratio	1.06		1.04		56397
	Pre Test in SD Units	0.89	0.003	0.87	0.003	
	Pre Test	0.94	0.003	0.90	0.003	
6 to 7	Post/Pre Test SD Ratio	0.98		0.97		57460
	Pre Test in SD Units	0.91	0.003	0.90	0.003	
	Pre Test	0.89	0.003	0.88	0.003	
7 to 8	Post/Pre Test SD Ratio	1.15		1.08		59901
	Pre Test in SD Units	0.90	0.003	0.91	0.003	
	Pre Test	1.04	0.003	0.99	0.003	

Figure 2.6.1: Pretest Coefficient on Mathematics

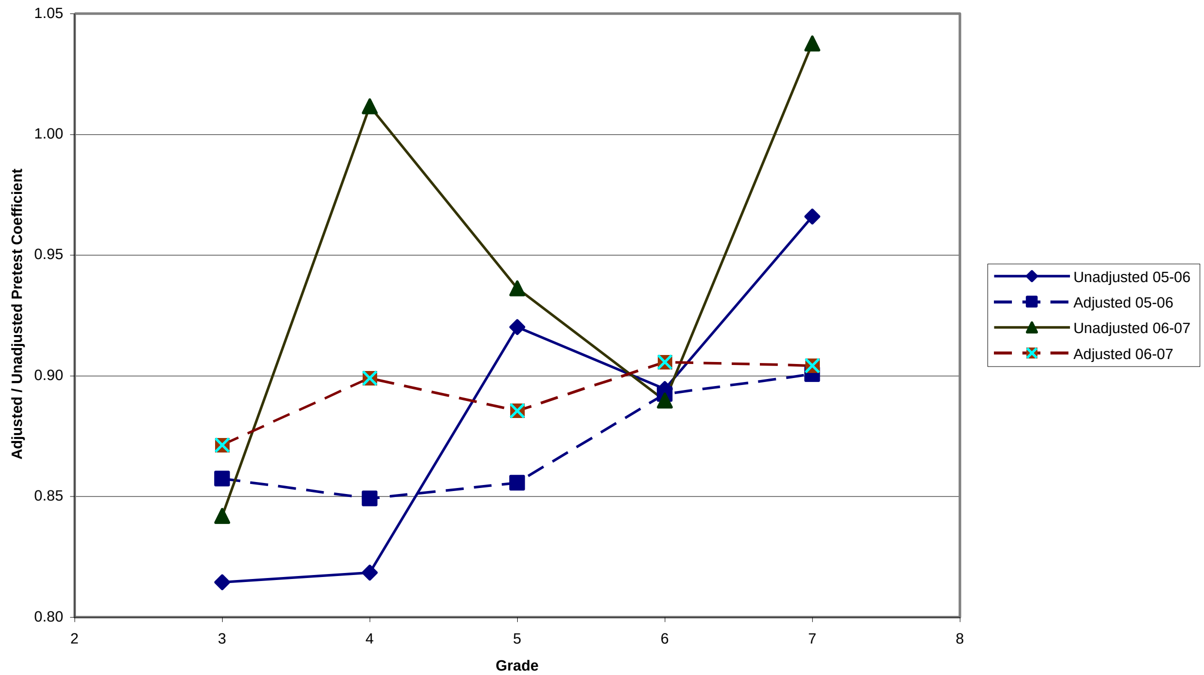
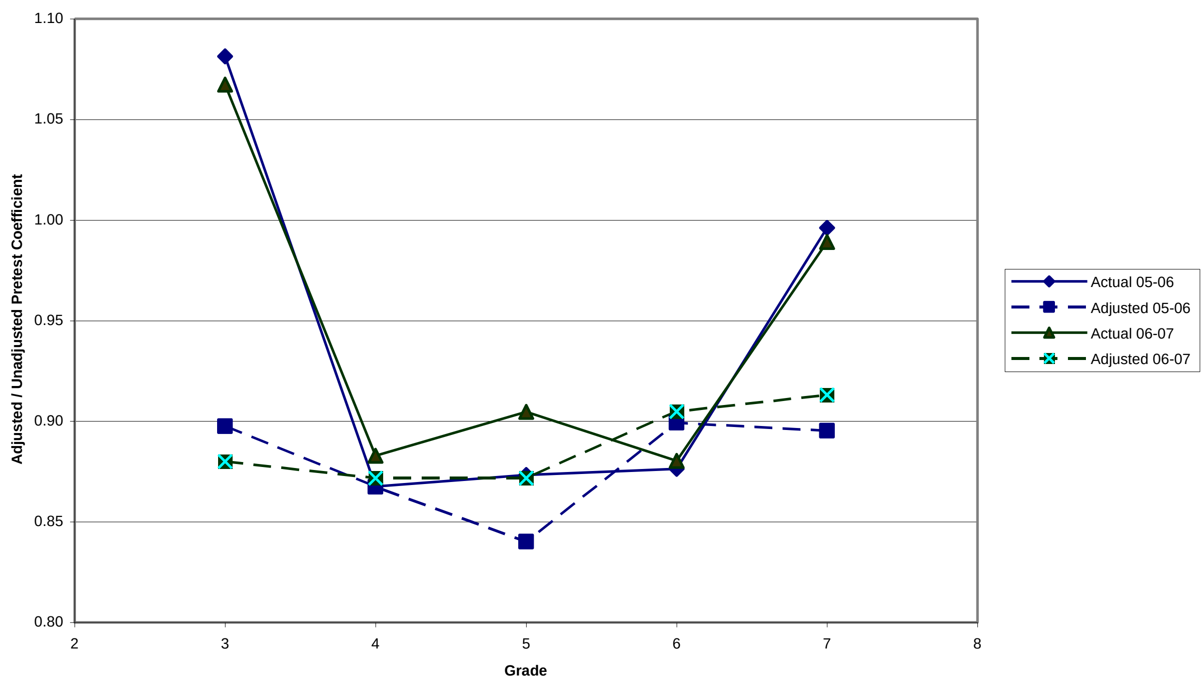


Figure 2.6.2: Pretest Coefficient on Reading



Value-Added Model: Estimates of Slope Coefficients for Demographic Variables

Demographic variables (X) in the value-added model capture differences in achievement growth across students. Demographic variables included in the state value-added model are gender, race/ethnicity, free lunch status, English language learners and disability status. More specifically, the variables are as follows:

- Gender: Male is the omitted categorical variable.
- Race/Ethnicity: White is the omitted categorical variable. African American, Asian, Hispanic and Native American are four race categorical variables included in the estimation.
- Free Lunch Indicator: Students who do not qualify for free or reduced price lunch are the omitted categorical variable. Students who qualify for either free and/or reduced price lunch are classified as one category and included in the estimation.
- English Language Learners: Native English speakers and students who are no longer English language learners are classified as one group and omitted. English language learner students, irrespective of their language mastery level, are classified as one category and included in the estimation.
- Disability Indicator: Students who do not have any kind of disability are omitted. Students with disabilities are classified into one category and included in the estimation.

Table 2.6.2 and Table 2.6.3 present the estimates of slope coefficients for these demographic variables for Growth Year 1 and Growth Year 2, respectively. In Table 2.6.2, grades are in column 1. Demographic variable names are in column 2. Estimates from the mathematics value-added model are in column 3, which is divided in three more columns of which the first one presents the estimates, the second one presents the standard errors, and the last one presents the p-values. Estimates from the reading value-added model are in column 4, which is divided in three more columns, of which first one presents the estimates, the second one presents the standard errors, and the last one presents the p-values. The fifth column presents the overall average of the categorical variables (which is also the ratio of the corresponding category for the corresponding grade; for example, 0.49 in the first row and 5th column means that 49% of the matched data for Grade 3-4 in Growth Year 1 is female, hence 0.49 times sixth column gives the number of female students for this particular cohort and so on), and finally the sixth column presents the number of students used in the estimation. Table 2.6.3 is constructed similarly.

The slopes of categorical variables should be interpreted relative to the corresponding omitted categorical variable. For example, the slope coefficient reports for female, *ceteris paribus*, how much the growth in test scores of female students differs from that of male students. For example, in Table 2.6.2, the slope coefficient for female is -0.58. Holding all other variables constant, the growth in test scores of female students are 0.58 scale score points lower than that of male students[10]. The coefficient estimates of the first three rows from Growth Year 1 in Grade 3 to 4 from Table 2.6.2 are presented and interpreted below;

[10] Note that the reported demographic coefficients do not necessarily represent causal effects. For example, female and male students may differ in ways (not included in the model) that are related to growth in student achievement.

- Female coefficient is -0.58, standard error is 0.22 and p-value is 0.01. Since p-value is less than 0.05, the estimate is statistically significant at the significance level of 5%[11]. Holding all other variables constant, the growth in test scores of female students is 0.58 scale score units lower than that of male students.
- African American coefficient is -4.24, standard error 0.54 and p-value is less than 0.005. Since p-value (<0.005) is less than 0.05, the estimate is statistically significant at the significance level of 5%. Holding all other variables constant, the growth in test scores of African American students is 4.24 scale score units lower than that of white students.
- Hispanic coefficient is -0.91, standard error 0.57 and p-value is 0.11. Since p-value is more than 0.05, the estimate is statistically insignificant at the significance level of 5%[12].

Estimates of slope coefficients for demographic characteristics across grades and cohorts are not as stable as one would expect. For example, we see that the slope coefficient for Asian students is -1.79 in Growth Year 1 for Grade 3-4 and 5.26 in Growth Year 2 for Grade 4-5 (same cohort of students) and, as seen from the Tables 2.6.2 and 2.6.3, both estimates have very low standard errors. For this particular case, the test administered in Growth Year 1 might have items particularly difficult for Asian students[13], hence it might result in lower scores for Asian students. Low scores in Growth Year 1 (for Grade 3-4) combined with an absence of such an item in Growth Year 2 might easily end up in big positive slope coefficient for Asian students in Growth Year two (for Grade 4-5). Instability of estimates between different cohorts can be (at least partially) attributed to cohort effects. However, attributing all instances of instabilities of estimates across grades and cohorts to these effects might not be right.

The only positive instance of African American coefficient in Growth Year 1 for Grade 7-8 is 0.60 and its standard error is relatively large. Hence it is not statistically significant different from zero (at the significance level of 5%.) and its p-value is 0.23. For mathematics tests, the only positive instance of female coefficient in Growth Year 1 for Grade 4-5 is 1.36 and it is statistically significantly different from zero.

Some of the demographic variables included in our model may serve, in part, as proxy variables for unobserved factors such as income and parental education that are not measured in the state data. This fact does not imply that the value-added effects included in the model are biased. The value-added effects will be unbiased if the “set” of demographic coefficients included in the model capture (predict) all systematic differences across schools in achievement growth.

[11] More precisely, the t-test can reject the hypothesis that this particular coefficient is 0 at the asymptotic 5% level.

[12] More precisely, the t-test cannot reject the hypothesis that this particular coefficient is 0 at the asymptotic 5% level.

[13] For example, a very simple probability question that uses a Poker game as its frame might be an impossible question to solve by somebody who has no idea of what a Poker game is.

Table 2.6.2: Estimates of Slope Coefficients for Demographic Variables, Growth Year 1

Nov 2005 - Nov 2006										
Grades	Variable	Math			Reading			Mean	N	
		Estimate	Std Err	Pr > |t|	Estimate	Std Err	Pr > |t|			
3 to 4	Female	-0.58	0.22	0.01	-2.87	0.23	0.00	0.49	53891	26502
	African American	-4.24	0.54	0.00	-2.07	0.58	0.00	0.10		5424
	Hispanic	-0.91	0.57	0.11	0.56	0.61	0.36	0.06		3337
	Asian	-1.79	0.72	0.01	-0.37	0.77	0.63	0.03		1748
	Native American	-1.27	1.07	0.23	-2.99	1.14	0.01	0.01		749
	Free Lunch Indicator	-1.82	0.28	0.00	-2.36	0.30	0.00	0.31		16970
	English Language Learner	-1.07	0.70	0.13	-2.56	0.75	0.00	0.04		2270
	Disability Indicator	-8.52	0.36	0.00	-4.28	0.39	0.00	0.11		6123
4 to 5	Female	1.36	0.21	0.00	3.83	0.22	0.00	0.49	55872	27335
	African American	-1.39	0.52	0.01	-3.88	0.55	0.00	0.10		5690
	Hispanic	0.26	0.55	0.63	-1.62	0.59	0.01	0.07		3666
	Asian	4.11	0.67	0.00	-1.71	0.72	0.02	0.03		1931
	Native American	-0.88	1.00	0.38	-3.35	1.07	0.00	0.01		778
	Free Lunch Indicator	-2.33	0.27	0.00	-2.79	0.29	0.00	0.32		17691
	English Language Learner	-2.03	0.64	0.00	0.07	0.69	0.92	0.05		2675
	Disability Indicator	-7.00	0.33	0.00	-6.53	0.37	0.00	0.12		6782
5 to 6	Female	-0.34	0.21	0.10	5.51	0.23	0.00	0.49	56088	27459
	African American	-1.99	0.52	0.00	-4.15	0.58	0.00	0.10		5717
	Hispanic	0.16	0.55	0.77	-1.93	0.61	0.00	0.06		3626
	Asian	3.85	0.68	0.00	-3.33	0.75	0.00	0.03		1902
	Native American	-2.60	0.98	0.01	-1.24	1.10	0.26	0.01		766
	Free Lunch Indicator	-2.70	0.27	0.00	-3.01	0.30	0.00	0.32		17747
	English Language Learner	-0.44	0.65	0.49	-2.27	0.72	0.00	0.05		2566
	Disability Indicator	-6.23	0.34	0.00	-7.60	0.39	0.00	0.12		6909
6 to 7	Female	-1.99	0.18	0.00	-2.90	0.21	0.00	0.49	58955	29047
	African American	-3.31	0.44	0.00	-2.77	0.52	0.00	0.10		6044
	Hispanic	0.01	0.48	0.98	-0.37	0.57	0.51	0.06		3411
	Asian	-0.10	0.60	0.87	-0.27	0.71	0.70	0.03		1917
	Native American	-0.53	0.82	0.52	-0.22	0.96	0.82	0.01		872
	Free Lunch Indicator	-2.04	0.23	0.00	-1.73	0.28	0.00	0.31		18117
	English Language Learner	-1.86	0.60	0.00	-0.96	0.70	0.17	0.04		2284
	Disability Indicator	-4.84	0.30	0.00	-2.44	0.37	0.00	0.13		7400
7 to 8	Female	-1.64	0.20	0.00	-1.11	0.24	0.00	0.49	60768	29630
	African American	0.60	0.50	0.23	-1.21	0.59	0.04	0.10		5955
	Hispanic	-1.09	0.53	0.04	1.28	0.63	0.04	0.06		3358
	Asian	3.28	0.64	0.00	5.46	0.76	0.00	0.03		2067
	Native American	-1.67	0.90	0.06	-4.00	1.07	0.00	0.02		927
	Free Lunch Indicator	-2.11	0.26	0.00	-1.87	0.31	0.00	0.30		17985
	English Language Learner	1.50	0.68	0.03	4.79	0.81	0.00	0.03		2047
	Disability Indicator	-4.24	0.34	0.00	-4.41	0.41	0.00	0.13		7697

Table 2.6.3: Estimates of Slope Coefficients for Demographic Variables, Growth Year 2

Nov 2006 - Nov 2007										
Grades	Variable	Math			Reading			Mean	N	
		Estimate	Std Err	Pr > |t|	Estimate	Std Err	Pr > |t|			
3 to 4	Female	-2.29	0.24	0.00	-1.90	0.24	0.00	0.49	55538	27258
	African American	-2.92	0.58	0.00	-3.58	0.58	0.00	0.10		5631
	Hispanic	-0.94	0.61	0.12	-0.41	0.61	0.50	0.08		4248
	Asian	2.49	0.77	0.00	0.00	0.78	1.00	0.03		1831
	Native American	-1.65	1.13	0.14	-0.35	1.13	0.75	0.01		763
	Free Lunch Indicator	-0.70	0.30	0.02	-2.62	0.31	0.00	0.34		19030
	English Language Learner	-2.50	0.68	0.00	0.02	0.68	0.98	0.06		3357
	Disability Indicator	-8.04	0.38	0.00	-4.58	0.39	0.00	0.12		6603
4 to 5	Female	-0.75	0.24	0.00	0.53	0.23	0.02	0.49	55379	27197
	African American	-0.04	0.58	0.94	-4.91	0.55	0.00	0.10		5603
	Hispanic	0.16	0.61	0.79	-1.21	0.59	0.04	0.07		4008
	Asian	5.26	0.77	0.00	-0.32	0.74	0.67	0.03		1895
	Native American	-1.21	1.15	0.29	0.63	1.11	0.57	0.01		745
	Free Lunch Indicator	-1.38	0.31	0.00	-1.48	0.29	0.00	0.33		18334
	English Language Learner	-0.70	0.70	0.32	1.67	0.68	0.01	0.06		3128
	Disability Indicator	-5.44	0.38	0.00	-5.22	0.38	0.00	0.12		6855
5 to 6	Female	-0.41	0.21	0.05	-1.08	0.23	0.00	0.49	56397	27529
	African American	-4.93	0.52	0.00	-4.86	0.57	0.00	0.10		5555
	Hispanic	-1.04	0.55	0.06	0.31	0.60	0.60	0.07		3914
	Asian	1.51	0.67	0.02	-2.07	0.73	0.00	0.04		2013
	Native American	-2.12	1.01	0.03	-1.56	1.10	0.15	0.01		763
	Free Lunch Indicator	-1.48	0.27	0.00	-2.37	0.30	0.00	0.33		18351
	English Language Learner	0.75	0.63	0.23	-1.06	0.69	0.12	0.06		3127
	Disability Indicator	-9.48	0.34	0.00	-6.39	0.38	0.00	0.13		7211
6 to 7	Female	-1.47	0.19	0.00	-3.47	0.23	0.00	0.49	57460	28180
	African American	-4.30	0.46	0.00	-1.89	0.56	0.00	0.10		5745
	Hispanic	-1.29	0.49	0.01	0.45	0.59	0.45	0.07		3928
	Asian	-0.32	0.61	0.60	1.70	0.73	0.02	0.04		2019
	Native American	-0.32	0.90	0.72	-2.56	1.08	0.02	0.01		790
	Free Lunch Indicator	-1.62	0.25	0.00	-2.29	0.30	0.00	0.32		18555
	English Language Learner	-4.78	0.57	0.00	-0.86	0.69	0.21	0.05		2883
	Disability Indicator	-5.52	0.32	0.00	-1.90	0.40	0.00	0.12		7110
7 to 8	Female	-1.98	0.21	0.00	4.86	0.23	0.00	0.49	59901	29471
	African American	-4.84	0.51	0.00	-1.59	0.56	0.00	0.10		5971
	Hispanic	-1.89	0.55	0.00	-0.28	0.61	0.65	0.06		3624
	Asian	1.39	0.68	0.04	2.33	0.76	0.00	0.03		2043
	Native American	-0.45	0.95	0.64	-1.48	1.06	0.16	0.01		856
	Free Lunch Indicator	-2.19	0.27	0.00	-1.05	0.30	0.00	0.31		18455
	English Language Learner	0.31	0.66	0.64	3.28	0.73	0.00	0.04		2664
	Disability Indicator	-4.56	0.35	0.00	-6.47	0.39	0.00	0.13		7547

2.7 Better Value-Added Estimates Using Shrinkage Estimation

One of the interesting challenges in reporting and using estimates of value-added productivity indicators is that they are inevitably measured with error. The variance of this error is largely a function of the number of students included in the estimation. As a result, estimation error tends to be highest at the classroom level (not the focus of this report), relatively high at the school/grade level for small schools, and lowest at the school level (aggregated across grades) for large schools. Since there is substantial variation across the state in the size of districts and schools, the precision of estimated value-added indicators tends to vary a lot across districts and schools. This creates the following problem: the lowest and highest performing schools, based on typical (means-type) value-added estimates, tend to be schools with the fewest number of students. It is apparent that that these schools show up in the extremes of the estimated distribution of value-added indicators largely because they were very unlucky or lucky, in the sense of drawing large negative or large positive estimation errors.

Because of this problem, we prefer to estimate value-added indicators using an estimation technique called “ensemble” estimation or “shrinkage” estimation. The basic idea behind this estimation technique is that better estimates can be produced (in the sense of having lower expected mean square error (MSE)) by incorporating information into the estimation process on the variance (or spread) of the true parameters measured without error. Shrinkage is essentially a technique that moderates the values added of schools with very high or very low value-added where those extremes in value-added are the result of statistical noise. In practice, these techniques can be embedded in the value-added estimation itself, or they can be applied after unshrunk value-added has been computed.

Consider the following example. Suppose that there are two types of schools, small schools and large schools. The statistics associated with these schools are reported in the table below.

Table 2.7.1: Statistics for Hypothetical Sets of Small and Large Schools

	Small Schools	Large Schools
True Average Value-Added (centered around zero)	0.0	0.0
Variance in True School Performance	100	100
Standard Error of (Means-Type) Value-Added Estimate	8.2	5.0
Variance of Error of (Means-Type) Value-Added Estimate	66.7	25.0
Total Variance	166.7	125.0
Reliability of Means-Type Value-Added Estimate	0.60	0.80
True Range of VA Indicators (95% interval) (plus or minus)	19.6	19.6
Range of Estimated (Means-Type) VA Indicators (95% interval) (plus or minus)	25.3	21.9

As indicated in the table, we assume that the true mean and variance of small and large schools are the same. As a result, if the true value-added indicators are approximately normally distributed, the range of true value-added indicators for both small and large schools (with 95%

coverage) lies in the interval of ± 19.6 (since the indicators are centered around zero). The standard errors are larger for small schools, however (8.2 versus 5.0). Hence, the range of estimated (means-type) indicators is much larger for small schools than large schools (± 25.3 versus ± 21.9). The solution provided by the technique of shrinkage estimation is to shrink the means-type estimates toward the mean value-added effect by multiplying the estimated means-type indicator times the reliability of the estimate. As indicated in the table, the reliability of value-added estimates is lower for small schools (0.60 versus 0.80), so shrinkage estimation shrinks the "noisy" estimates for small schools more than the (less noisy) estimates for large schools. The end result is that shrinkage estimates are better estimates of means-type estimates in that they minimize expected mean squared error. In Section 2.9, we present estimates of means-type (unshrunk) estimates and shrinkage estimates.

2.8 The Reliability and Variance of Value-Added Productivity in Mathematics and Reading in Wisconsin

In this section we report on the statistical properties of the estimated statewide value-added indicators in reading and mathematics for growth years 1 (2005-2006) and 2 (2006-2007). We address three major questions:

a) Is variability in school and district productivity important from a policy perspective?
b) Are the value-added indicators estimated with sufficient precision so that it is possible to reliably distinguish the difference between high and low performing schools?
c) Do differences across districts in average performance reflect systematic differences in the capacity of districts to generate schools at a given productivity level, or are these differences simply the result of essentially random (unsystematic) generation of schools of given quality?

In answering these questions we draw on evidence for the entire state as well as evidence from Madison and Milwaukee.

In order to answer the first question, we need a standard by which to evaluate the magnitude of a change in value-added productivity in a given year. We present data for four different standards:

- A year of growth (measured by the average grade-to-grade growth over the two growth years in our data, 2005-2006 and 2006-2007). This statistic varies by grade and across subjects (mathematics and reading).
- The basic to proficient gap in scale score points.
- The basic to advanced gap in scale score points.
- The student-level effect size (the standard deviation of post achievement) in each grade and year.

Note that the last three standards are cumulative attainment (rather than growth) standards, so that it might be more appropriate to use these standards to evaluate the magnitude of changes in value-added across multiple years/grades.

We measure the magnitude of value-added productivity effects by the difference in productivity for a single year of instruction between the low and high performing schools statewide and in Madison and Milwaukee.[14] In order to err on the conservative side in estimating this quantity, we compute the "low-to-high" productivity difference as four times the noise-corrected standard deviation of value-added productivity, rather than the actual difference between the low and high-performing schools.[15] The rationale for using four times the standard deviation is that the resulting range encompasses more than 95% of the schools. If value-added

[14] It could be argued that a measure of the magnitude of value-added effects should include differences between low and high-performing classrooms rather than differences between low and high-performing schools. The former quantity is certainly larger. We use numbers based on school-level productivity in this report because the state data does not include information at the classroom level.

[15] We use the noise-corrected standard deviation rather than the actual standard deviation, because the actual standard deviation is biased upward due to statistical estimation error. See the discussion on the statistical reliability of value-added indicators later in this section.

productivity is approximately normally distributed, this corresponds to a range of productivity from approximately the 2nd to the 98th percentiles of value-added productivity.

As reported in Tables 2.8.1 and 2.8.2, the low-to-high productivity gap differs between the state (as a whole) and Madison and Milwaukee, with the highest productivity gap occurring in Milwaukee and the lowest gap occurring in Madison. For the state as a whole the productivity gap ranges from 19 to 30 scale score points in mathematics and 16 to 21 scale score points in reading. If we accept that scale score points in mathematics and reading can be compared in a very rough manner, then our evidence suggests that productivity in mathematics is less consistent across the state than productivity in reading. Interestingly, differences between mathematics and reading productivity are much less pronounced in Madison and Milwaukee.

As indicated in Tables 2.8.1 and 2.8.2, productivity gaps of the magnitude discussed above (19-30 points in mathematics, 16-21 points in reading) are quite large relative to all four of the standards presented in the tables. In fact, the low-to-high productivity gap typically is larger than an entire year of student achievement growth in both mathematics and reading in almost all grades. Similarly, the low-to-high productivity gap typically exceeds the gap between basic and proficient status in mathematics and is greater than 50% of the low-to-high productivity gap in reading. In short, it is evident that variability in school and district productivity is important from a policy perspective. In the next section we discuss the results for Madison and Milwaukee in greater detail. For our purposes here, simply note that the low-to-high productivity gap is very large for Milwaukee, as much as double the statewide gap in third grade.

Table 2.8.1: Magnitude of Value-Added Productivity Parameters in Mathematics

	Mathematics									
	Growth Year 1 (Nov 2005 – Nov 2006)					Growth Year 2 (Nov 2006 – Nov 2007)				
	3 to 4	4 to 5	5 to 6	6 to 7	7 to 8	3 to 4	4 to 5	5 to 6	6 to 7	7 to 8
Standard Deviation of VA Productivity (Noise Corrected)										
Wisconsin	6.79	6.92	7.20	5.41	5.86	7.10	7.65	7.07	4.83	5.06
Madison	3.56	5.60	5.85	5.37	3.05	8.16	3.06	5.75	2.93	2.59
Milwaukee	11.39	8.53	10.22	6.38	6.53	11.34	9.54	8.53	6.57	9.40
Low to High VA Productivity										
Wisconsin	27.14	27.69	28.81	21.64	23.45	28.42	30.61	28.29	19.32	20.25
Madison	14.24	22.41	23.39	21.48	12.21	32.66	12.26	23.02	11.71	10.34
Milwaukee	45.55	34.12	40.88	25.53	26.12	45.35	38.16	34.13	26.30	37.58
Basic to Proficient Gap	15.00	17.00	18.00	21.00	24.00	15.00	17.00	18.00	21.00	24.00
Basic to Advanced Gap	60.00	63.00	60.00	68.00	75.00	60.00	63.00	60.00	68.00	75.00
One Year of Growth	33.91	27.02	27.02	23.81	10.82	33.91	27.02	27.02	23.81	10.82
Student Level Effect Size (Standard Deviation of Post Achievement)	41.93	43.28	44.75	42.69	47.31	44.18	47.63	45.58	43.86	48.64

Table 2.8.2: Magnitude of Value-Added Productivity Parameters in Reading

	Reading									
	Growth Year 1 (Nov 2005 – Nov 2006)					Growth Year 2 (Nov 2006 – Nov 2007)				
	3 to 4	4 to 5	5 to 6	6 to 7	7 to 8	3 to 4	4 to 5	5 to 6	6 to 7	7 to 8
Standard Deviation of VA Productivity (Noise Corrected)										
Wisconsin	4.73	4.71	5.24	4.33	4.37	5.18	4.72	5.35	3.96	3.96
Madison	4.71	3.45	5.33	5.75	5.63	4.92	4.91	2.83	2.88	4.24
Milwaukee	7.69	8.23	8.63	6.09	6.43	8.28	8.16	8.21	6.08	5.58
Low to High VA Productivity										
Wisconsin	18.92	18.86	20.97	17.32	17.49	20.72	18.86	21.39	15.83	15.85
Madison	18.84	13.79	21.33	23.00	22.51	19.69	19.66	11.30	11.51	16.95
Milwaukee	30.75	32.91	34.53	24.38	25.72	33.12	32.62	32.82	24.31	22.30
Basic to Proficient Gap	36.00	44.00	43.00	39.00	33.00	36.00	44.00	43.00	39.00	33.00
Basic to Advanced Gap	72.00	93.00	96.00	96.00	89.00	72.00	93.00	96.00	96.00	89.00
One Year of Growth	19.10	8.14	18.65	11.42	15.55	19.10	8.14	18.65	11.42	15.55
Student Level Effect Size (Standard Deviation of Post Achievement)	43.72	45.35	48.00	46.44	50.95	46.04	45.19	47.79	47.12	50.57

Tables 2.8.3, 2.8.4, and 2.8.5 report on the reliability of estimated value-added indicators statewide and in Madison and Milwaukee. At the state level the reliabilities range from 77% to 85% in mathematics and from 60% to 70% in reading. The lower reliabilities in reading than mathematics reflect the fact that the variation in productivity is less in reading than in mathematics (if it is reasonable to compare numbers on the reading and mathematics scales). The average variance of statistical estimation error (noise) is roughly the same in both reading and mathematics. Reliabilities of this magnitude are sufficiently large to be able to accurately contrast the productivity of low and high performing schools. As indicated in Table 2.8.5, the reliabilities in both mathematics and reading in Milwaukee are quite large, ranging from 81% to 90% in mathematics and 70% to 82% in reading. This is because variation in school/district productivity in Milwaukee is much larger than in the rest of the state.

Finally, Tables 2.8.3, 2.8.4, and 2.8.5 indicate that there are systematic differences in the capacity of districts to generate schools at a given productivity level. Although the results vary somewhat from grade to grade, the share of total (noise corrected) productivity variation due to systematic district differences is typically about 20% in mathematics and 30% in reading. This implies that differences in school/district productivity are due to differences in district-wide policies and resources as well as differences in school-level policies and resources.

Table 2.8.3: Estimates of the Reliability and Variance of District and Within-District School Effects in Wisconsin

	Mathematics									
	Growth Year 1 (Nov 2005 – Nov 2006)					Growth Year 2 (Nov 2006 – Nov 2007)				
	3 to 4	4 to 5	5 to 6	6 to 7	7 to 8	3 to 4	4 to 5	5 to 6	6 to 7	7 to 8
Variance of district VA effects	9.50	16.20	20.39	0.0	9.68	2.54	18.75	17.09	3.24	0.25
Average variance of VA within-district school effects	36.54	31.72	31.51	29.28	24.70	47.94	39.79	32.94	20.09	25.39
Variance of total school/district VA effects	46.05	47.93	51.89	29.28	34.38	50.48	58.55	50.03	23.33	25.64
Average variance of statistical estimation error (noise)	13.07	11.75	11.31	5.69	5.94	15.22	14.89	11.47	6.34	6.44
Total variance, including statistical estimation error	59.11	59.68	63.20	34.97	40.32	65.70	73.44	61.50	29.67	32.08
Average reliability of total school/district VA effects	0.78	0.80	0.82	0.84	0.85	0.77	0.80	0.81	0.79	0.80
	Reading									
	Growth Year 1 (Nov 2005 – Nov 2006)					Growth Year 2 (Nov 2006 – Nov 2007)				
	3 to 4	4 to 5	5 to 6	6 to 7	7 to 8	3 to 4	4 to 5	5 to 6	6 to 7	7 to 8
Variance of district VA effects	4.92	7.35	10.19	8.67	1.19	5.98	7.11	13.60	9.12	0.0
Average variance of VA within-district school effects	17.46	14.89	17.29	10.07	17.91	20.84	15.13	15.00	6.56	15.70
Variance of total school/district VA effects	22.37	22.23	27.48	18.75	19.11	26.82	22.24	28.60	15.67	15.70
Average variance of statistical estimation error (noise)	14.88	13.55	14.10	7.86	8.40	15.28	13.81	13.60	9.17	7.99
Total variance, including statistical estimation error	37.26	35.78	41.58	26.61	27.51	42.10	36.05	42.20	24.84	23.69
Average reliability of total school/district VA effects	0.60	0.62	0.66	0.70	0.69	0.64	0.62	0.68	0.63	0.66

Table 2.8.4: Estimates of the Reliability and Variance of District and Within-District School Effects in Madison

	Madison									
	Math 2005-2006					Math 2006-2007				
	3	4	5	6	7	3	4	5	6	7
Variance of VA within-district school effects	12.67	31.39	34.18	28.82	9.31	66.66	9.39	33.12	8.58	6.69
Average variance of statistical estimation error (noise)	10.64	10.22	10.39	3.55	4.19	12.44	12.10	10.08	3.52	4.43
Total variance, including statistical estimation error	23.31	41.61	44.58	32.38	13.50	79.09	21.50	43.20	12.10	11.11
Reliability	0.54	0.75	0.77	0.89	0.69	0.84	0.44	0.77	0.71	0.60
	Read 2005-2006					Read 2006-2007				
	3	4	5	6	7	3	4	5	6	7
Variance of VA within-district school effects	22.18	11.89	28.43	33.06	31.66	24.22	24.15	7.98	8.28	17.95
Average variance of statistical estimation error (noise)	12.11	11.78	12.97	4.91	5.93	12.48	11.23	11.96	5.09	5.49
Total variance, including statistical estimation error	34.29	23.67	41.40	37.97	37.58	36.70	35.38	19.94	13.37	23.44
Reliability	0.65	0.50	0.69	0.87	0.84	0.66	0.68	0.40	0.62	0.77

Table 2.8.5: Estimates of the Reliability and Variance of District and Within-District School Effects in Milwaukee

	Milwaukee									
	Math 2005-2006					Math 2006-2007				
	3	4	5	6	7	3	4	5	6	7
Variance of VA within-district school effects	129.67	72.76	104.45	40.75	42.64	128.52	91.02	72.82	43.23	88.27
Average variance of statistical estimation error (noise)	15.16	13.04	13.61	8.17	9.98	17.28	17.54	14.45	9.82	10.92
Total variance, including statistical estimation error	144.83	85.80	118.07	48.92	52.62	145.80	108.56	87.27	53.04	99.19
Reliability	0.90	0.85	0.88	0.83	0.81	0.88	0.84	0.83	0.81	0.89
	Read 2005-2006					Read 2006-2007				
	3	4	5	6	7	3	4	5	6	7
Variance of VA within-district school effects	59.08	67.69	74.52	37.14	41.35	68.58	66.51	67.33	36.95	31.09
Average variance of statistical estimation error (noise)	17.26	15.04	16.98	11.29	14.12	17.35	16.27	17.14	14.19	13.54
Total variance, including statistical estimation error	76.35	82.73	91.50	48.43	55.47	85.92	82.78	84.47	51.14	44.63
Reliability	0.77	0.82	0.81	0.77	0.75	0.80	0.80	0.80	0.72	0.70

2.9 Value-Added Productivity Estimates for Milwaukee, Madison and Waunakee

This section presents the estimates of the school/district productivity parameters, district average and consistency of the school/district productivity parameters for Milwaukee, Madison and Waunakee. Due to probable form effects as discussed in Section 2.5, it would be misleading to interpret the "trend" of value-added productivity as reflecting absolute growth of indicators over time. If we assume the state average of value-added does not differ much from one year to another then the reported change in the relative value added would be approximately equal to the change in absolute value-added. Hence school/district value-added indicators can only be used to compare the productivity of schools *relative* to other schools in the same year (for a discussion of the horizontal equating see Section 2.5). Therefore we provide two independent sets of value added productivity estimates for the two growth years.

Figure 2.9.1 and Figure 2.9.2 illustrate the estimates of the school/district productivity parameters for Milwaukee and Madison relative to the state mean in scale score units. The left and right edges of the boxes in the graphs represent 25th and 75th percentile of the school/district productivity parameters in the district. The white dots represent the district averages and the black dots are individual school/district productivity parameters. As seen in the figure, shrinkage estimates of school/district productivity parameters are shrunk toward the estimated district averages. Similarly, shrinkage estimates of district productivity are shrunk toward the state average. As discussed in Section 2.7, the amount of shrinkage is data driven and depends on three factors (a) the estimation error (noise), (b) true variability among the school/district productivity parameters, $\omega^2_{\alpha kl}$, and (c) true variability among the district productivity parameters, ω^2_{δ}.

Figure 2.9.1 and Figure 2.9.2 also illustrates the variability in school/district effects. As seen in the graphs, the school/district productivity parameters are less dispersed from the district average in Madison than Milwaukee.

Figure 2.9.1: Value-Added Productivity Estimates for Milwaukee and Madison, Growth Year 1

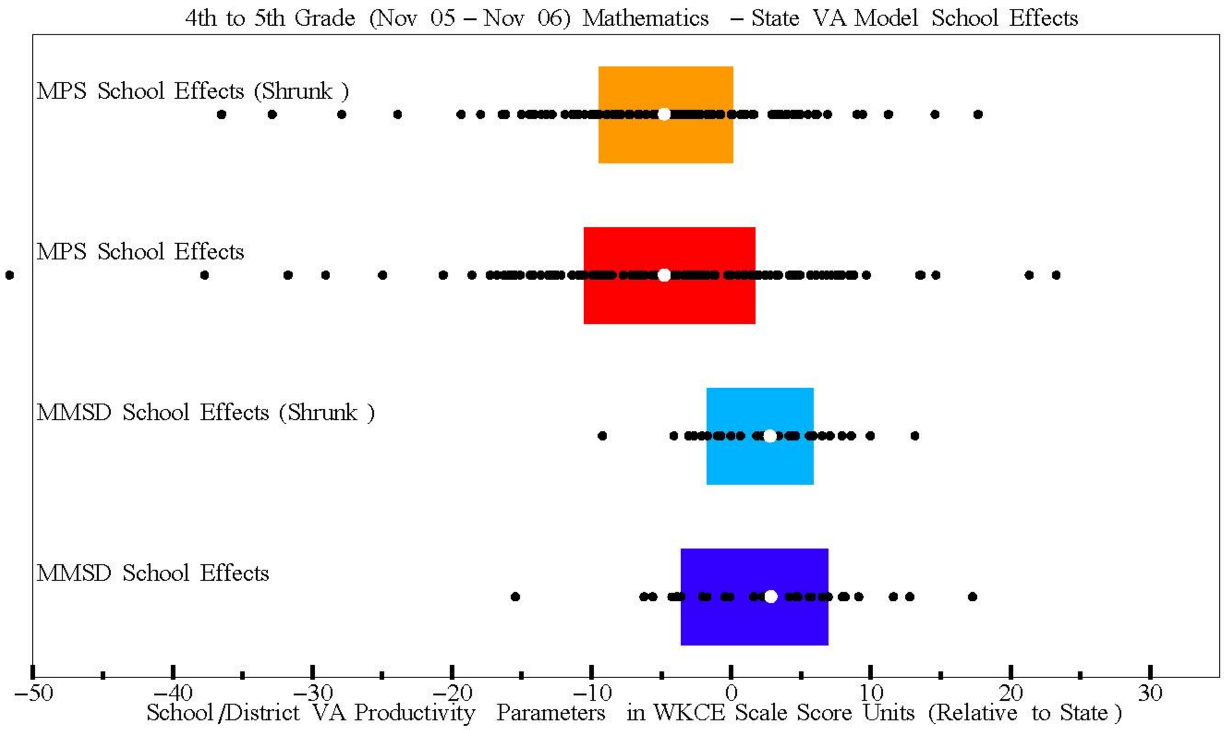
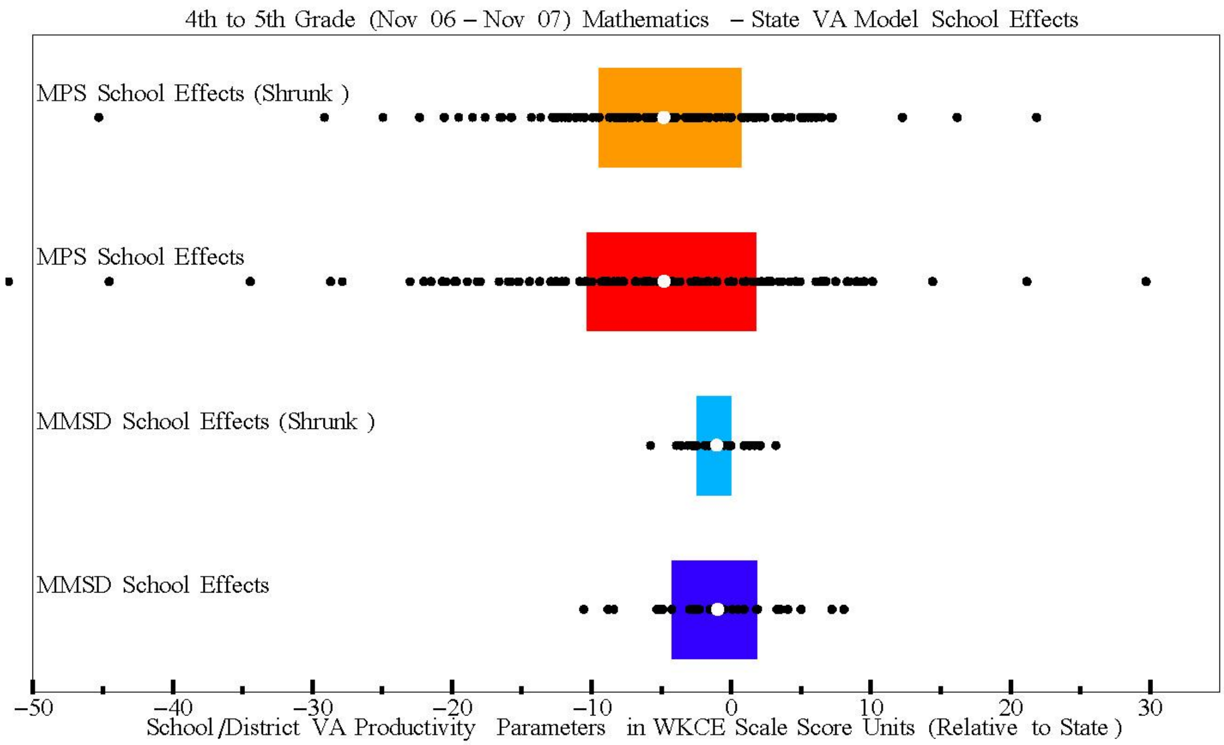


Figure 2.9.2: Value-Added Productivity Estimates for Milwaukee and Madison, Growth Year 2



Estimates from the mathematics value-added model are presented in the first part of the Table 2.9.1; the table reports the district average for the school/district parameters, and the standard error and standard deviation of school productivity within districts. As seen in the table, the district average of Madison is higher than Milwaukee at all grades and in both growth years except from Grade 3-4 in Growth Year 1 for mathematics and Grade 7-8 in Growth Year 1 for reading. For example, everything else equal, an average school in Madison contributes 2.77 more points to a student's mathematics scale score than an average school in the state in Growth Year 1 from Grade 4-5. On the other hand, everything else equal, an average school in Milwaukee contributes 4.79 fewer points to a student's mathematics scale score than an average school in the state in Growth Year 1 from Grade 4-5.

The variability in school/district productivity parameters in Madison is lower than the variability in school/district productivity in Milwaukee at all grades and in both growth years, as seen in Table 2.9.1 in the column named "district standard deviation." Hence Madison is more consistent in school productivity than Milwaukee in the sense that the school/district productivity parameters of Madison's schools are more densely populated around the district average than that of Milwaukee's[16]. Given that average productivity is higher in Madison than in Milwaukee and the variability of school productivity is much lower in Madison than in Milwaukee, it is not surprising that the lowest performing Madison schools tend to have higher productivity than the lowest performing Milwaukee schools (see Figures 2.9.1 and 2.9.2). On the other hand, despite the fact that average productivity is higher in Madison than in Milwaukee, the highest performing Milwaukee schools tend to have higher productivity than the highest performing Madison schools (see Figures 2.9.1 and 2.9.2).

[16] This variance parameter also captures differences in the variance of average classroom effects $\bar{v}_{C.klt}$ across schools and districts. The variance is a function of the number of classrooms per school (at a given grade level) and thus will tend to be relatively low in schools and districts where the number of classrooms per school (at a given grade level) is relatively high.

Table 2.9.1: District Value-Added Effects: Madison and Milwaukee

		Mathematics					
		Growth Year 1			Growth Year 2		
Grade	District	District Average	Standard Error	District Standard Deviation	District Average	Standard Error	District Standard Deviation
3	Madison	-3.48	0.59	3.56	0.78	0.61	8.16
4	Madison	2.77	0.59	5.60	-1.03	0.64	3.06
5	Madison	-0.95	0.60	5.85	3.84	0.59	5.75
6	Madison	0.62	0.50	5.37	2.06	0.51	2.93
7	Madison	2.53	0.58	3.05	0.66	0.43	2.59
3	Milwaukee	-0.66	0.40	11.39	-0.76	0.41	11.34
4	Milwaukee	-4.79	0.37	8.53	-4.83	0.43	9.54
5	Milwaukee	-6.22	0.39	10.22	-5.06	0.39	8.53
6	Milwaukee	-2.64	0.29	6.38	-2.32	0.36	6.57
7	Milwaukee	-0.14	0.39	6.53	-1.69	0.37	9.40

		Reading					
		Growth Year 1			Growth Year 2		
Grade	District	District Average	Standard Error	District Standard Deviation	District Average	Standard Error	District Standard Deviation
3	Madison	0.52	0.61	4.71	-0.49	0.63	4.92
4	Madison	3.36	0.61	3.45	2.59	0.60	4.91
5	Madison	0.90	0.65	5.33	0.82	0.63	2.83
6	Madison	1.01	0.63	5.75	0.91	0.64	2.88
7	Madison	1.35	0.64	5.63	1.32	0.54	4.24
3	Milwaukee	-1.89	0.41	7.69	-5.37	0.41	8.28
4	Milwaukee	-1.99	0.39	8.23	-4.33	0.41	8.16
5	Milwaukee	-4.17	0.42	8.63	-5.13	0.42	8.21
6	Milwaukee	-2.34	0.39	6.09	-2.77	0.43	6.08
7	Milwaukee	1.92	0.42	6.43	0.24	0.34	5.58

Table 2.9.2 reports the low performing schools in Milwaukee and Madison using the tier analyses. The shrunk estimates of the school/district productivity parameters are standardized, i.e. dividing the estimate by the standard deviation of the school/district productivity parameters, ω [17]. A school is then defined as a low performing school if its school/district productivity parameter is less than -0.5. Hence if a school's school/district productivity parameter is 0.5 (or more) standard deviation units lower than state mean (zero), then it is labeled as low-performing school. By definition, about 30% of schools in the state are low performing schools. From Table 2.9.2, we see that the percentage of low performing schools (the ratio of the number of schools labeled as low-performing to number of all schools in that district) is higher in Milwaukee than in Madison. In all but two instances, the percentage of low performing schools in Madison is lower than Milwaukee.

Table 2.9.2: Low Performing Schools: Madison and Milwaukee

		Mathematics			
		Growth Year 1		Growth Year 2	
Grade	District	Number of Low Performing Schools	Percentage of Low Performing Schools	Number of Low Performing Schools	Percentage of Low Performing Schools
3	Madison	16	59.26	8	29.63
4	Madison	2	7.41	3	11.11
5	Madison	10	37.04	2	7.41
6	Madison	3	27.27	1	9.09
7	Madison	0	0.00	2	18.18
3	Milwaukee	49	39.52	52	42.62
4	Milwaukee	73	58.40	71	58.20
5	Milwaukee	75	59.52	70	56.91
6	Milwaukee	49	50.00	59	60.20
7	Milwaukee	26	27.96	48	52.17
		Reading			
		Growth Year 1		Growth Year 2	
Grade	District	Number of Low Performing Schools	Percentage of Low Performing Schools	Number of Low Performing Schools	Percentage of Low Performing Schools
3	Madison	10	37.04	10	37.04
4	Madison	1	3.70	5	18.52
5	Madison	8	29.63	1	3.70
6	Madison	4	36.36	3	27.27
7	Madison	3	27.27	1	9.09
3	Milwaukee	61	49.19	84	68.85
4	Milwaukee	62	49.60	78	63.93
5	Milwaukee	79	62.70	77	62.60
6	Milwaukee	51	52.04	49	50.00
7	Milwaukee	21	22.58	27	29.35

[17] Since the state average of the shrunk estimates is re-centered around zero, dividing shrunk school/district productivity by the standard deviation of the school/district productivity makes their average 0 and variance 1.

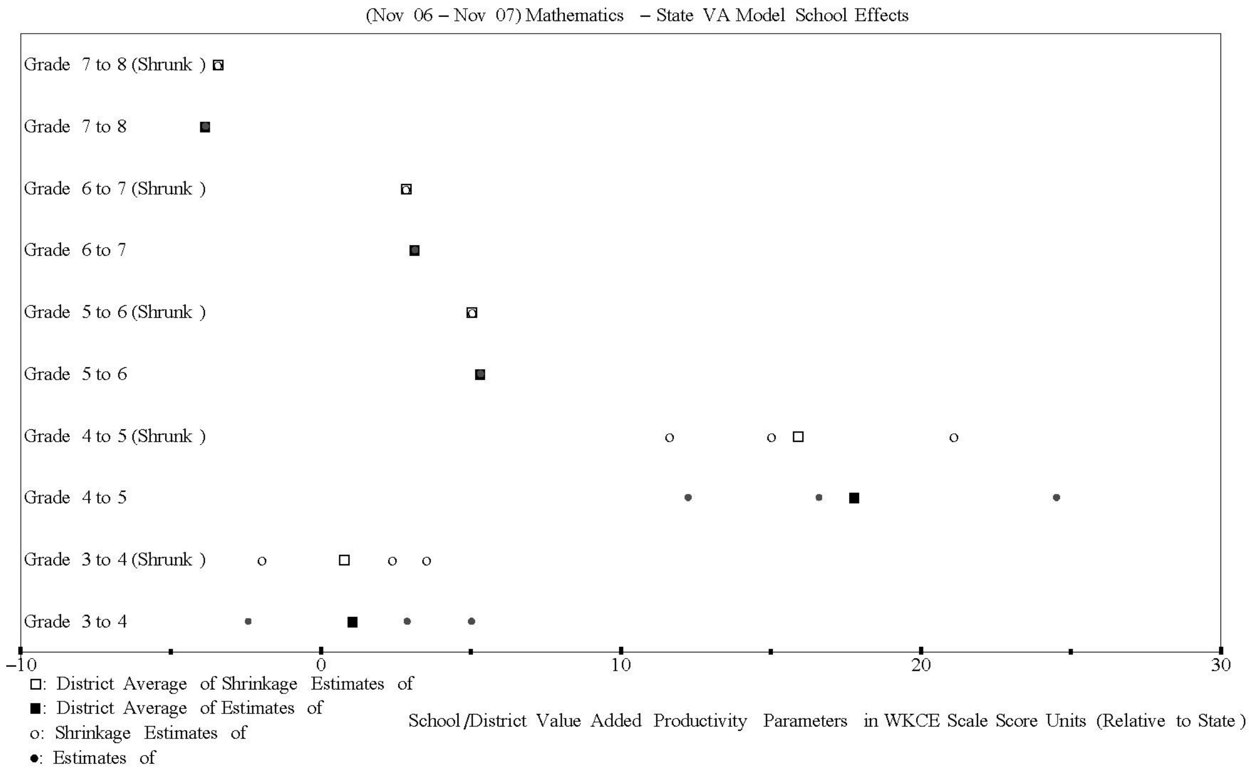
Productivity parameters for all Wisconsin schools and districts (regardless of the number of schools in the district) can be interpreted in the same way as the parameters discussed for Madison and Milwaukee. Below for example, we consider estimates for the Waunakee school district, a district with three schools with third grade students. Parameters from the mathematics value-added model are presented in the first part of the Table 2.9.3; the table reports the district average for the school/district parameters, and the standard error. Figure 2.9.3 illustrates the estimates of school productivity parameters for Waunakee. Black dots are the unshrunk estimates of school productivity parameters and the black square is the district average. White dots are shrinkage estimates of the school productivity parameters and the white square is the district average. As seen in Table 2.9.3 and Table 2.9.1, the district average of Waunakee mathematics' school/district productivity parameter in Growth Year 2 for Grade 3-4 is 0.79 (scale score points and relative to state) and it is approximately same as the district average of Madison mathematics' productivity parameter. It is larger than the district average of Milwaukee mathematics' productivity parameter which is -0.78.

One major difference between large and small districts is that it is not possible to reliably compute the variance of school productivity in given districts for small districts. Thus, we do not report the variance of school productivity for Waunakee.

Table 2.9.3: District Value-Added Effects: A Small District

	Mathematics				Reading			
	Growth Year 1		Growth Year 2		Growth Year 1		Growth Year 2	
Grade	District Average	Standard Error	District Average	Standard Error	District Average	Standard Error	District Average	Standard Error
3	-0.91	1.59	0.79	1.53	-1.36	1.44	-3.20	1.42
4	9.36	1.43	15.93	1.66	0.08	1.34	3.39	1.34
5	5.96	1.51	5.05	1.44	7.74	1.53	5.12	1.47
6	-1.93	1.32	2.85	1.32	1.48	1.39	0.77	1.39
7	-3.91	1.36	-3.42	1.36	2.61	1.45	-2.61	1.45

Figure 2.9.3: Value-Added Productivity Estimates for a Small District, Growth Year 2



2.10 Conclusions and Next Steps

This section identifies technical, policy and logistical conclusions and possible next steps for the project.

Technical Conclusions

A statewide value-added model differs systematically from separate value-added district models. The three important questions that need to be addressed are both challenges and advantages unique to the statewide value-added model. Specifically, these three questions are

1. Does average productivity (as represented by the mean parameter $\bar{\eta}_{.lt}$) differ across districts?
2. Does the consistency of school productivity (as represented by the variance parameter $\omega^2_{\alpha kl}$) differ across districts?
3. Do districts differ systematically in their capacity to create high and low performing schools? (Is $\omega^2_{\delta} > 0$?)

We address these three questions in two sections. First, in Section 2.8, the magnitude, variance, and reliability of the value-added estimates are presented. The numbers presented in the section imply that (a) the variability in school and district productivity is important from a policy perspective, (b) the value-added indicators are estimated with sufficient precision so that it is possible to reliably distinguish between high and low performing schools, and (c) differences across school/district productivity are due to differences in district-wide policies and resources as well as differences in school-level policies and resources. Item (c) is an answer for Question 3, above. Item (b), combined with the results in Section 2.9 answers the first two questions. The estimates from Madison, Milwaukee and Waunakee are presented in Section 2.9. In this section, it is seen that (a) average productivity between these three districts differs, and (b) estimates of variance of school/district productivity parameter are different between Milwaukee and Madison.[18]

Policy Conclusions

Value-added performance indicators allow for the analysis of growth tied directly to the impact of schools and districts. The current practice of using only status and/or simple growth models to analyze student growth limits a school or district's ability to attribute changes in student performance to specific programs or to the instruction provided by teachers, for example. From a public policy perspective, value-added provides the necessary statistical grounding to more accurately determine causal reasons behind low and high performing schools and districts and can more precisely direct schools and districts to an appropriate intervention.

---

[18] The following questions can also readily be answered: 1) What is the value-added productivity of schools within each district? (see the report templates in Appendix C for estimates, see Section 2.2 for a definition) 2) How do district value-added indicator systems best make use of state-level data? (see discussion of three questions unique to the statewide value added model, i.e. Section 2.2, 2.8, 2.9 and 2.10) 3) Can state-level value-added data be used for program evaluation to identify the most effective district and state programs and policies? (see the discussion in Section 2.8.)

While a likely preferred outcome for all districts is for all schools to measure as high performing and for there to be little variance among a district's schools, district officials and policymakers will need the proper tools to guide their districts and schools from all variations (low performing, low variance; low performing, high variance; high performing, high variance; and high performing, low variance) towards reaching and maintaining the goal of being a high performing district with low variation between schools. Clearly district officials and policymakers would not want to sacrifice high performance for a reduction in variance. At the same time, high average district performance with high variance – meaning that some schools are extremely low performing – is not acceptable. Managing this balancing act and making recommendations for best practices for using value-added data will be a subject of further study.

Technical Next Steps

For the sake of simplicity and to emphasize the importance of the variability in variance in school productivity between districts, three groups of districts are classified: Milwaukee, Madison, and the remainder of the state. Each of these groupings has its own variance estimate of school productivity parameters. What follows are potential next steps to improve the model:

- Grouping methods can be improved by relying on similarities that might be related to instructional practices.
- Several statistical techniques from shrinkage literature might be applied to noisy individual estimates of the variance of school/district productivity parameters in small districts; this would enable more precise estimates to be obtained for smaller districts.
- Modeling the variance of the districts by using the characteristics of districts is another option.

We believe that a combination of these possible extensions of variance models will result in a richer, more policy relevant model that will guide future work and become an important part of the statewide value-added models.

Finally, from a technical and public policy perspective it is important to investigate why the estimates of demographic coefficients and value-added productivity parameters might change, more than one would reasonably expect, from one year to another and/or from one grade to the next. Future work will aim to distinguish the genuine changes in parameters from random, or lucky, changes; and later, to control for the random, or lucky, changes in parameters so that the noise in change pattern is minimized.

Logistical Next Steps

Over the course of the project VARC staff have met with representatives from the CESAs, specifically CESA #1, CESA #2 (a project partner), CESA #6, and the CESA School Improvement Services (SIS) group. The purpose of these meetings has been to provide district administrators and CESA staff with a general technical understanding of value-added and to facilitate a discussion about the policy implications and practical implications of bringing a statewide value-added system into production.

CESA #2 and VARC have developed a pilot model for generating value-added metrics for Wisconsin districts and schools and providing the necessary professional development for district staff. In this model VARC will generate the value-added metrics using WKCE data. CESA #2 will deliver the value-added metrics to districts wanting to participate and will provide professional development around how to interpret and use the data. Technical assistance for interpreting the data will be provided by VARC. While CESA #2 will be responsible for delivering value-added professional development to their member districts, VARC staff will assist in both the development and delivery of the professional development.

The impetus for this pilot model originated with the three categories of responses identified by participants at the January 2008 planning meeting for this project – see the bottom of page 3.

1. What is the value-added productivity of schools within each district?
2. How do district value-added indicator systems best make use of state-level data?
3a. Can state-level value-added data be used for program evaluation?
3b. Can state-level value-added data be used to identify the most effective district and state programs and policies?

Building from these categories, a significant portion of the professional development provided by CESA #2 will include guidance around how to use value-added data within a consortium of districts or within an individual school district. Districts may first learn how to use the data to identify schools that are consistently low or high performing based on their value-added data. What more, for example, are district or school officials able to learn about student learning from value-added data, compared to basic WKCE data?

Once these schools are identified districts may decide to direct extra resources to the low performing schools or decide to study the high performing schools to see if their practices might be implemented within other schools. Districts will also receive guidance around how to use value-added data with other available data. How, for example, would a district use their value-added data in conjunction with their WKCE data or classroom assessments to identify areas needing extra support?

As a final example, large districts or a consortium of districts may decide to provide additional data to allow for program evaluation. Milwaukee, for example, has used value-added data to evaluate the implementation of the Read 180 program. Similarly, a group of smaller districts using a certain math program may decide to pool their data to be analyzed under a value-added lens. The professional development program will include a discussion of how value-added can be used to evaluate locally implemented programs to statewide initiatives, provided the data is available.

Appendix A. Technical Description of "Core" Value-Added Model

In this appendix we briefly describe a "core" value-added model for a given grade that includes the features listed below[19]. Most, if not all, value-added models produce value-added parameters of the type included in this model.

- School/district value-added productivity effects η_{klt}.
- Statewide value-added productivity effects π_t.
- Two years of longitudinal attainment data for each student.
- A posttest on pretest parameter λ_t. This parameter allows for situations where the variances of the posttest and pretest variables may be atypical.
- Control for measurement in prior achievement Y_{1it-1}.
- Demographic variables X_{it} to capture differences across students in achievement growth.

The core value-added model is defined by the following equation:

$$Y_{2it} = \xi + \lambda_t Y_{1it-1} + \pi_t + \beta_t' X_{it} + \sum_k \sum_l \eta_{klt} S_{iklt} + \varepsilon_{it} \qquad (1)$$

where the variables, parameters, and indices in the model are defined in Appendix Table A.1 and the grade descriptors are omitted, for simplicity. The model is structured so that the value-added effect η_{klt} captures educational productivity for the pretest grade since the period from the date of the assessment (mid-November) to the end of the school year (typically early June) encompasses most of the grade.

[19] Note that by aggregating grade level school/district value added productivity parameters, one can obtain the school level school/district productivity parameters.

Table A.1: Variables and Parameters in Value-Added Model

i	Student identifier
k	Within-district school identifier
l	District identifier
t	Year of posttest score
g	Grade (not explicitly included in above model)
Y_{2it}	Posttest score in year t
Y_{1it-1}	Pretest score in year *(t-1)* (prior year)
X_{it}	Student demographic characteristics (vector)
S_{iklt}	Student indicator, or fractional measure of enrollment, in school k, in district l, in year t
λ_t	Coefficient on pretest score
β_t	Coefficient (vector) for demographic characteristics
ξ	Intercept
ε_{it}	Student level error component

See Table 2.2.1 in the main report for a definition of the value-added parameters in this model.

This value-added model is an example of a "T2" value-added model in that it requires two years of longitudinal attainment data for each student and thus is sufficient to measure growth in student achievement over a single growth year. In this report we also discuss results based on models based on two and three years of longitudinal data.[20] Note that all parameters are allowed to vary by year, including the slope parameters λ_t and β_t.

The Coefficient on Prior Achievement: λ_{gt}

One of the important features of the value-added model considered above is that it allows for the possibility that the coefficients on prior achievement (λ_{gt}) may differ across grades and years and may not equal one, a parameter restriction that is imposed in some value-added models. The model would be simpler to estimate if it was appropriate to impose the parameter restriction $\lambda = 1$, but there are at least four factors that could make this restriction invalid. First, λ could be less than one if the "stock" of knowledge, skill, and achievement captured by student assessments is not totally "durable," but rather is subject to decay. Let δ = annual durability rate of student achievement so that the annual decay rate equals $(1-\delta)$ (Meyer, 2006).

Second, λ could differ from one if school resources are allocated differentially to students as a function of true prior achievement, as captured by a resource allocation parameter ρ. If resources are tilted relatively toward low achieving students – a remediation

[20] Note that since statewide testing begins in third grade in Wisconsin and most other states, only two years of (up-to-date) attainment data are available to estimate value-added models of achievement growth from third to fourth grade.

strategy – then $\rho < 1$. The opposite would be true if resources were tilted toward high achieving students (Meyer, 2006). Combining these two factors yields a coefficient on prior achievement equal to:

$$\lambda^* = \delta + \rho \tag{2}$$

Third, λ could differ from one (or from λ^*, as defined above) if posttest and pretest scores are measured on different scales, perhaps because the assessments administered in different grades are from different vendors and scored on different test scales or due to instability in the variability of test scores across grades and years. In this case, the coefficient on prior achievement partially reflects the difference in scale units between the pretest and posttest. Fourth, the different methods used to score assessments could, in effect, transform posttest and pretest scores so that the relationship between post and prior achievement is nonlinear. In this case, a linear value-added model might still provide a reasonably accurate approximation of the achievement growth process, but the coefficient on prior achievement (as in the case of the third point) is affected by the test scaling.

To see this, consider a value-added model with the same structure as equation (1), defined in terms of latent unobserved test scores z_2 (true latent post achievement) and z_1 (true latent prior achievement)[21]:

$$z_{2it} = \xi^* + \lambda_t^* z_{1it-1} + \pi_t^* + \beta_t^{*\prime} X_{it} + \sum_k \sum_l \eta_{klt}^* S_{iklt} + \varepsilon_{it}^* \tag{3}$$

The parameters in this model are distinguished from the parameters of equation (1) by the superscript "*." True scores (measured without error), corresponding to measured pretest and posttest scores, are given by (possibly nonlinear) transformations of the latent achievement scores:

$$\begin{aligned} y_{1it-1} &= f_{1t-1}(z_{1it-1}) \\ y_{2it} &= f_{2t}(z_{2it}) \end{aligned} . \tag{4}$$

Note that the transformation functions could be the same, if the properties of the scoring/scaling algorithm are similar at different grades and years. Latent prior achievement is correspondingly given by:

$$z_{1i} = f_{1t-1}^{-1}(y_{1i}) \tag{5}$$

where f_{1t-1}^{-1} is the inverse function.

[21] In this section we define the model in terms of test scores measured without error in order to simplify the discussion. At the estimation stage, it is necessary to employ an estimation procedure that addresses measurement error in prior achievement.

Given equations (3) – (5), the value-added model, written in terms of true scores y_{2i} and y_{1i} is given by:

$$y_{2it} = f_{2t}\left[\xi^* + \lambda^* f_{1t-1}^{-1}(y_{1it-1}) + \pi_t^* + \beta_t^{*\prime} X_i + \sum_k \sum_l \eta_{klt}^* S_{iklt} + \varepsilon_{it}^*\right] \tag{6}$$

This model is not actionable since the transformation functions are unknown, but it can be approximated by a linear (Taylor series) approximation around the district (or state) means of the regressors:

$$y_{2it} = \bar{f}_{2t} + \left(\frac{m_{2t}}{m_{1t}}\right)\lambda_t^*(y_{1it} - \bar{\bar{y}}_{1t}) + m_{2t}\left[\pi_t^* + \beta_t^{*\prime}(X_{it} - \bar{\bar{X}}_t) + \sum_k \sum_l \eta_{klt}^* S_{iklt} + \varepsilon_{it}^*\right] \tag{7}$$

where $\bar{\bar{y}}_{1t}$ = the district mean of prior achievement in year *t*, $\bar{\bar{X}}_t$ = the district mean of student characteristics in year *t*, $\bar{f}_{2t}$ = the function f_{2t} evaluated at district means, and $m_{2t} \equiv \bar{f}_{2t}'$ and $m_{1t-1} \equiv \bar{f}_{1t-1}'$ (scale multipliers) are the first derivatives of the functions f_{2t} and f_{1t-1} evaluated at district means, respectively.[22] The key result is that equation (7) is equivalent to the linear value-added model defined by equation (1); the parameters of this model adjust in response to the particular scaling algorithms used to score/scale assessments. In particular, the coefficient on prior achievement is the product of the ratio of two scale multipliers (m_{2t} / m_{1t-1}) and the latent pretest coefficient, and the latent school performance effect and the other components of the model are multiplied by the posttest multiplier m_{2t}:

$$\lambda_t = \left(\frac{m_{2t}}{m_{1t-1}}\right)\lambda_t^* \tag{8}$$

$$\begin{aligned} \pi_t &= m_{2t}\pi_t^* \\ \eta_{klt} &= m_{2t}\eta_{klt}^* \\ \beta_t &= m_{2t}\beta_t^* \\ \varepsilon_{it} &= m_{2t}\varepsilon_{it}^* \end{aligned} \tag{9}$$

The degree to which the latent parameters are affected by implicit scale transformation depends on the shape of the scaling functions f_t. The bottom line is that the parameters of a value-added model are not invariant to the scaling algorithms used to score the pretest and posttest assessments.

In summary, we have considered four factors that could make it problematic to impose the parameter restriction that the coefficient in prior achievement are identical in all grades and years and equal to particular value (such as $\lambda = 1$): durability/decay in achievement, differential resource allocation, differences in the pretest and posttest test scales, and nonlinearity in the test scaling algorithm.

[22] The district means of the other variables are zero, given the normalizations used in the model.

One of the major implications of the above analysis is that the units of pretest and posttest scores may differ from a hypothetical ideal situation due to explicit differences in the two scales or unintended nonlinearities in test scaling to irregularities. It may be possible to detect whether the units of pretest and posttest scores differ from a hypothetical ideal situation by using information on the standard deviation of pretest and posttest scores to interpret the estimated coefficients on prior achievement. To do so, we note that the test scale multipliers defined above can be expressed as ratios of the standard deviation of measured achievement s_{gt} to the (unobserved) standard deviation of achievement on the hypothetical scale σ_{gt}:

$$\begin{aligned} m_{2gt} &= s_{gt} / \sigma_{gt} \\ m_{1g-1,t-1} &= s_{g-1,t-1} / \sigma_{g-1,t-1} \end{aligned} \tag{10}$$

Substituting these expressions into equation (8) yields:

$$\lambda_{gt} = M_{gt} \lambda_{gt}^{**} \tag{11}$$

where M_{gt} = the ratio of the standard deviations of measured achievement:

$$M_{gt} = \frac{s_{gt}}{s_{g-1,t-1}} \tag{12}$$

and λ_{gt}^{**} is given by:

$$\lambda_{gt}^{**} = \left(\frac{\sigma_{g-1,t-1}}{\sigma_{gt}} \right) \lambda_{gt}^{*} \tag{13}$$

Note that λ_{gt}^{**} is equal to the parameter that would be estimated if all posttest and pretest scores were transformed to scales with identical standard deviations. Both M_{gt} and λ_{gt}^{**} can be calculated given estimates of λ_{gt}. We conjecture that λ_{gt}^{**} may vary only slightly across grades and years. If so, plots of estimates of λ_{gt}^{**} with respect to grade or year should be relatively "smooth." Estimates of λ_{gt}. on the other hand may be quite unstable due to the scaling issues discussed above. The ratio of test score standard deviations M_{gt} provides a direct measure of scaling irregularities as they affect the variability of test scores. Empirical evidence on the issue of scaling stability is presented later in this report.

Appendix B. The Average Change in Statewide Gain is Approximately Equal to the Change in Statewide Value-Added Productivity

Using the value-added model defined in Section 2.2, the average statewide gain from year *(t-1)* to *(t)* at a given grade level is given by:

$$G_t = \bar{Y}_{2.t} - \bar{Y}_{1.t-1} = \xi + \pi_t + (\lambda_t - 1)\bar{Y}_{1.t-1} + \beta_t' \bar{X}_{.t} \tag{1}$$

where the bar over each variable (and the dot replacing the *i* index) signifies that the variable is a state mean. The change in statewide gain from posttest year *s* to *t* is similarly given by:

$$\Delta_{st} = G_t - G_s = (\pi_t - \pi_s) + (C_t - C_s) \tag{2}$$

where C_t, a cohort variable, is defined as:

$$C_t = (\lambda_t - 1)\bar{Y}_{1.t-1} + \beta_t' \bar{X}_{.t} \tag{3}$$

The cohort variables will typically not change much from year to year so that the change in statewide gain approximately equals the change in statewide productivity, as asserted:

$$\Delta_{st} = G_t - G_s \approx (\pi_t - \pi_s) \tag{4}$$

The change in statewide gain in tier units is obtained by dividing gain Δ_{st} by the standard deviation of school productivity in the baseline year ω.

Appendix C. Understanding School and District Value-Added Reports

As part of the Wisconsin Value-Added Demonstration Project, district and school report templates were developed to provide value-added information to practitioners in an easily digestible format. Templates were developed for examples of large (Milwaukee Public Schools – MPS, Appendix C.1), medium (Madison Metropolitan School District – MMSD, Appendix C.2) and small (Waunakee Community School District – WCSD, Appendix C.3) school districts. For each of the three districts one elementary school was randomly selected. Value-added data is presented in tabular and graphic formats for each WCKE tested grade-level at the selected school. Note that at MPS and MMSD both school and district level data is presented for grades 3, 4 and 5, but for WCSD school data is not presented for grade 5 due to the selected elementary school being a K-4 school.

While there is a separate template for examples of each of the three size categories of schools – small, medium and large, there is no difference is how the value-added data is reported. Each series of reports begins first with a tabular presentation of the school and district data, followed by a graphical presentation of district value-added (by grade-level) against the state average, a graphical presentation of school value-added (by grade-level) against the state average, a graphical presentation of district value-added and attainment data (by grade-level) against the state averages, and finally a graphical presentation of school value-added and attainment data (by grade-level) against the state averages.

Following the presentation of the district and school report templates, Appendix C.4 contains an example of an individual student growth report.

The information below is provided as guidance for how best to utilize information in the reports.

What is Value-Added Data?

A value-added model is another type of growth model. It is distinguished from simpler types of growth models by the use of statistical methods. The aim of the statistical analysis is to isolate the effect of a particular district, school, grade or classroom on student achievement. In other words, value-added is a way of recognizing the contribution a school has made to the student’s academic growth.

By controlling for the following factors, value-added analysis allows for an “apples to apples” comparison to be made between schools. For example, a district might look at the performance of two schools with very different demographics but be able to compare their growth as if their demographics were similar.

Reference Points

For both the reference and analysis sections, the following comparisons are made in terms of scale scores:

District Compared to State

- A comparison of the district's value added in scale score points to the state value-added average on the horizontal axis. The district's state percentile rank is shown by the data point on the vertical axis.

School Compared to State

- A comparison of the school's value added in scale score points to the state value-added average on the horizontal axis. The school's state percentile rank is shown by the data point on the vertical axis.

For What Factors Does Value-added Analysis Control?

Value-added analysis controls for the demographic composition of schools (for example, economic status, race/ethnicity, gender, disability status). These controls allow for fairer growth comparisons to be made. Controlling for demographic factors make possible the measurement of differences in growth across demographic groups.

What Can You do With Value-added Data?

- Identify high-performing districts and schools
- Identify districts and schools in need of improvement
- Identify trends that are evident over time by district or school
- Identify where incentives should be directed, such as schools, individuals or teams

What are "Confidence Intervals"?

The confidence intervals presented are similar to the ones you see with poll results. We are 95% certain that the "true" amount of achievement growth lies within the confidence interval bands. The larger the sample size (number of student scores included in the calculation), the shorter the confidence interval band (less error). The smaller the sample size, the longer the confidence interval band (more error). The lesser the variability in student scores, the smaller the confidence interval band; the greater the variability in student test scores, the larger the confidence interval band. For example, a value-added metric based on the growth of five hundred students will be more precise than one based on the growth of five students. Additionally, the greater the concentration of student scores around the average, the more precise the metric.

What is the Rationale for Quadrant (Value-added Data Plotted Against Attainment Data) Analysis?

A useful application of value-added data is to compare it in relation to the percentage of students who score proficient or advanced on the WKCE (attainment). This type of comparison results in a school or district falling into 1 of 4 different quadrants.

Low Value Added High Attainment (3)	High Value Added High Attainment (1)
Low Value Added Low Attainment (4)	High Value Added Low Attainment (2)

Placement of schools or districts into one of the four quadrants allows for a quick understanding of each school's growth and attainment. Schools in Quadrant 1 (high value added, high attainment) are both above average in growth, and above average in attainment. Schools in Quadrant 2 (high value added, low attainment) are experiencing high growth but are still low attaining schools. Schools in Quadrant 3 (low value added, high attainment) are experiencing lower than average growth but are still high attaining schools. Schools in Quadrant 4 (low value added, low attainment) are both below average in growth, and below average in attainment.

Appendix C.1 Milwaukee: School and District Reports

School - Meir

Mathematics	Value Added				Attainment	
School Year 2006-07	Value Added	Confidence Interval	N	State Percentile	WCKE % Proficient	Quadrant
Grade 3	2.88	-2.78 to 8.54	91	64.9	87.9	1
Grade 4	2.40	-3.34 to 8.14	90	64.0	86.7	1
Grade 5	-5.81	-10.96 to -0.66	89	24.1	78.7	3
School Year 2005-06	Value Added	Confidence Interval	N	State Percentile	WCKE % Proficient	Quadrant
Grade 3	-4.33	-9.5 to 0.84	90	28.0	91.1	3
Grade 4	0.44	-4.62 to 5.5	92	54.1	85.9	1
Grade 5	-3.07	-8.15 to 2.01	93	35.5	87.1	3

Reading	Value Added				Attainment	
School Year 2006-07	Value Added	Confidence Interval	N	State Percentile	WCKE % Proficient	Quadrant
Grade 3	3.08	-2.6 to 8.76	91	67.4	94.5	1
Grade 4	0.24	-5.31 to 5.79	90	50.8	95.6	1
Grade 5	7.12	1.49 to 12.75	89	86.1	94.4	1
School Year 2005-06	Value Added	Confidence Interval	N	State Percentile	WCKE % Proficient	Quadrant
Grade 3	3.98	-1.55 to 9.51	90	72.7	97.8	1
Grade 4	-0.76	-6.19 to 4.67	92	44.5	95.7	3
Grade 5	1.44	-4.24 to 7.12	93	59.7	96.8	1

District - Milwaukee

Mathematics	Value Added				Attainment	
School Year 2006-07	Value Added	Confidence Interval	N	State Percentile	WCKE % Proficient	Quadrant
Grade 3	-0.79	-1.69 to 0.11	5333	44.3	46.4	4
Grade 4	-4.81	-5.71 to -3.91	5175	29.6	52.2	4
Grade 5	-5.07	-5.89 to -4.25	5143	26.1	45.9	4
School Year 2005-06	Value Added	Confidence Interval	N	State Percentile	WCKE % Proficient	Quadrant
Grade 3	-0.70	-1.54 to 0.14	5103	44.3	43.3	4
Grade 4	-4.81	-5.59 to -4.03	5624	25.8	43.8	4
Grade 5	-6.25	-7.07 to -5.43	5436	19.8	40.3	4

Reading	Value Added				Attainment	
School Year 2006-07	Value Added	Confidence Interval	N	State Percentile	WCKE % Proficient	Quadrant
Grade 3	-5.59	-6.49 to -4.69	5333	14.4	64.1	4
Grade 4	-4.38	-5.26 to -3.5	5175	19.1	63.7	4
Grade 5	-5.34	-6.24 to -4.44	5143	12.7	65.3	4
School Year 2005-06	Value Added	Confidence Interval	N	State Percentile	WCKE % Proficient	Quadrant
Grade 3	-2.00	-2.9 to -1.1	5103	31.6	63.1	4
Grade 4	-1.92	-2.76 to -1.08	5624	40.6	61.7	4
Grade 5	-4.24	-5.14 to -3.34	5436	20.5	62.4	4

Note: Value Added is derived from the WCKE tests held each November. School Year 2005-06 Value Added is derived from the November 2005 and November 2006 WKCE tests; School Year 2006-07 Value Added is derived from the November 2006 and November 2007 WCKE tests.

3rd Grade value-added measures growth from the November 3rd Grade test point through the November 4th Grade test point.
4th Grade value-added measures growth from the November 4th Grade test point through the November 5th Grade test point.
5th Grade value-added measures growth from the November 5th Grade test point through the November 6th Grade test point.

Report Key

Value Added

- This value is a measure that is equal to the number of extra points scored by students in a school or district (that is attributable to the school or district) on a test relative to observationally similar students across the state or district. For example, if a school's value added is +3, then students at the school scored three points higher on the test then similar students across the state. If a school's value added is -3, then students at the school scored three points lower relative to similar students. An average school or district will have a relative value added of zero when compared to the state.

Confidence Interval

- This reflects the range of possible value-added scores. As the number of students measured (n) increases, the confidence interval generally decreases – note the smaller District-level confidence intervals versus the School-level confidence intervals.

N

- Number of students tested at a school.

State Percentile

- The value-added state percentile rank for a school or district.

Attainment

- This is the percentage of students scoring proficient or advanced on the WKCE for a school or district.

Quadrant

- This is a value from 1 to 4 based upon a combination of value-added and attainment data for the school or district.

Graphical Presentation

A graphical presentation of each school and district's value-added data follows.

District Compared to the State 2005-06
District Compared to the State 2006-07
School Compared to the State 2005-06
School Compared to the State 2006-07

For both the reference and analysis sections, the following comparisons are made in terms of scale scores:

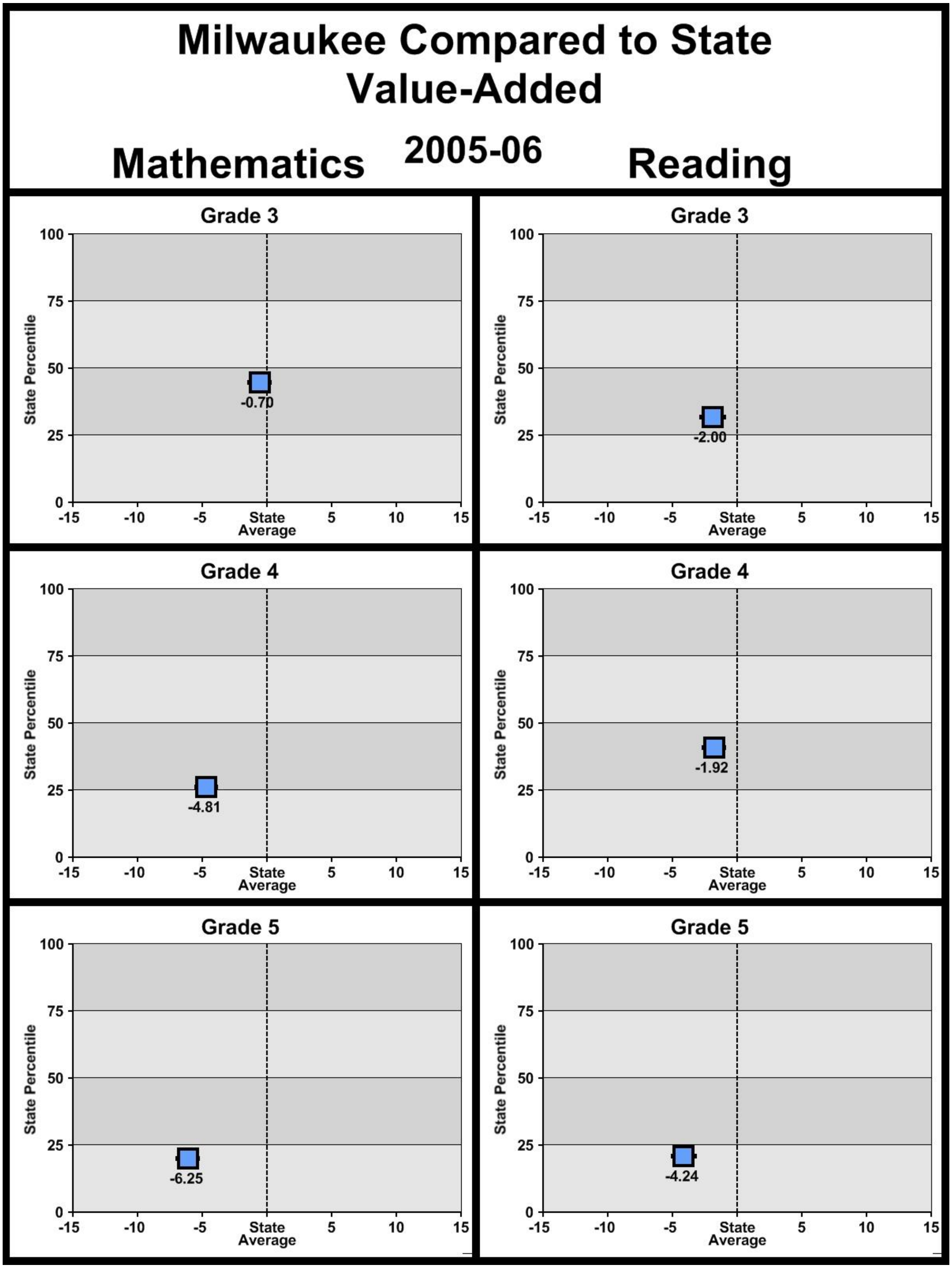
District Compared to State

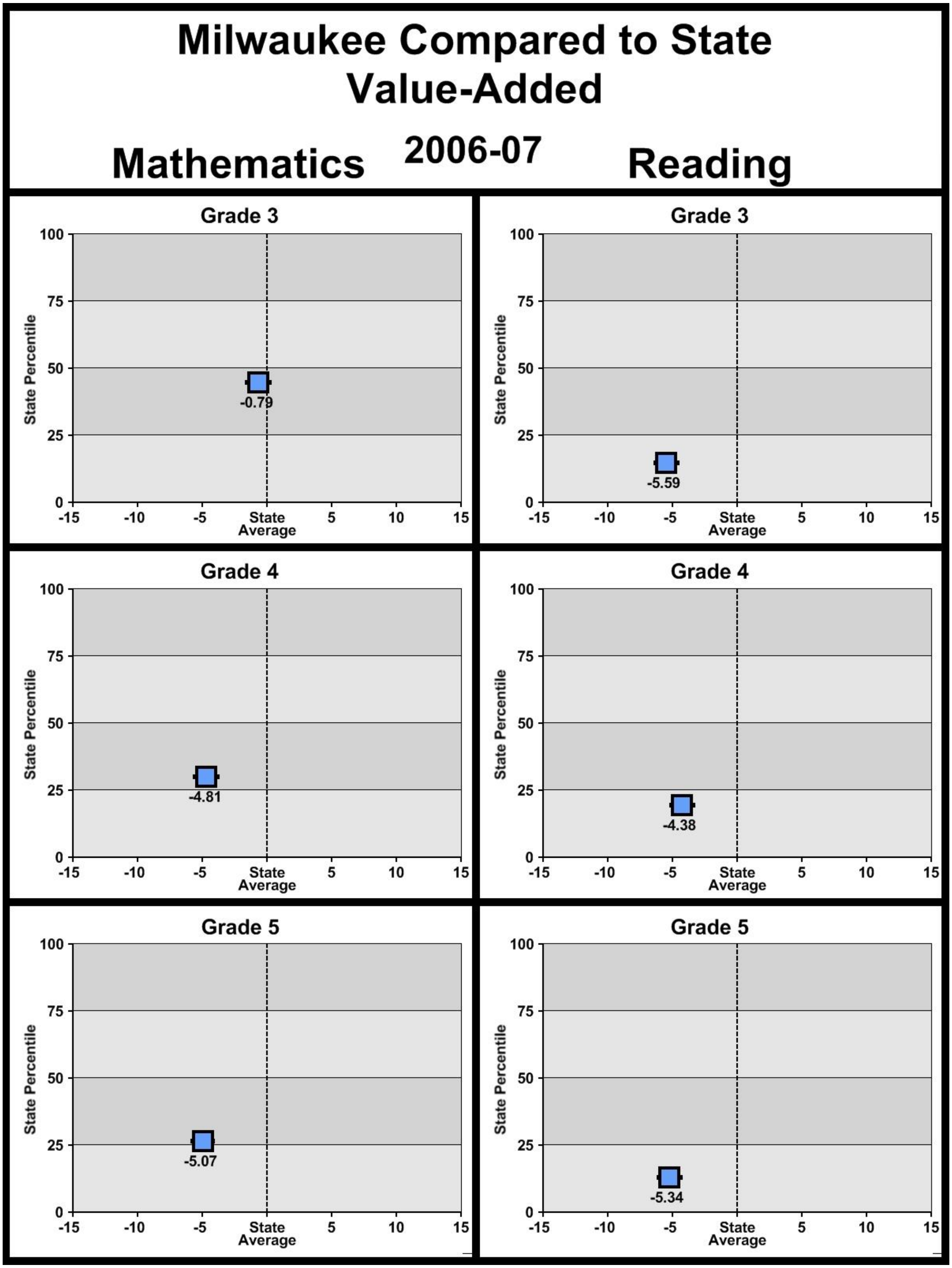
- This graphic shows the district's value added in scale score points compared to the state value-added average on the horizontal axis. The district's state percentile rank is shown by the data point on the vertical axis.

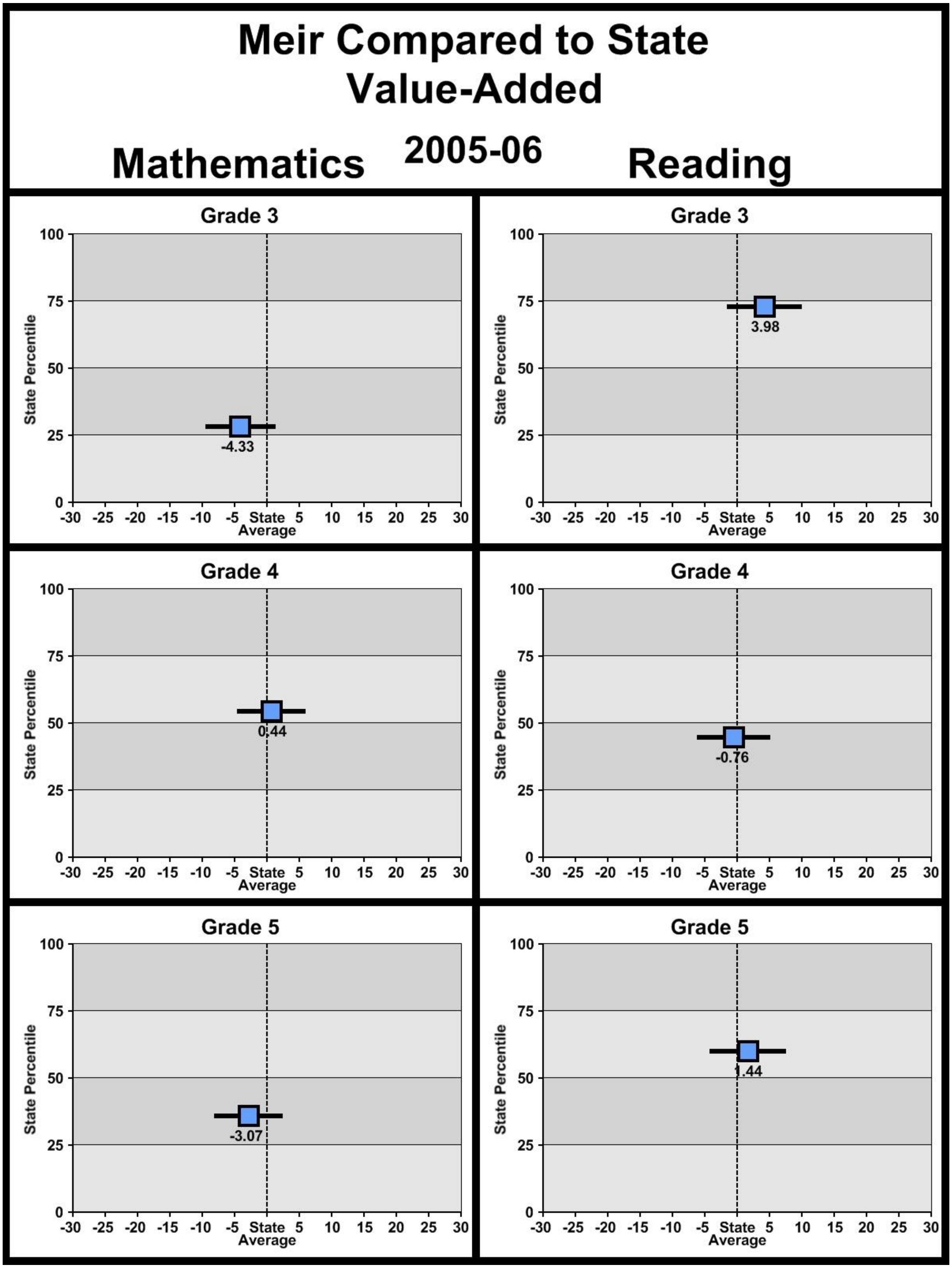
School Compared to State

- This graphic shows the school's value added in scale score points compared to the state value-added average on the horizontal axis. The school's state percentile rank is shown by the data point on the vertical axis.

Each comparison includes a column for mathematics and a column for reading. Each column has three rows providing value-added data for each grade third through fifth.

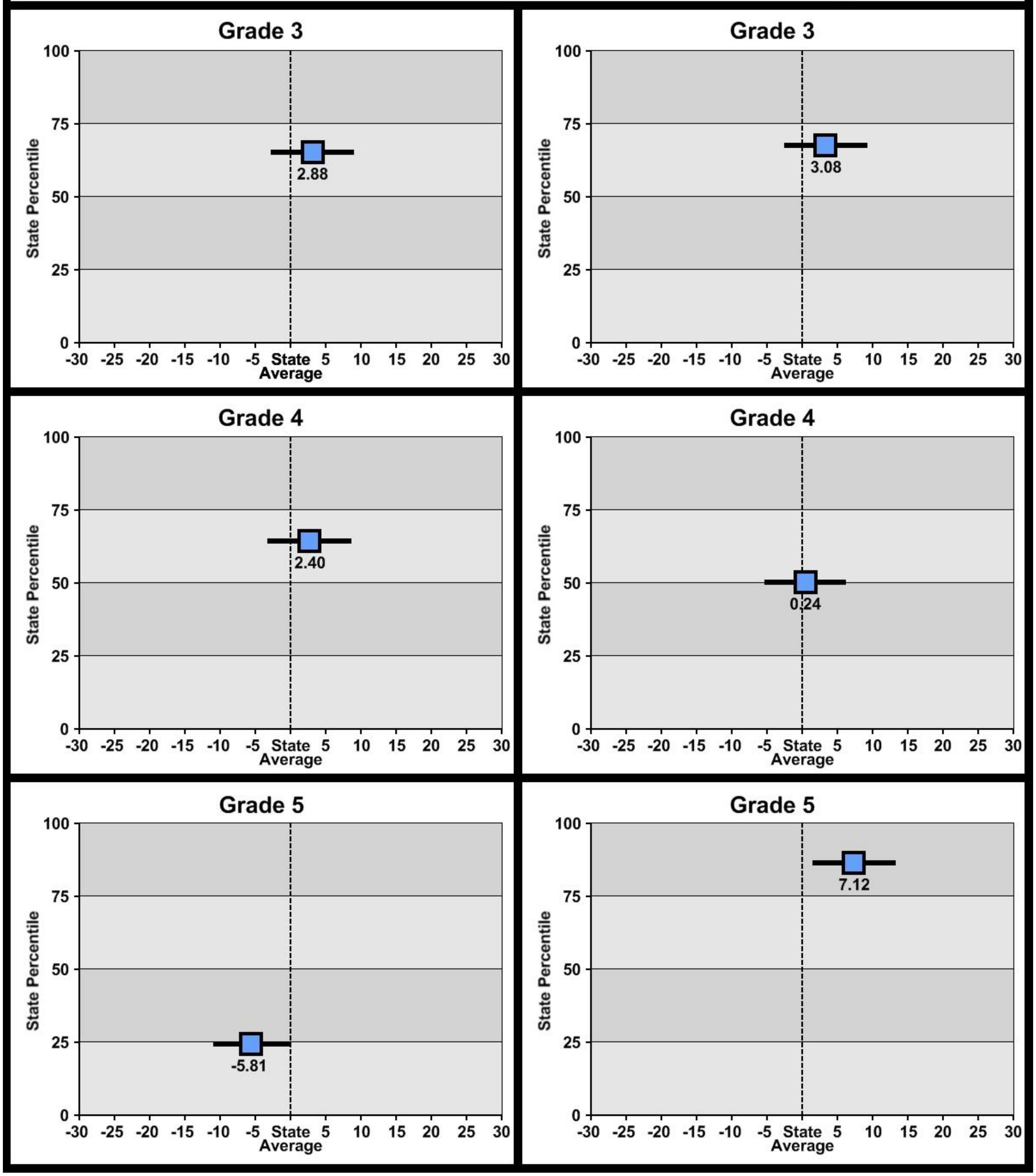

Milwaukee Compared to State
Value-Added
Mathematics
2005-06
Reading
Grade 3
State Percentile
-0.70
-2.00
Grade 4
-4.81
-1.92
Grade 5
-6.25
-4.24
100
75
50
25
0
-15
-10
-5
State Average
5
10
15


Milwaukee Compared to State
Value-Added
2006-07
Mathematics
Reading
Grade 3
-0.79
Grade 3
-5.59
Grade 4
-4.81
Grade 4
-4.38
Grade 5
-5.07
Grade 5
-5.34
State Percentile
State Average


Meir Compared to State
Value-Added
Mathematics
2005-06
Reading
Grade 3
State Percentile
100
75
50
25
0
-30 -25 -20 -15 -10 -5 State 5 10 15 20 25 30
Average
-4.33
3.98
Grade 4
0.44
-0.76
Grade 5
-3.07
1.44

Meir Compared to State
Value-Added

Mathematics 2006-07 Reading



Analysis Section

The information below will guide your interpretation of the following plots:

- District / State Averages, Value-Added and Attainment Sectors
- School / State Averages, Value-Added and Attainment Sectors

- Graphic Layout
    - Each data point represents a school or district and is determined by plotting a school or district's value-added score against the school or district's percent proficient/advanced on the WKCE (attainment). Your district (District / State Averages) or school (School / State Averages) is represented by a red diamond .
    - The state average for both value added and percent proficient/advanced provides the structure upon which the four quadrants are distinguished.
    - Schools and districts fall into one of the four different quadrants.
    - The gray shaded areas above and below, and to the left and the right of the state average lines represent one standard deviation away from that line.
    - The AMO (Annual Measurable Objectives), a measurement used to determine compliance with the federal No Child Left Behind Act (NCLB), is included.

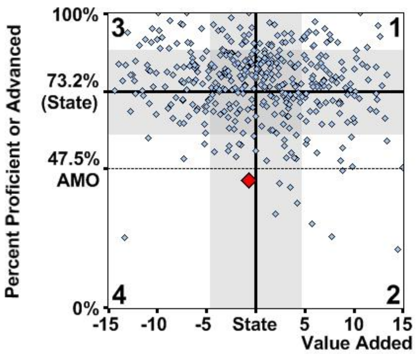
- Quadrant Analysis
    - Schools, or (districts) in Quadrant 1 (high value added, high attainment) are both above average in growth and above average in attainment.
    - Schools, or (districts) in Quadrant 2 (high value added, low attainment) are above average in growth and below average in attainment.
    - Schools, or (districts) in Quadrant 3 (low value added, high attainment) are below average in growth and above average in attainment.

o Schools, or (districts) in Quadrant 4 (low value added, low attainment) are both below average in growth and below average in attainment.

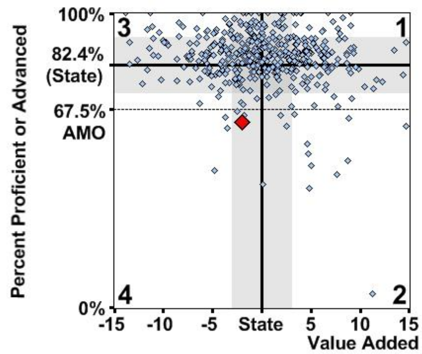
Milwaukee / State Averages
Value-Added and Attainment Quadrants

Mathematics 2005-06 Reading

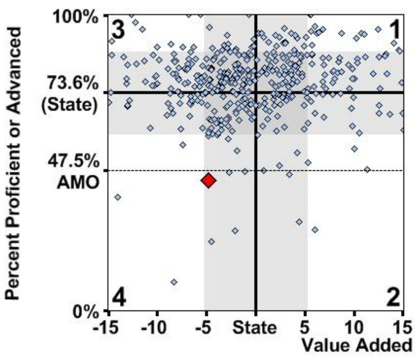
Grade 3



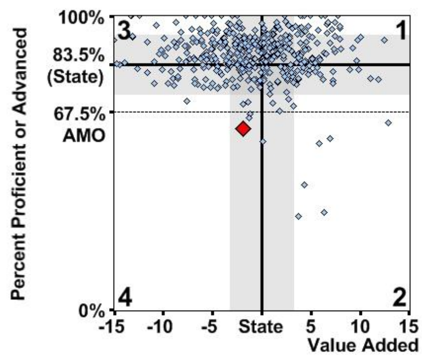
Grade 3



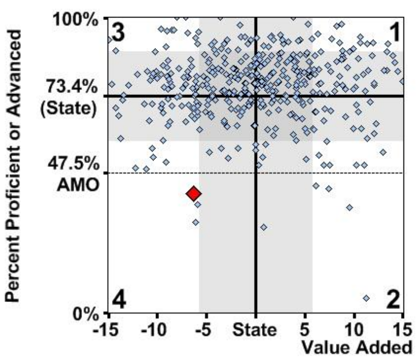
Grade 4



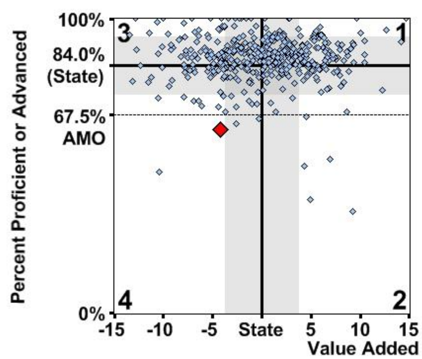
Grade 4



Grade 5



Grade 5

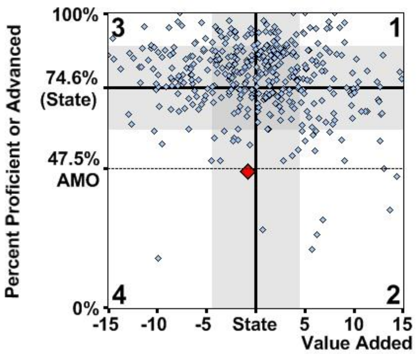


Quadrant 1 - High Value Added, High Attainment
Quadrant 2 - High Value Added, Low Attainment
Quadrant 3 - Low Value Added, High Attainment
Quadrant 4 - Low Value Added, Low Attainment

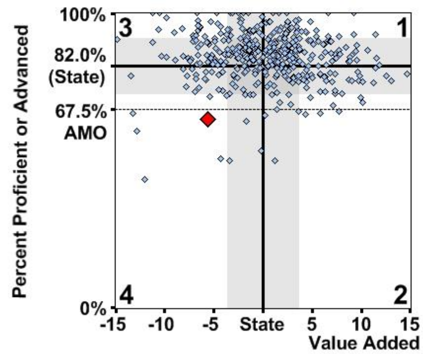
Milwaukee / State Averages
Value-Added and Attainment Quadrants

Mathematics 2006-07 Reading

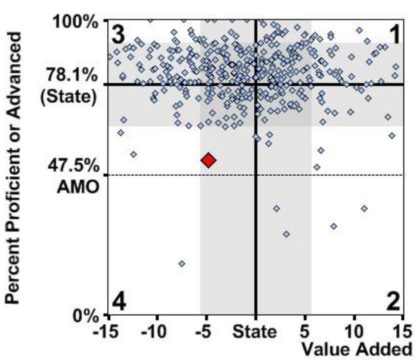
Grade 3



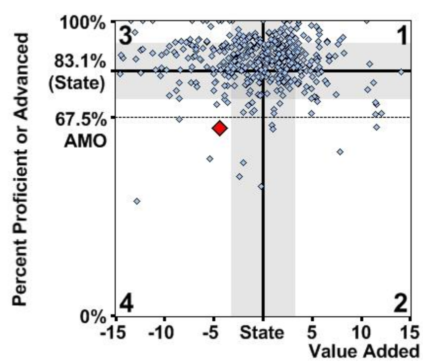
Grade 3



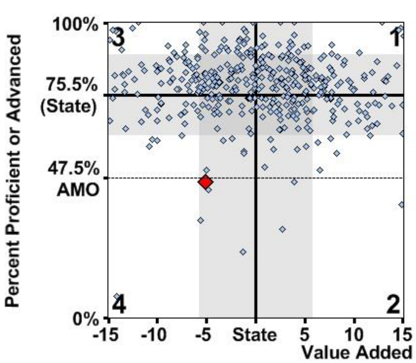
Grade 4



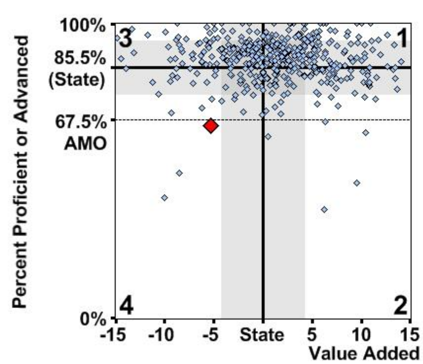
Grade 4



Grade 5



Grade 5

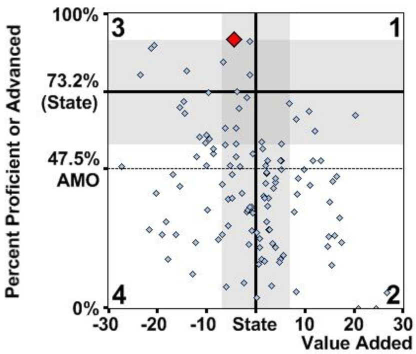


Quadrant 1 - High Value Added, High Attainment
Quadrant 2 - High Value Added, Low Attainment
Quadrant 3 - Low Value Added, High Attainment
Quadrant 4 - Low Value Added, Low Attainment

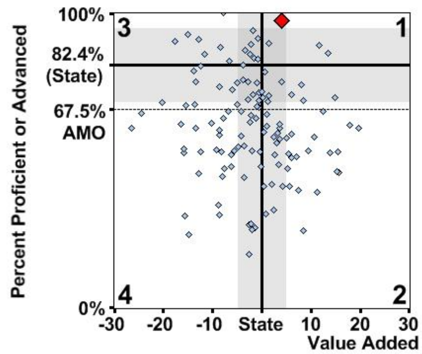
Meir / State Averages
Value-Added and Attainment Quadrants

Mathematics 2005-06 Reading

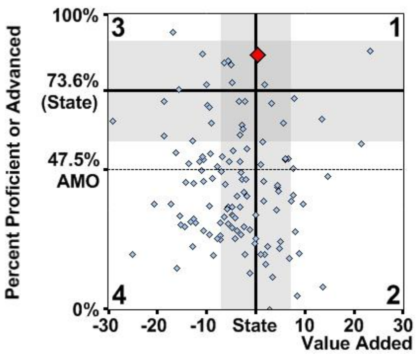
Grade 3



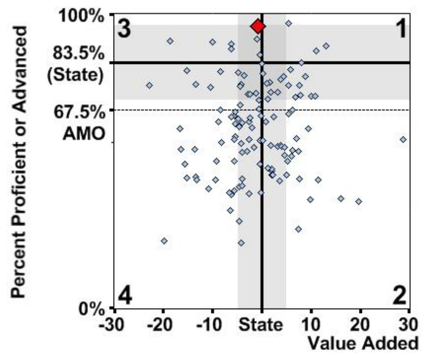
Grade 3



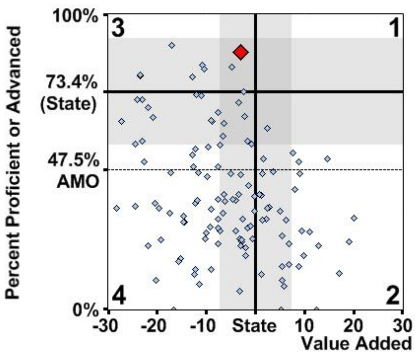
Grade 4



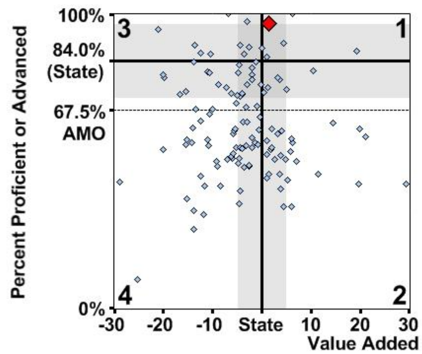
Grade 4



Grade 5



Grade 5

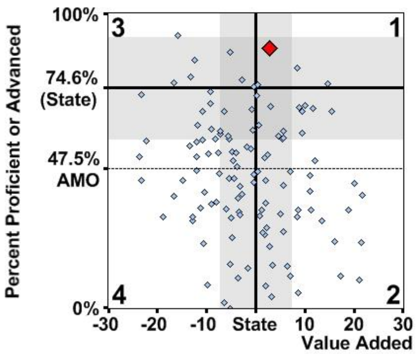


Quadrant 1 - High Value Added, High Attainment
Quadrant 2 - High Value Added, Low Attainment
Quadrant 3 - Low Value Added, High Attainment
Quadrant 4 - Low Value Added, Low Attainment

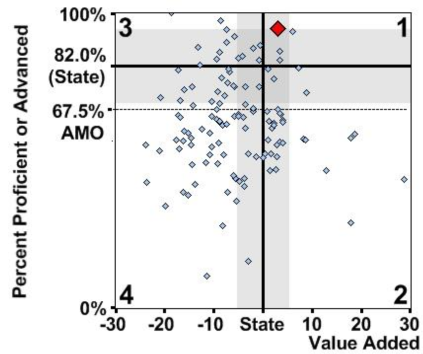
Meir / State Averages
Value-Added and Attainment Quadrants

Mathematics 2006-07 Reading

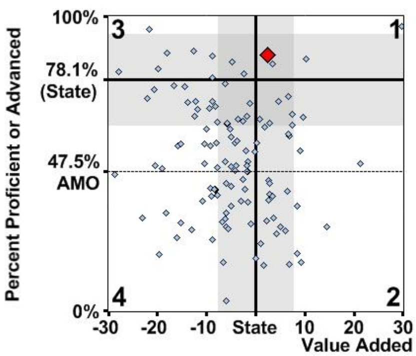
Grade 3



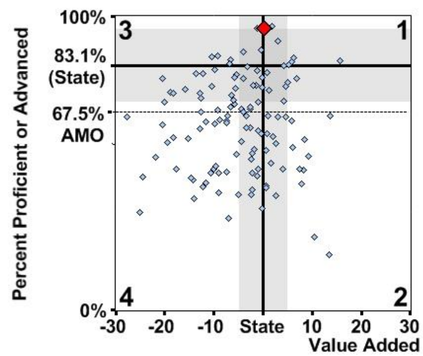
Grade 3



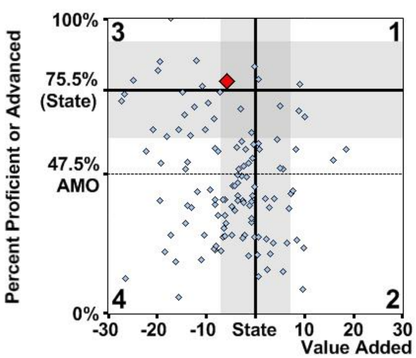
Grade 4



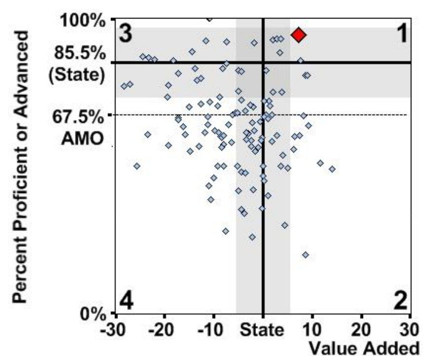
Grade 4



Grade 5



Grade 5



Quadrant 1 - High Value Added, High Attainment
Quadrant 2 - High Value Added, Low Attainment
Quadrant 3 - Low Value Added, High Attainment
Quadrant 4 - Low Value Added, Low Attainment

Appendix C.2 Madison: School and District Reports

School - Shorewood Hills							District - Madison					
Mathematics	**Value Added**				**Attainment**		**Value Added**				**Attainment**	
School Year 2006-07	**Value Added**	**Confidence Interval**	**N**	**State Percentile**	**WCKE % Proficient**	**Quadrant**	**Value Added**	**Confidence Interval**	**N**	**State Percentile**	**WCKE % Proficient**	**Quadrant**
Grade 3	23.82	15.33 to 32.31	49	98.9	87.8	1	0.80	-0.53 to 2.13	1629	52.6	70.8	2
Grade 4	3.51	-3.31 to 10.33	72	69.1	98.6	1	-0.99	-2.3 to 0.32	1648	51.1	73.8	4
Grade 5	14.07	7.15 to 20.99	58	96.1	96.6	1	3.93	2.73 to 5.13	1607	69.3	70.5	2
School Year 2005-06	**Value Added**	**Confidence Interval**	**N**	**State Percentile**	**WCKE % Proficient**	**Quadrant**	**Value Added**	**Confidence Interval**	**N**	**State Percentile**	**WCKE % Proficient**	**Quadrant**
Grade 3	-0.74	-9.25 to 7.77	49	95.0	97.3	3	-3.67	-4.9 to -2.44	1572	26.6	73.3	3
Grade 4	12.78	6.16 to 19.4	64	93.7	89.0	1	2.84	1.64 to 4.04	1537	69	72.9	2
Grade 5	6.11	0.23 to 11.99	72	79.9	97.2	1	-0.98	-2.2 to 0.24	1513	43.5	74.9	3
Reading	**Value Added**				**Attainment**		**Value Added**				**Attainment**	
School Year 2006-07	**Value Added**	**Confidence Interval**	**N**	**State Percentile**	**WCKE % Proficient**	**Quadrant**	**Value Added**	**Confidence Interval**	**N**	**State Percentile**	**WCKE % Proficient**	**Quadrant**
Grade 3	10.48	1.97 to 18.99	49	95.0	95.9	1	-0.66	-1.99 to 0.67	1629	42.7	75.6	4
Grade 4	13.60	7.03 to 20.17	72	98.7	98.6	1	2.73	1.48 to 3.98	1648	77.3	77.0	2
Grade 5	-2.21	-9.76 to 5.34	58	35.2	94.8	3	0.67	-0.64 to 1.98	1607	43.6	79.0	2
School Year 2005-06	**Value Added**	**Confidence Interval**	**N**	**State Percentile**	**WCKE % Proficient**	**Quadrant**	**Value Added**	**Confidence Interval**	**N**	**State Percentile**	**WCKE % Proficient**	**Quadrant**
Grade 3	4.10	-2.41 to 10.61	75	73.6	97.3	1	0.48	-0.83 to 1.79	1572	52.1	78.5	2
Grade 4	12.25	5.14 to 19.36	64	97.6	95.3	1	3.65	2.36 to 4.94	1537	80	81.2	2
Grade 5	8.65	2.08 to 15.22	72	92.1	98.6	1	0.88	-0.47 to 2.23	1513	56.2	83.5	2

Note: Value Added is derived from the WCKE tests held each November. School Year 2005-06 Value Added is derived from the November 2005 and November 2006 WKCE tests; School Year 2006-07 Value Added is derived from the November 2006 and November 2007 WCKE tests.

3rd Grade value-added measures growth from the November 3rd Grade test point through the November 4th Grade test point.
4th Grade value-added measures growth from the November 4th Grade test point through the November 5th Grade test point.
5th Grade value-added measures growth from the November 5th Grade test point through the November 6th Grade test point.

Report Key

Value Added

- This value is a measure that is equal to the number of extra points scored by students in a school or district (that is attributable to the school or district) on a test relative to observationally similar students across the state or district. For example, if a school's value added is +3, then students at the school scored three points higher on the test then similar students across the state. If a school's value added is -3, then students at the school scored three points lower relative to similar students. An average school or district will have a relative value-added of zero when compared to the state.

Confidence Interval

- This reflects the range of possible value-added scores. As the number of students measured (n) increases, the confidence interval generally decreases – note the smaller District-level confidence intervals versus the School-level confidence intervals.

N

- Number of students tested at a school.

State Percentile

- The value-added state percentile rank for a school or district.

Attainment

- This is the percentage of students scoring proficient or advanced on the WKCE for a school or district.

Quadrant

- This is a value from 1 to 4 based upon a combination of value-added and attainment data for the school or district.

Graphical Presentation

A graphical presentation of each school and district's value-added data follows.

District Compared to the State 2005-06
District Compared to the State 2006-07
School Compared to the State 2005-06
School Compared to the State 2006-07

For both the reference and analysis sections, the following comparisons are made in terms of scale scores:

District Compared to State

- This graphic shows the district's value added in scale score points compared to the state value-added average on the horizontal axis. The district's state percentile rank is shown by the data point on the vertical axis.

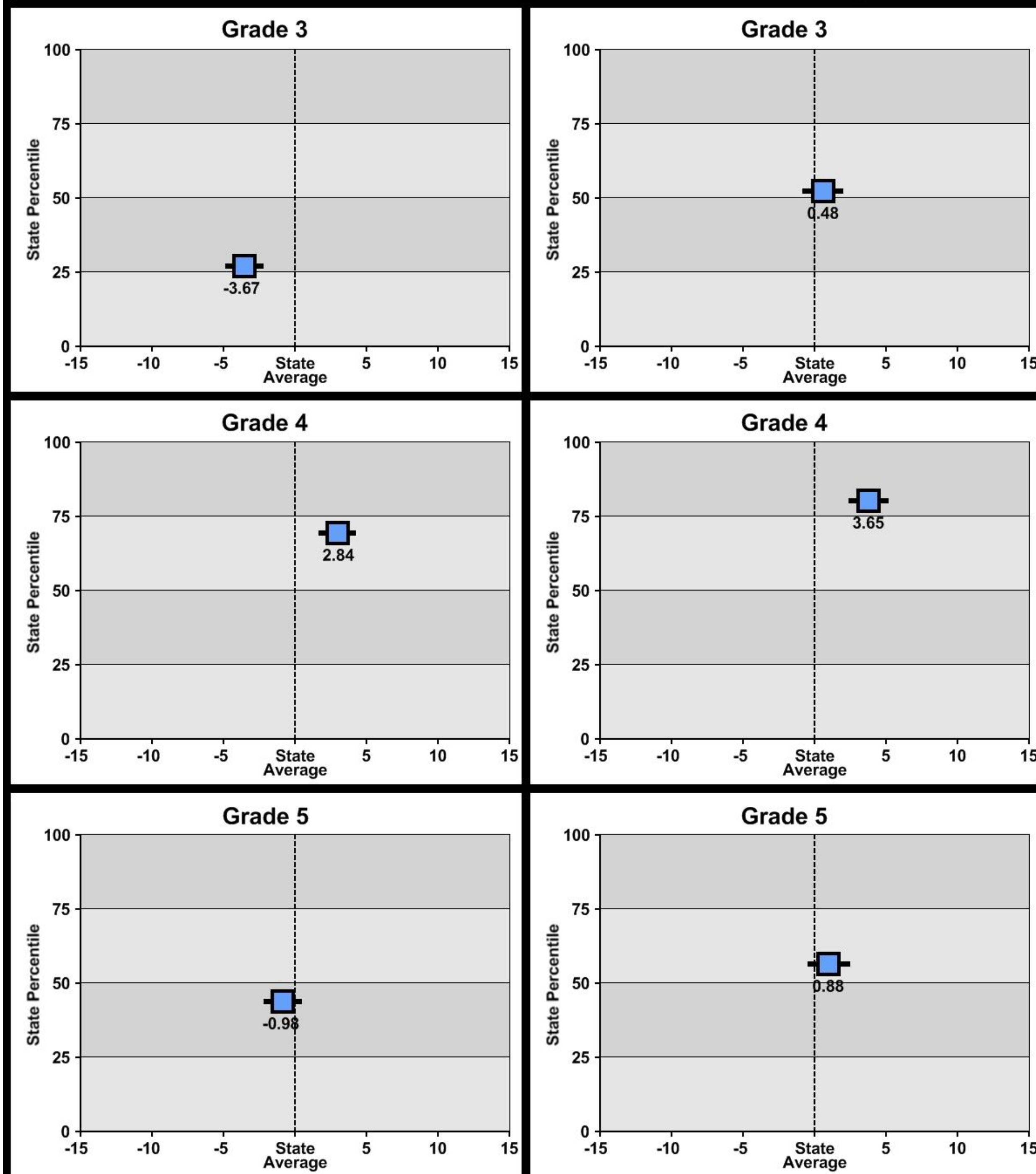
School Compared to State

- This graphic shows the school's value added in scale score points compared to the state value-added average on the horizontal axis. The school's state percentile rank is shown by the data point on the vertical axis.

Each comparison includes a column for mathematics and a column for reading. Each column has three rows providing value-added data for each grade – third through fifth.

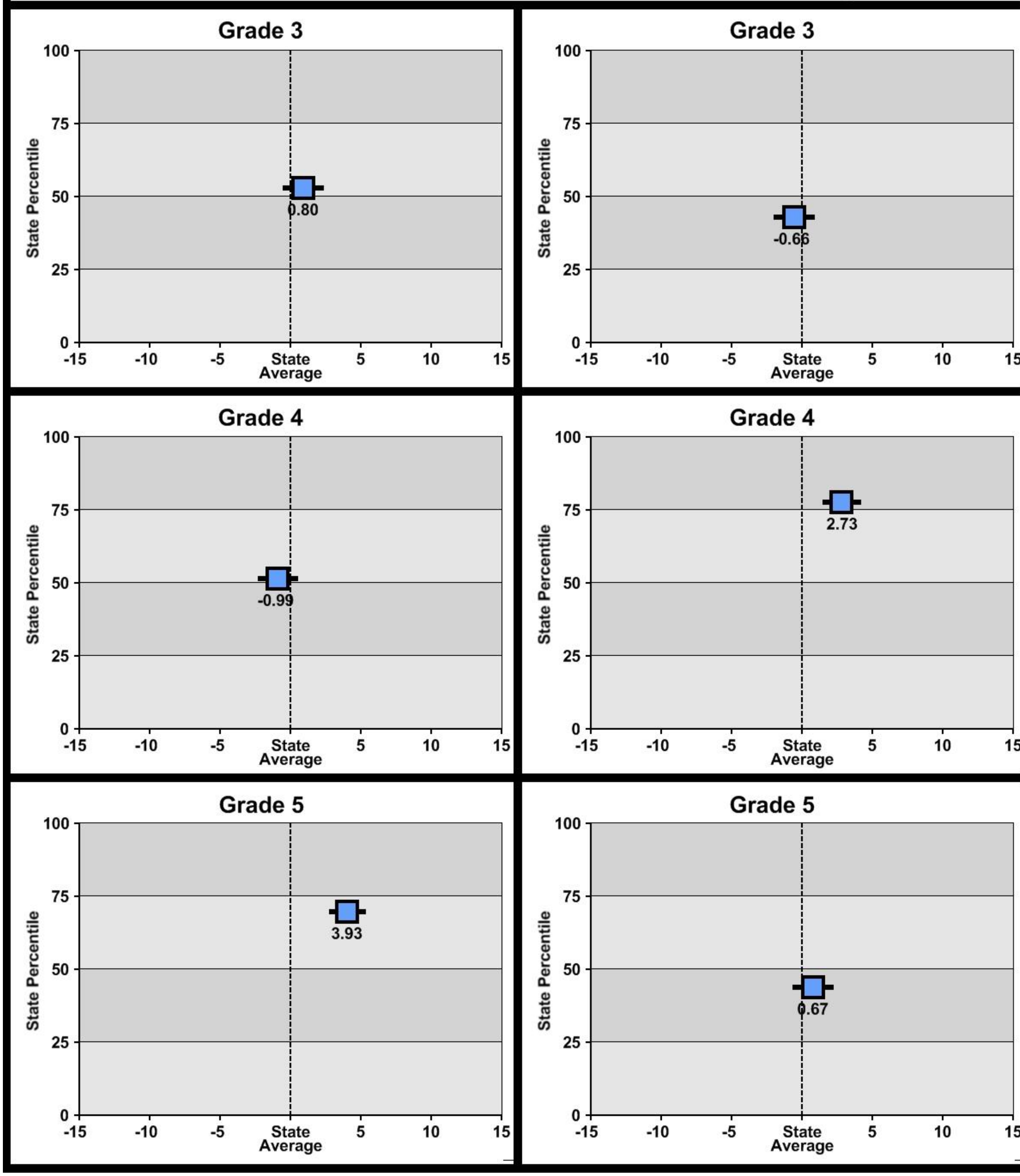
Madison Compared to State
Value-Added

Mathematics **2005-06** **Reading**



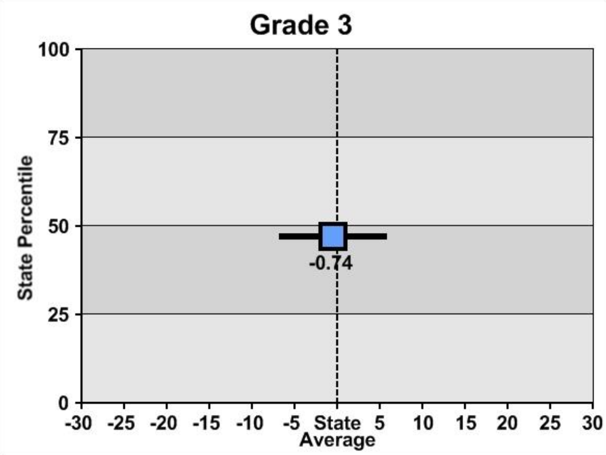
Madison Compared to State
Value-Added

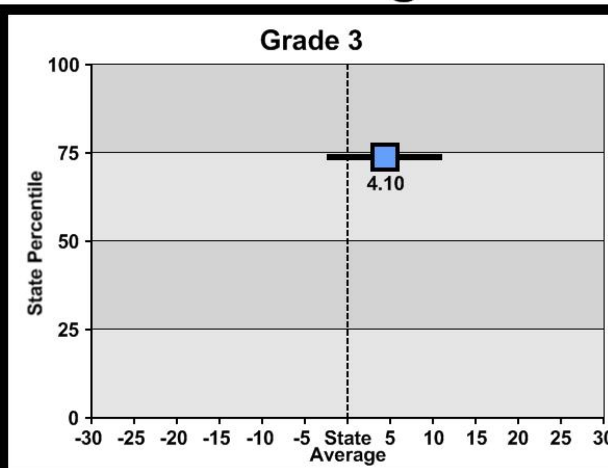
Mathematics 2006-07 Reading

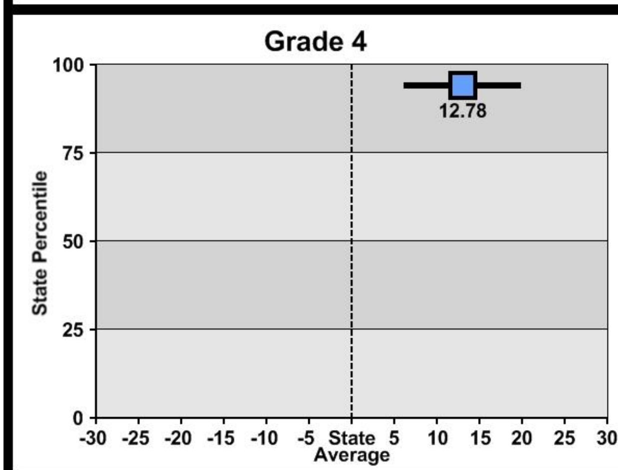


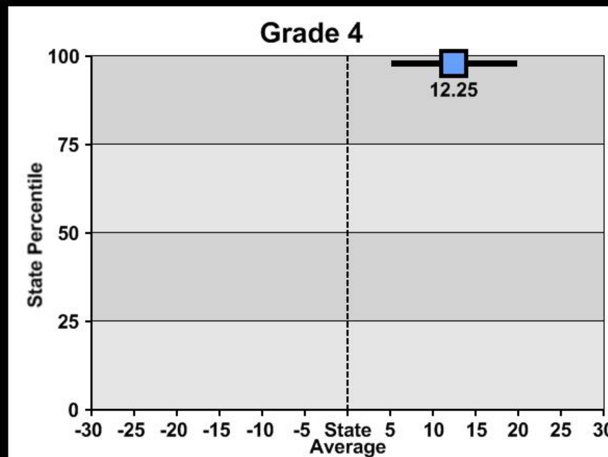
Shorewood Hills Compared to State Value-Added

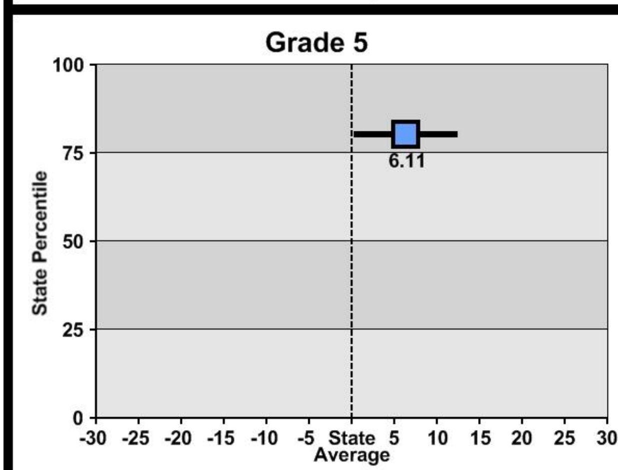
Mathematics 2005-06 Reading

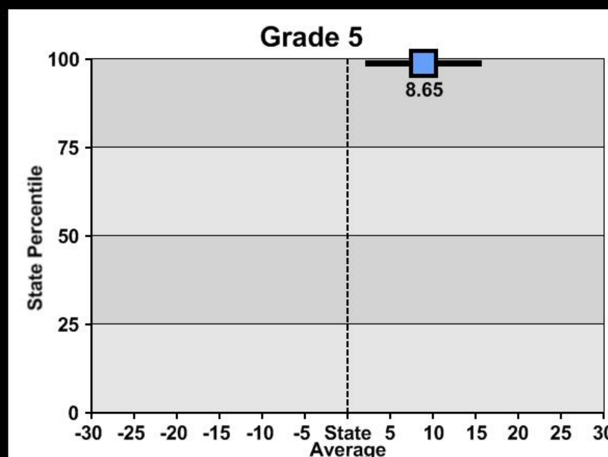






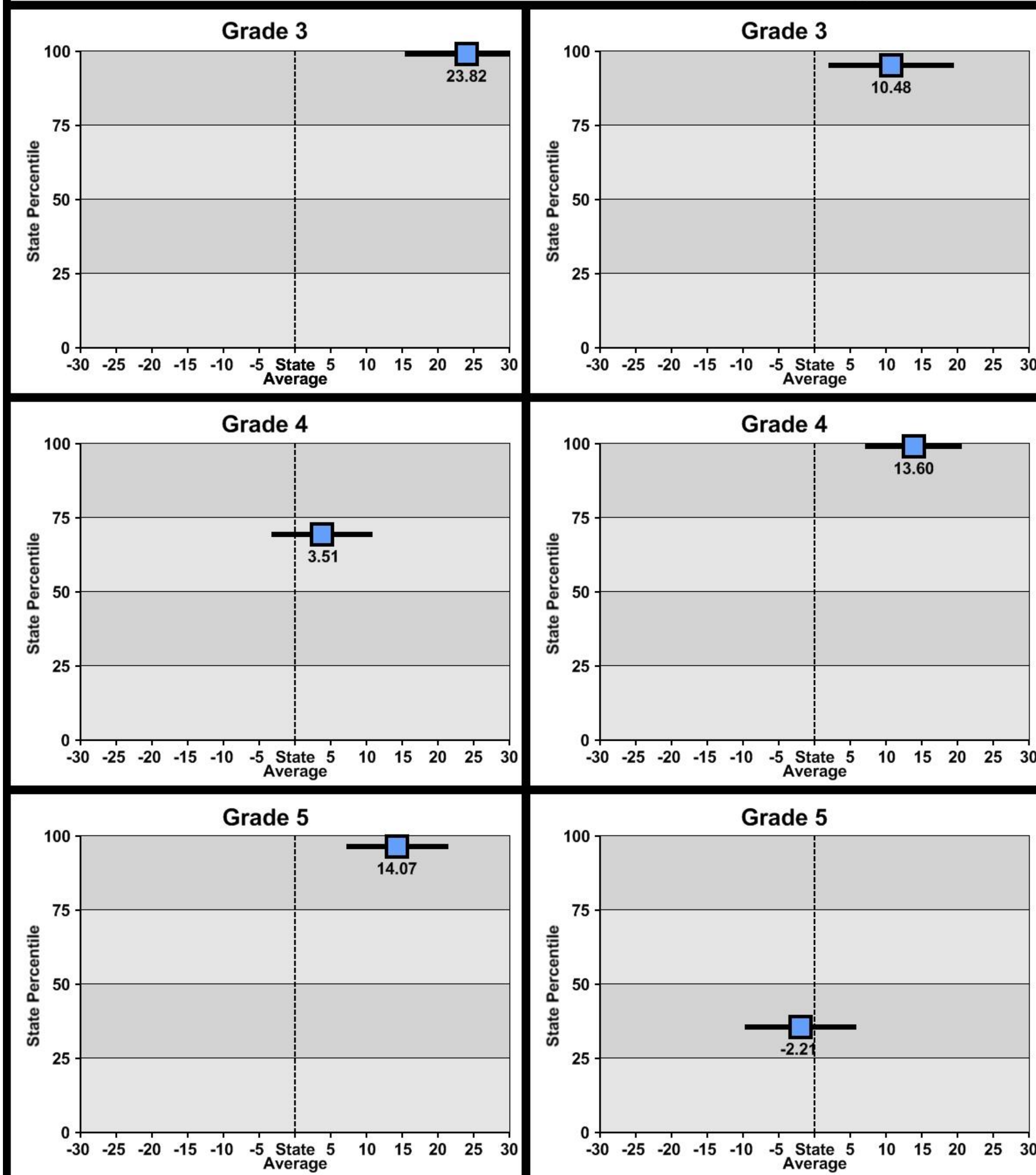






Shorewood Hills Compared to State Value-Added

Mathematics 2006-07 Reading



Analysis Section

The information below will guide your interpretation of the following plots:

- District / State Averages, Value-Added and Attainment Sectors
- School / State Averages, Value-Added and Attainment Sectors

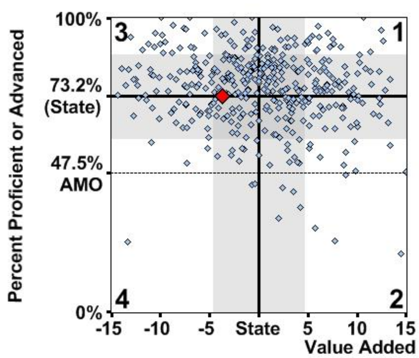
- Graphic Layout
  - Each data point represents a school or district and is determined by plotting a school or district's value-added score against the school or district's percent proficient/advanced on the WKCE (attainment). Your district (District / State Averages) or school (School / State Averages) is represented by a red diamond .
  - The state average for both value added and percent proficient/advanced provides the structure upon which the four quadrants are distinguished.
  - Schools and districts will fall into one of the four different quadrants.
  - The gray shaded areas above and below, and to the left and the right of the state average lines represent one standard deviation away from that line.
  - The AMO (Annual Measurable Objectives), a measurement used to determine compliance with the federal No Child Left Behind Act (NCLB), is included.

- Quadrant Analysis
  - Schools, or (districts) in Quadrant 1 (high value added, high attainment) are both above average in growth and above average in attainment.
  - Schools, or (districts) in Quadrant 2 (high value added, low attainment) are above average in growth and below average in attainment.
  - Schools, or (districts) in Quadrant 3 (low value added, high attainment) are below average in growth and above average in attainment.
  - Schools, or (districts) in Quadrant 4 (low value added, low attainment) are both below average in growth and below average in attainment.

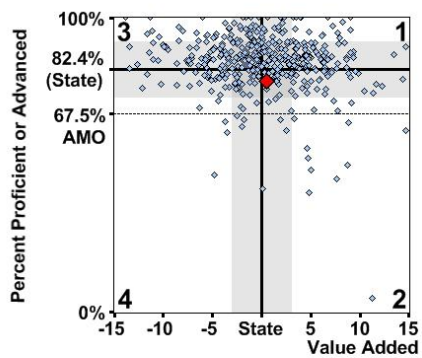
Madison / State Averages
Value-Added and Attainment Quadrants

Mathematics 2005-06 Reading

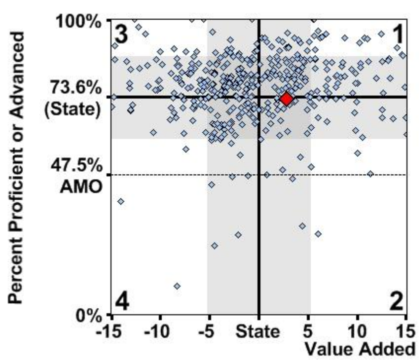
Grade 3



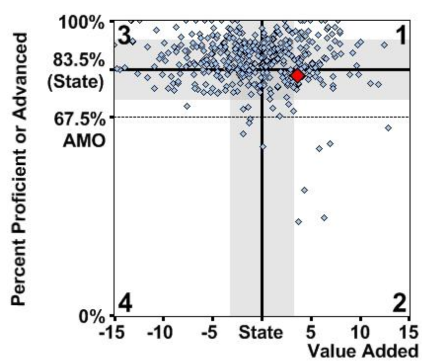
Grade 3



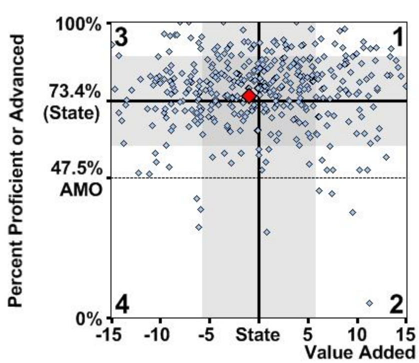
Grade 4



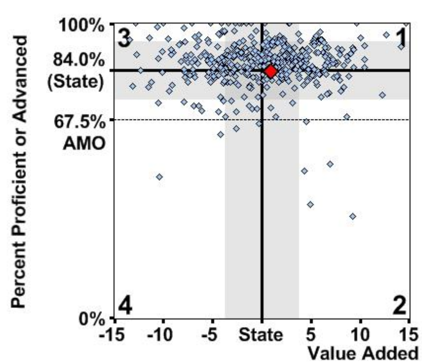
Grade 4



Grade 5



Grade 5

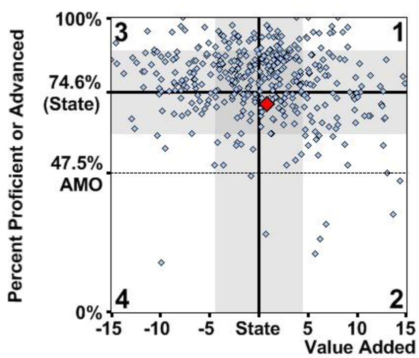


Quadrant 1 - High Value Added, High Attainment
Quadrant 2 - High Value Added, Low Attainment
Quadrant 3 - Low Value Added, High Attainment
Quadrant 4 - Low Value Added, Low Attainment

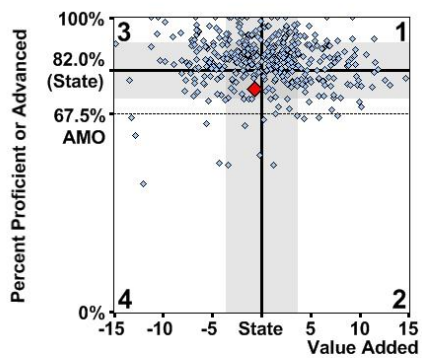
Madison / State Averages
Value-Added and Attainment Quadrants

Mathematics 2006-07 Reading

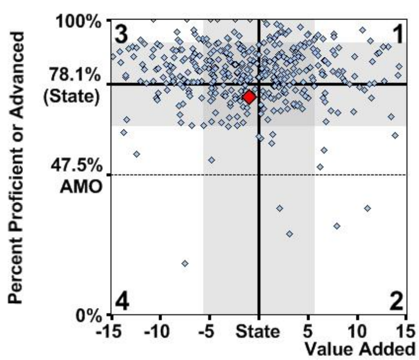
Grade 3



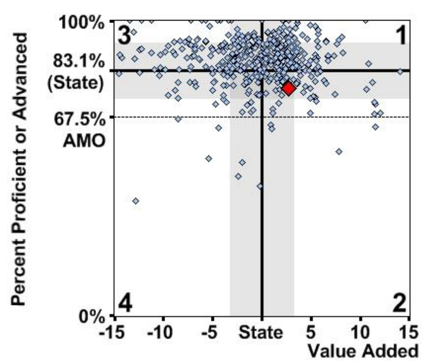
Grade 3



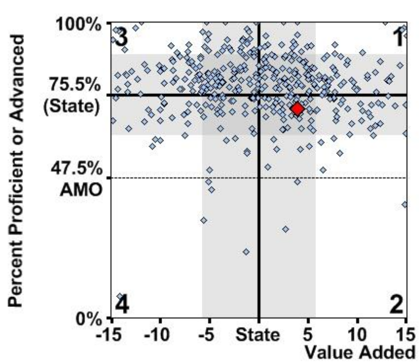
Grade 4



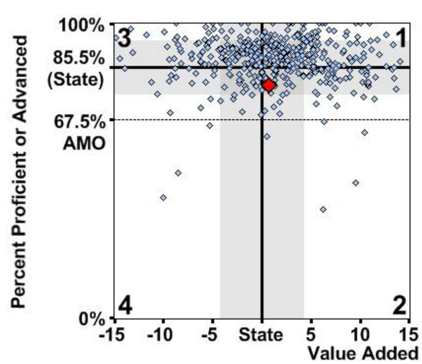
Grade 4



Grade 5



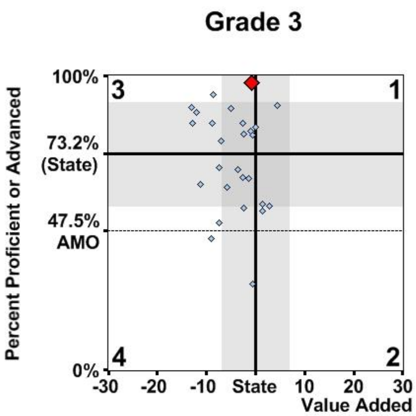
Grade 5

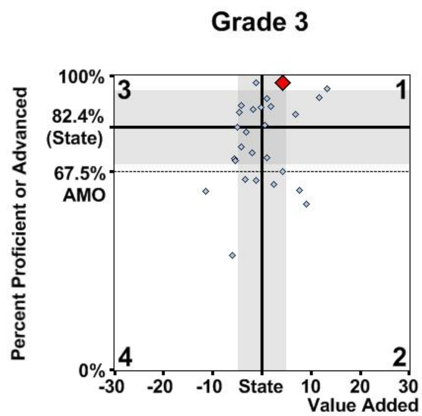


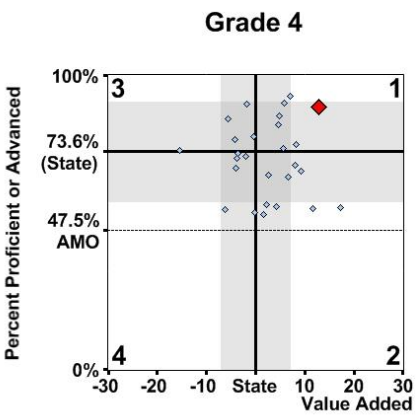
Quadrant 1 - High Value Added, High Attainment
Quadrant 2 - High Value Added, Low Attainment
Quadrant 3 - Low Value Added, High Attainment
Quadrant 4 - Low Value Added, Low Attainment

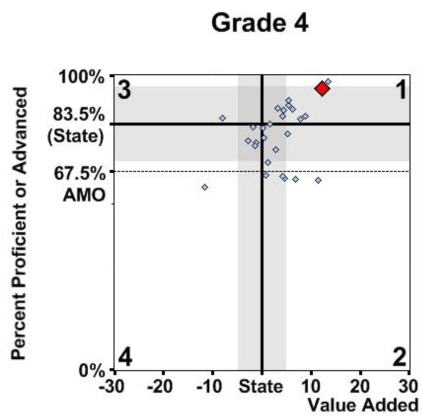
Shorewood Hills / State Averages
Value-Added and Attainment Quadrants

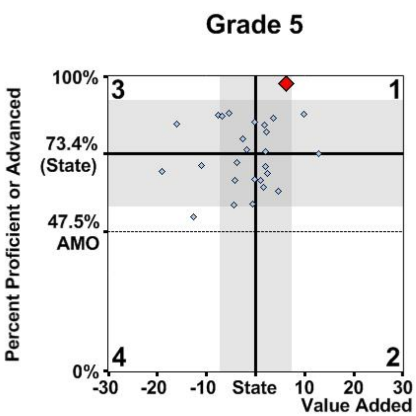
Mathematics 2005-06 Reading

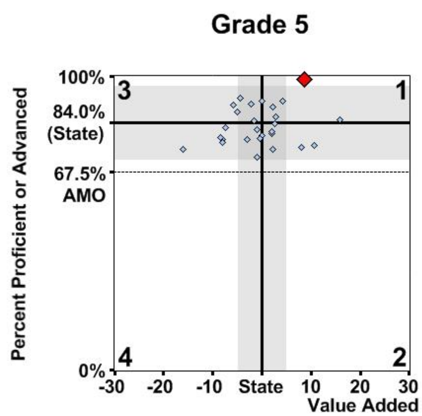










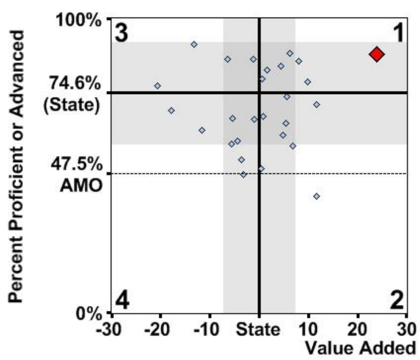


Quadrant 1 - High Value Added, High Attainment
Quadrant 2 - High Value Added, Low Attainment
Quadrant 3 - Low Value Added, High Attainment
Quadrant 4 - Low Value Added, Low Attainment

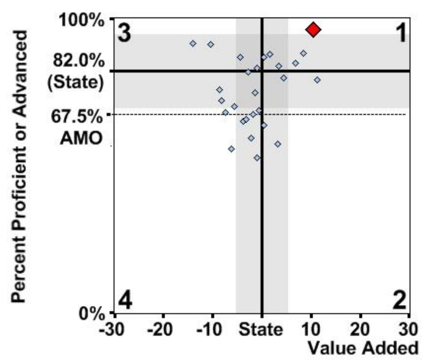
Shorewood Hills / State Averages
Value-Added and Attainment Quadrants

Mathematics 2006-07 Reading

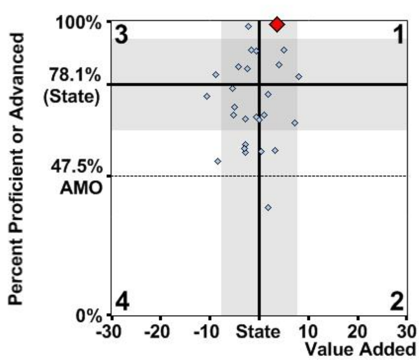
Grade 3



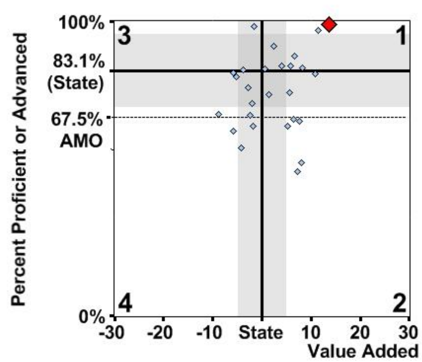
Grade 3



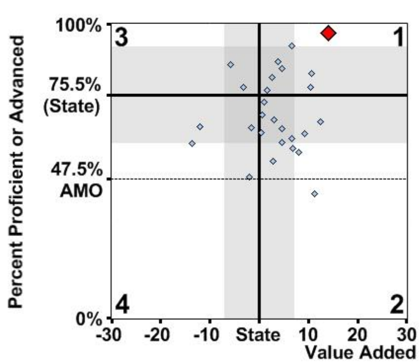
Grade 4



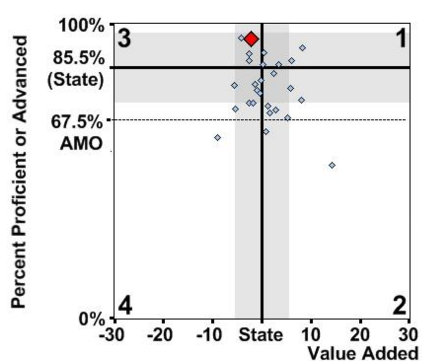
Grade 4



Grade 5



Grade 5



Quadrant 1 - High Value Added, High Attainment
Quadrant 2 - High Value Added, Low Attainment
Quadrant 3 - Low Value Added, High Attainment
Quadrant 4 - Low Value Added, Low Attainment

Appendix C.3 Waunakee: School and District Reports

	School – Prairie Elementary						District - Waunakee					
Mathematics	**Value Added**				**Attainment**		**Value Added**				**Attainment**	
School Year 2006-07	**Value Added**	**Confidence Interval**	**N**	**State Percentile**	**WCKE % Proficient**	**Quadrant**	**Value Added**	**Confidence Interval**	**N**	**State Percentile**	**WCKE % Proficient**	**Quadrant**
Grade 3	-2.41	-8.05 to 3.21	93	37.2	96.8	3	1.07	-2.57 to 4.71	217	54.7	95.9	1
Grade 4	16.60	10.37 to 22.83	74	96.7	93.2	1	17.80	14.16 to 21.44	214	98.2	90.4	1
Grade 5	N/A	N/A	N/A	N/A	N/A	N/A	5.32	2.35 to 8.3	260	76.9	92.0	1
School Year 2005-06	**Value Added**	**Confidence Interval**	**N**	**State Percentile**	**WCKE % Proficient**	**Quadrant**	**Value Added**	**Confidence Interval**	**N**	**State Percentile**	**WCKE % Proficient**	**Quadrant**
Grade 3	2.13	-2.49 to 6.76	113	59.8	91.2	1	-1.17	-4.72 to 2.39	188	40.6	92.2	3
Grade 4	9.54	5.5 to 13.58	139	88.6	91.4	1	10.17	7.12 to 13.21	241	92.2	91.9	1
Grade 5	N/A	N/A	N/A	N/A	N/A	N/A	6.29	3.17 to 9.42	229	80.9	94.4	1
Reading	**Value Added**				**Attainment**		**Value Added**				**Attainment**	
School Year 2006-07	**Value Added**	**Confidence Interval**	**N**	**State Percentile**	**WCKE % Proficient**	**Quadrant**	**Value Added**	**Confidence Interval**	**N**	**State Percentile**	**WCKE % Proficient**	**Quadrant**
Grade 3	-3.44	-9.07 to 2.2	93	28.8	95.7	3	-4.27	-7.91 to -0.62	217	20.3	95.5	3
Grade 4	2.59	-3.41 to 8.59	74	67.3	95.9	1	4.56	1.06 to 8.07	214	87.5	93.6	1
Grade 5	N/A	N/A	N/A	N/A	N/A	N/A	5.56	2.32 to 8.8	260	77.4	93.5	1
School Year 2005-06	**Value Added**	**Confidence Interval**	**N**	**State Percentile**	**WCKE % Proficient**	**Quadrant**	**Value Added**	**Confidence Interval**	**N**	**State Percentile**	**WCKE % Proficient**	**Quadrant**
Grade 3	1.55	-3.39 to 6.48	113	59.4	97.3	1	-2.07	-5.86 to 1.73	188	31.1	95.6	3
Grade 4	1.26	-3.08 to 5.59	139	58.4	97.8	1	0.20	-3.06 to 3.47	241	57.0	95.5	1
Grade 5	N/A	N/A	N/A	N/A	N/A	N/A	8.96	5.47 to 12.44	229	95.3	89.7	1

Note: Value Added is derived from the WCKE tests held each November. School Year 2005-06 Value Added is derived from the November 2005 and November 2006 WKCE tests; School Year 2006-07 Value Added is derived from the November 2006 and November 2007 WCKE tests.

3rd Grade value-added measures growth from the November 3rd Grade test point through the November 4th Grade test point.
4th Grade value-added measures growth from the November 4th Grade test point through the November 5th Grade test point.
5th Grade value-added, not computed since this is a K-4 school.

Report Key

Value Added

- This value is a measure that is equal to the number of extra points scored by students in a school or district (that is attributable to the school or district) on a test relative to observationally similar students across the state or district. For example, if a school's value added is +3, then students at the school scored three points higher on the test then similar students across the state. If a school's value added is -3, then students at the school scored three points lower relative to similar students. An average school or district will have a relative value-added of zero when compared to the state.

Confidence Interval

- This reflects the range of possible value-added scores. As the number of students measured (n) increases, the confidence interval generally decreases – note the smaller District-level confidence intervals versus the School-level confidence intervals.

N

- Number of students tested at a school.

State Percentile

- The value-added state percentile rank for a school or district.

Attainment

- This is the percentage of students scoring proficient or advanced on the WKCE for a school or district.

Quadrant

- This is a value from 1 to 4 based upon a combination of value-added and attainment data for the school or district.

Graphical Presentation

A graphical presentation of each school and district's value-added data follows.

District Compared to the State 2005-06
District Compared to the State 2006-07
School Compared to the State 2005-06
School Compared to the State 2006-07

For both the reference and analysis sections, the following comparisons are made in terms of scale scores:

District Compared to State

- This graphic shows the district's value-added in scale score points compared to the state value-added average on the horizontal axis. The district's state percentile rank is shown by the data point on the vertical axis.

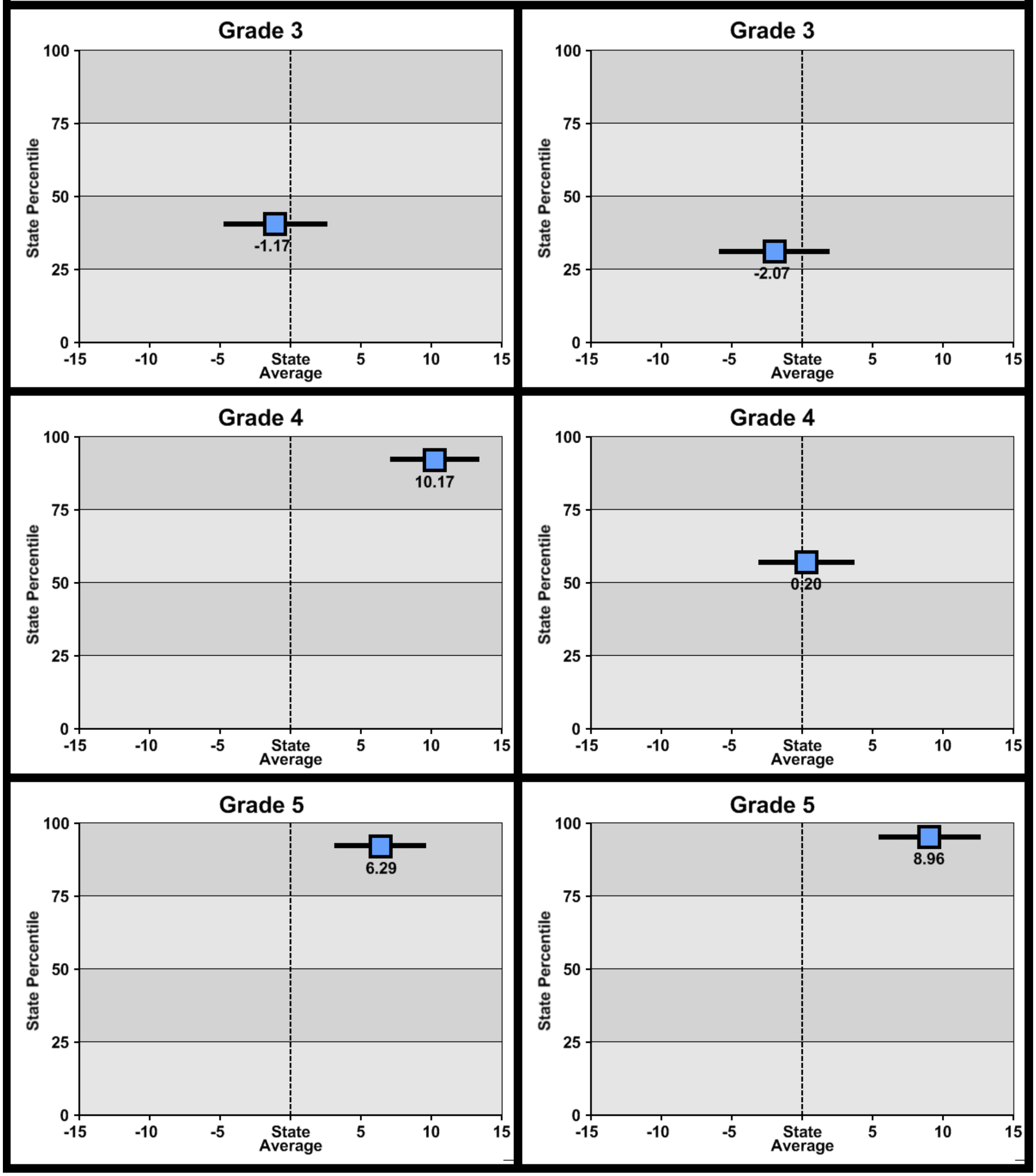
School Compared to State

- This graphic shows the school's value-added in scale score points compared to the state value-added average on the horizontal axis. The school's state percentile rank is shown by the data point on the vertical axis.

Each comparison includes a column for Mathematics and a column for Reading. Each column has three rows providing value-added data for each grade – third through fifth.

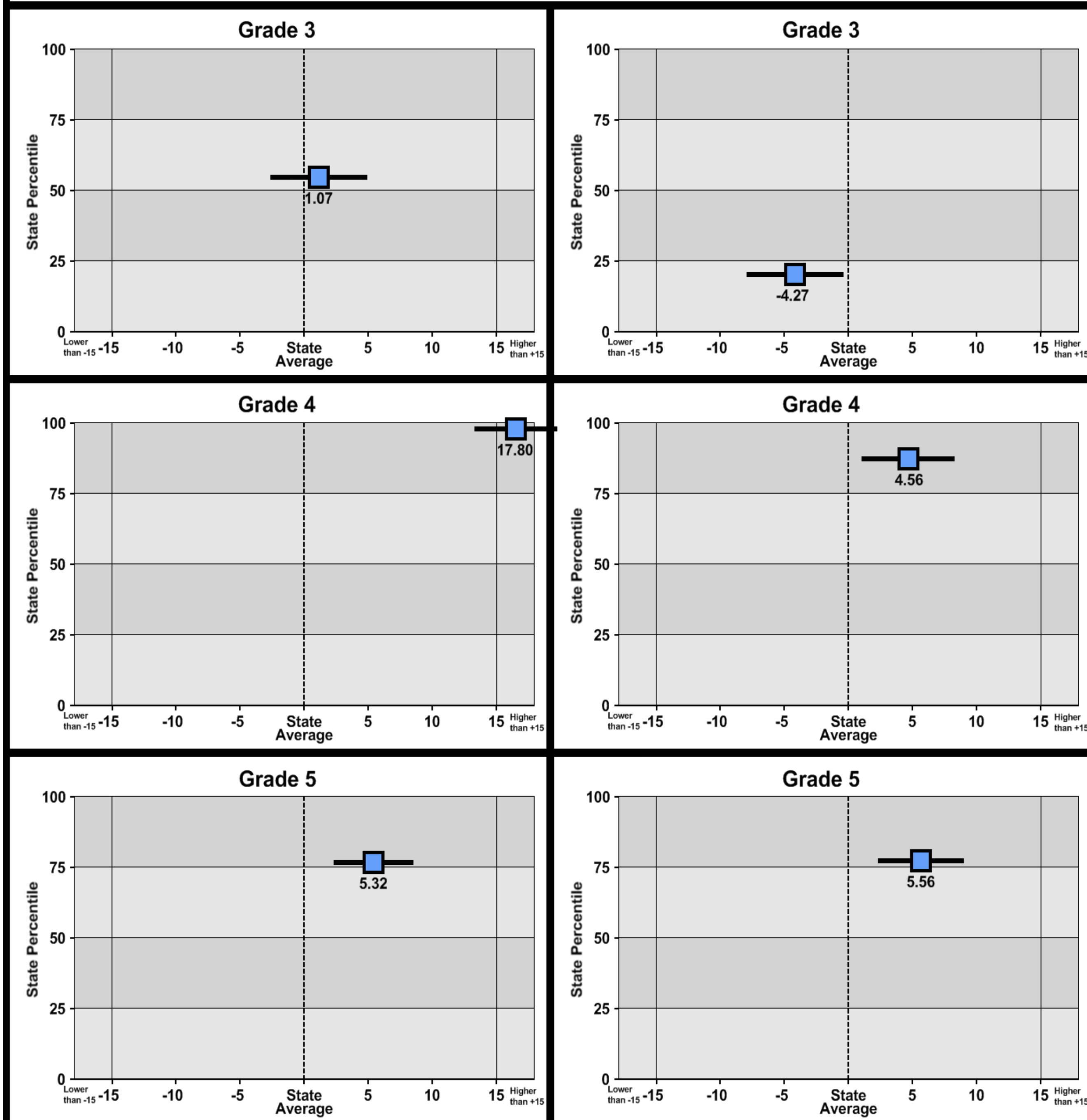
Waunakee Compared to State Value-Added

Mathematics **2005-06** **Reading**



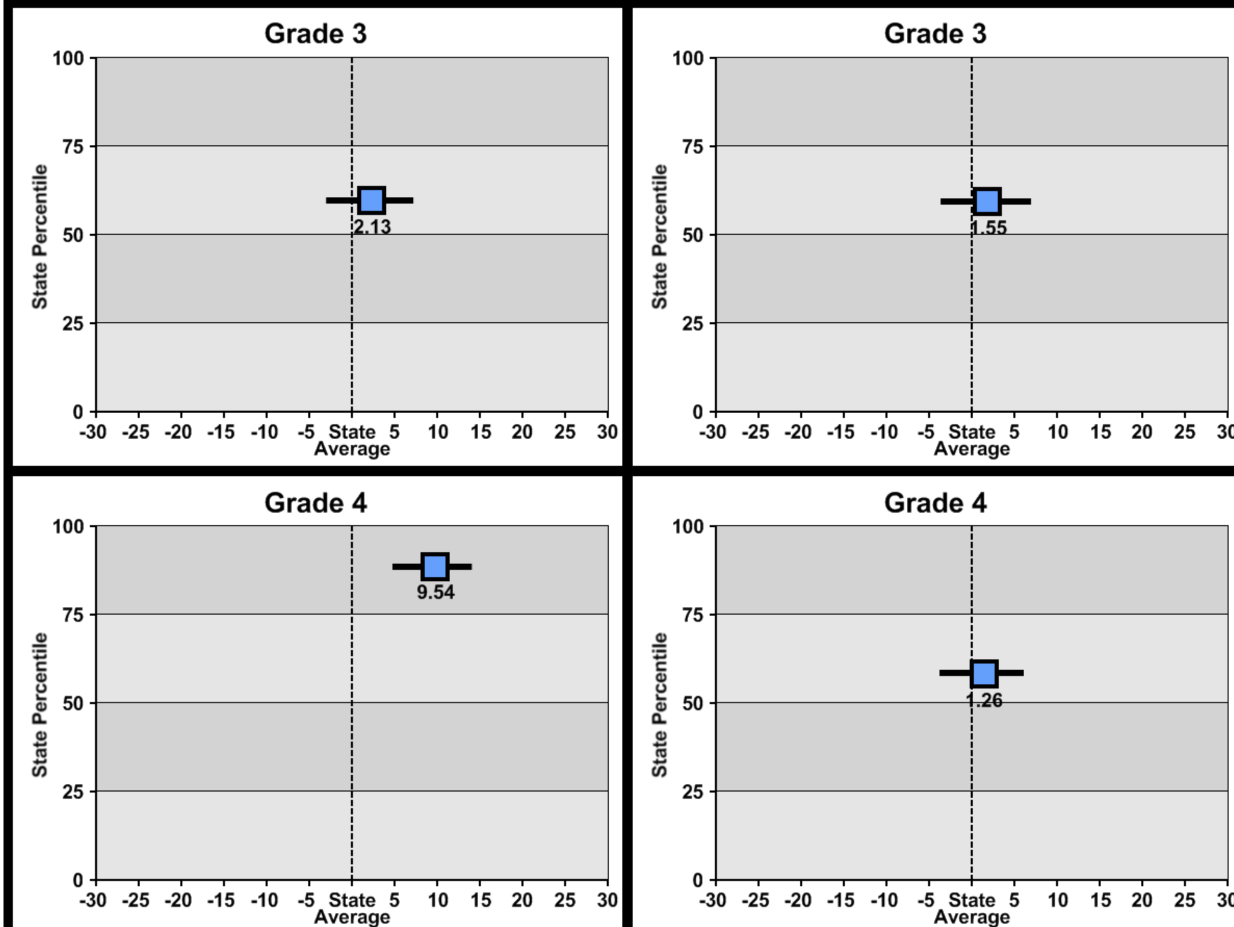
Waunakee Compared to State
Value-Added

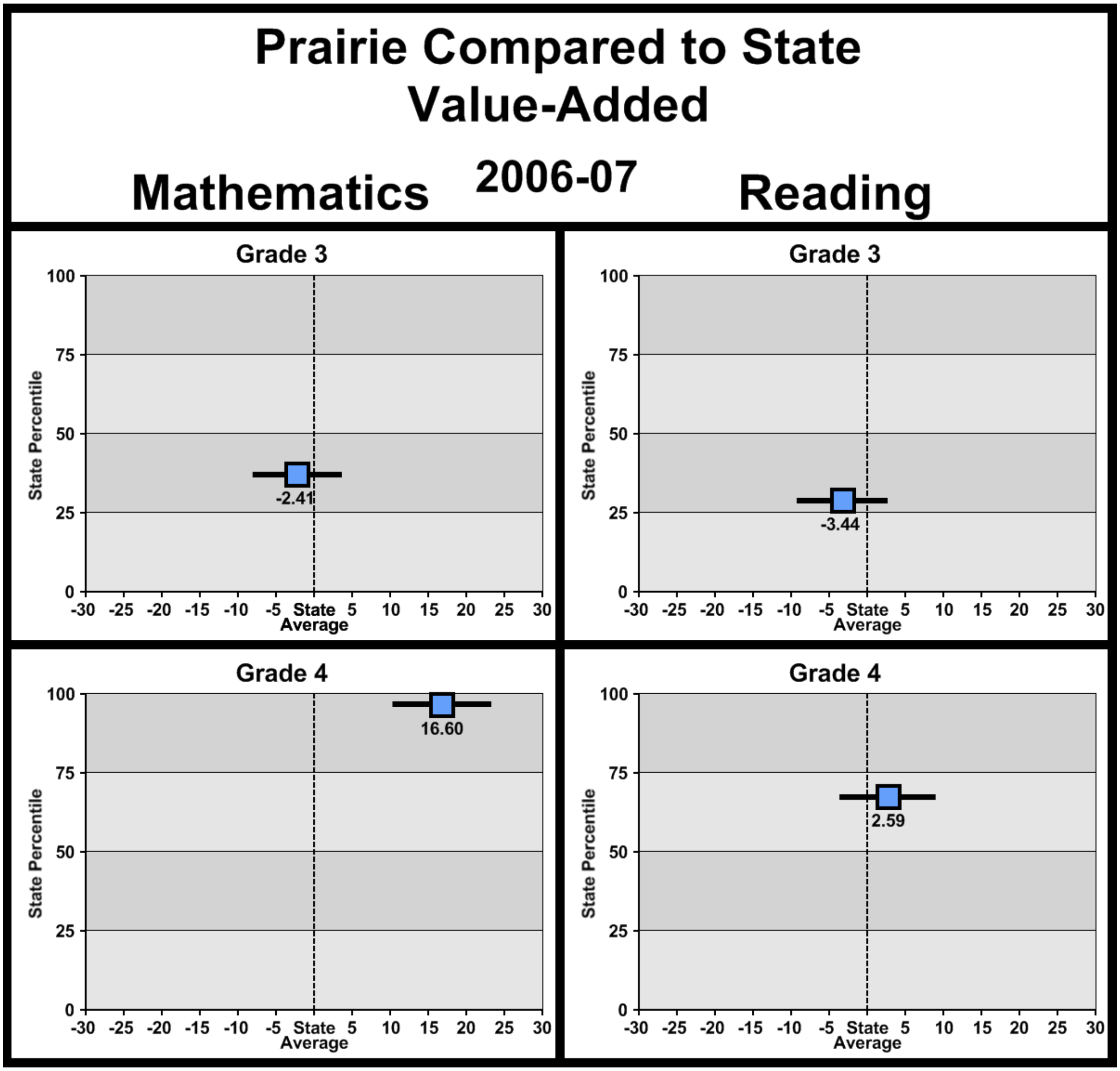
Mathematics 2006-07 Reading



Prairie Compared to State
Value-Added

Mathematics **2005-06** **Reading**




Prairie Compared to State
Value-Added
Mathematics
2006-07
Reading
Grade 3
-2.41
Grade 3
-3.44
Grade 4
16.60
Grade 4
2.59
State Percentile
100
75
50
25
0
-30 -25 -20 -15 -10 -5 State Average 5 10 15 20 25 30

Analysis Section

The information below will guide your interpretation of the following plots:

- District / State Averages, Value-Added and Attainment Sectors
- School / State Averages, Value-Added and Attainment Sectors

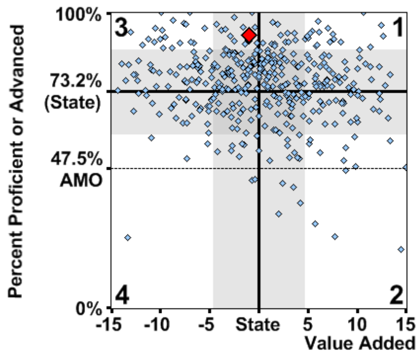
- Graphic Layout
  - Each data point represents a school or district and is determined by plotting a school or district's value-added score against the school or district's percent proficient/advanced on the WKCE (attainment). Your district (District / State Averages) or school (School / State Averages) is represented by a red diamond ◆.
  - The state average for both value added and percent proficient/advanced provides the structure upon which the four quadrants are distinguished.
  - Schools and districts will fall into one of the four different quadrants.
  - The gray shaded areas above and below, and to the left and the right of the state average lines represent one standard deviation away from that line.
  - The AMO (Annual Measurable Objectives), a measurement used to determine compliance with the federal No Child Left Behind Act (NCLB), is included.

- Quadrant Analysis
  - Schools, or (districts) in Quadrant 1 (high value added, high attainment) are both above average in growth and above average in attainment.
  - Schools, or (districts) in Quadrant 2 (high value added, low attainment) are above average in growth and below average in attainment.
  - Schools, or (districts) in Quadrant 3 (low value added, high attainment) are below average in growth and above average in attainment.
  - Schools, or (districts) in Quadrant 4 (low value added, low attainment) are both below average in growth and below average in attainment.

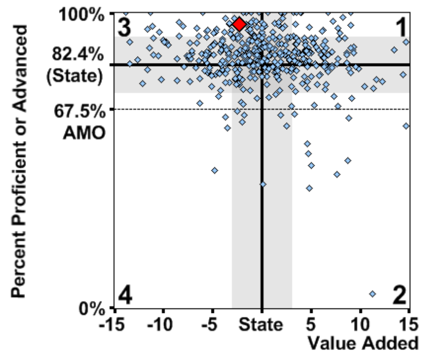
Waunakee / State Averages
Value-Added and Attainment Quadrants

Mathematics **2005-06** **Reading**

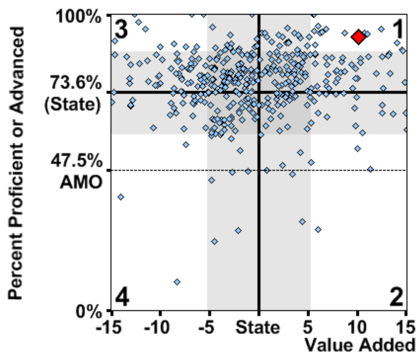
Grade 3



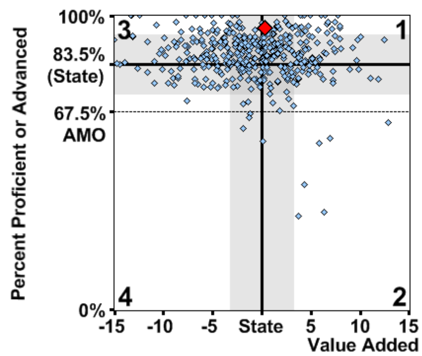
Grade 3



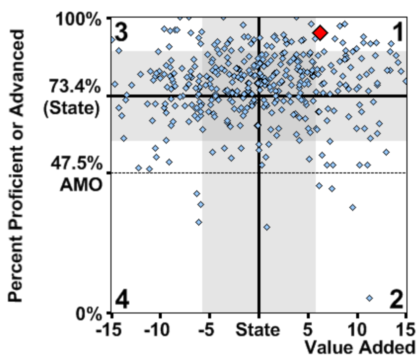
Grade 4



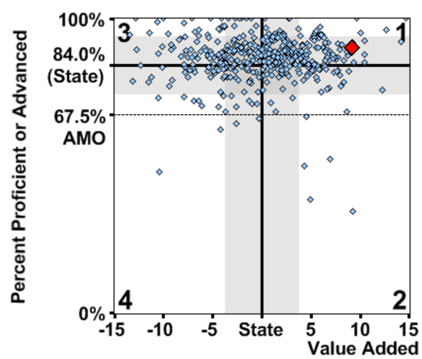
Grade 4



Grade 5



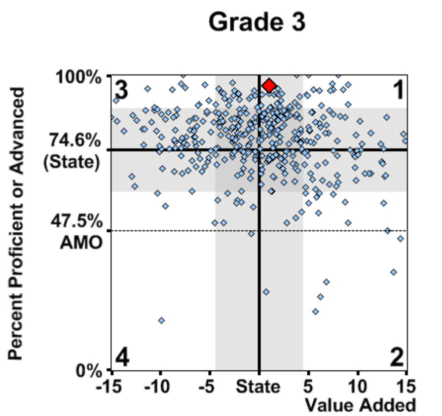
Grade 5

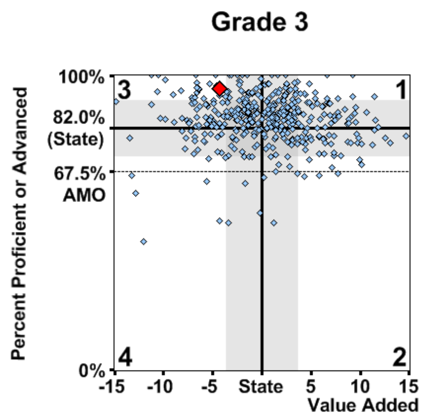


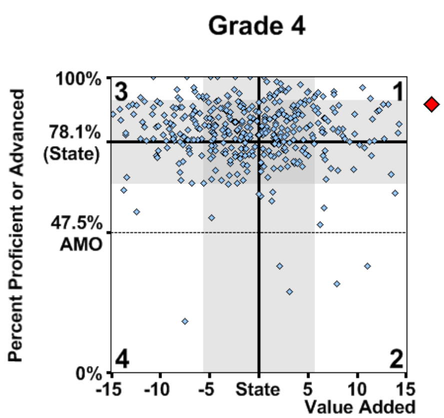
Quadrant 1 - High Value Added, High Attainment
Quadrant 2 - High Value Added, Low Attainment
Quadrant 3 - Low Value Added, High Attainment
Quadrant 4 - Low Value Added, Low Attainment

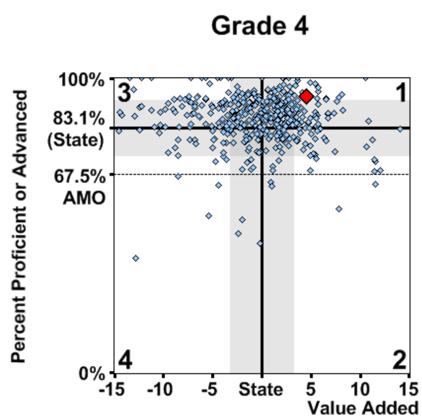
Waunakee / State Averages
Value-Added and Attainment Quadrants

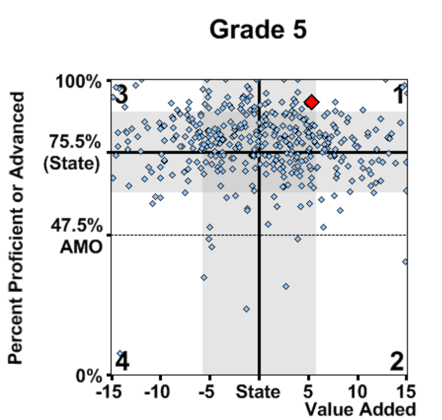
Mathematics **2006-07** **Reading**

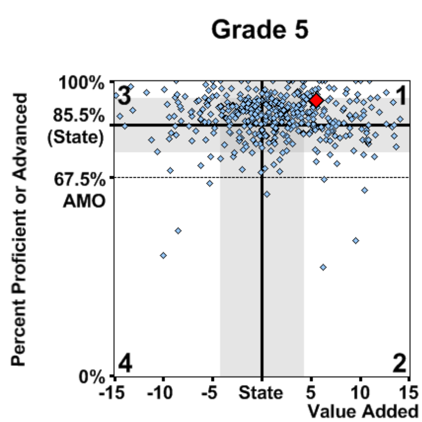










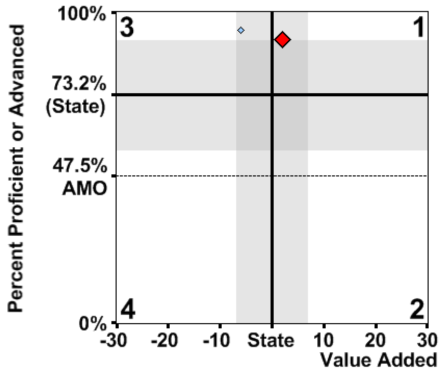


Quadrant 1 - High Value Added, High Attainment
Quadrant 2 - High Value Added, Low Attainment
Quadrant 3 - Low Value Added, High Attainment
Quadrant 4 - Low Value Added, Low Attainment

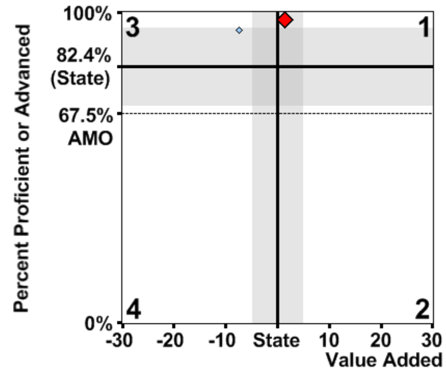
Prairie / State Averages
Value-Added and Attainment Quadrants

Mathematics 2005-06 Reading

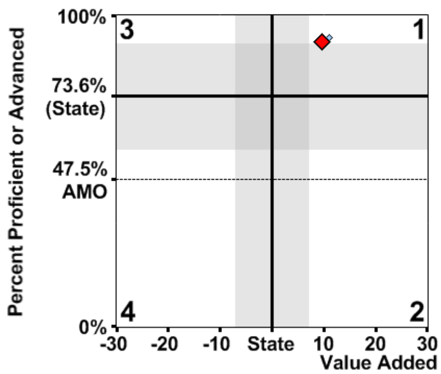
Grade 3



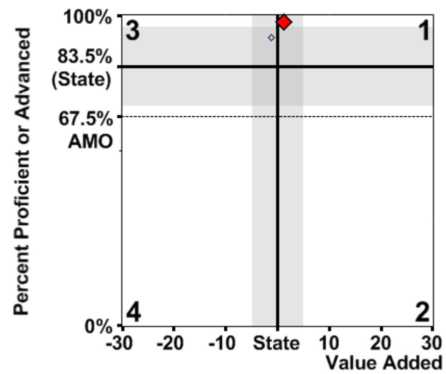
Grade 3



Grade 4



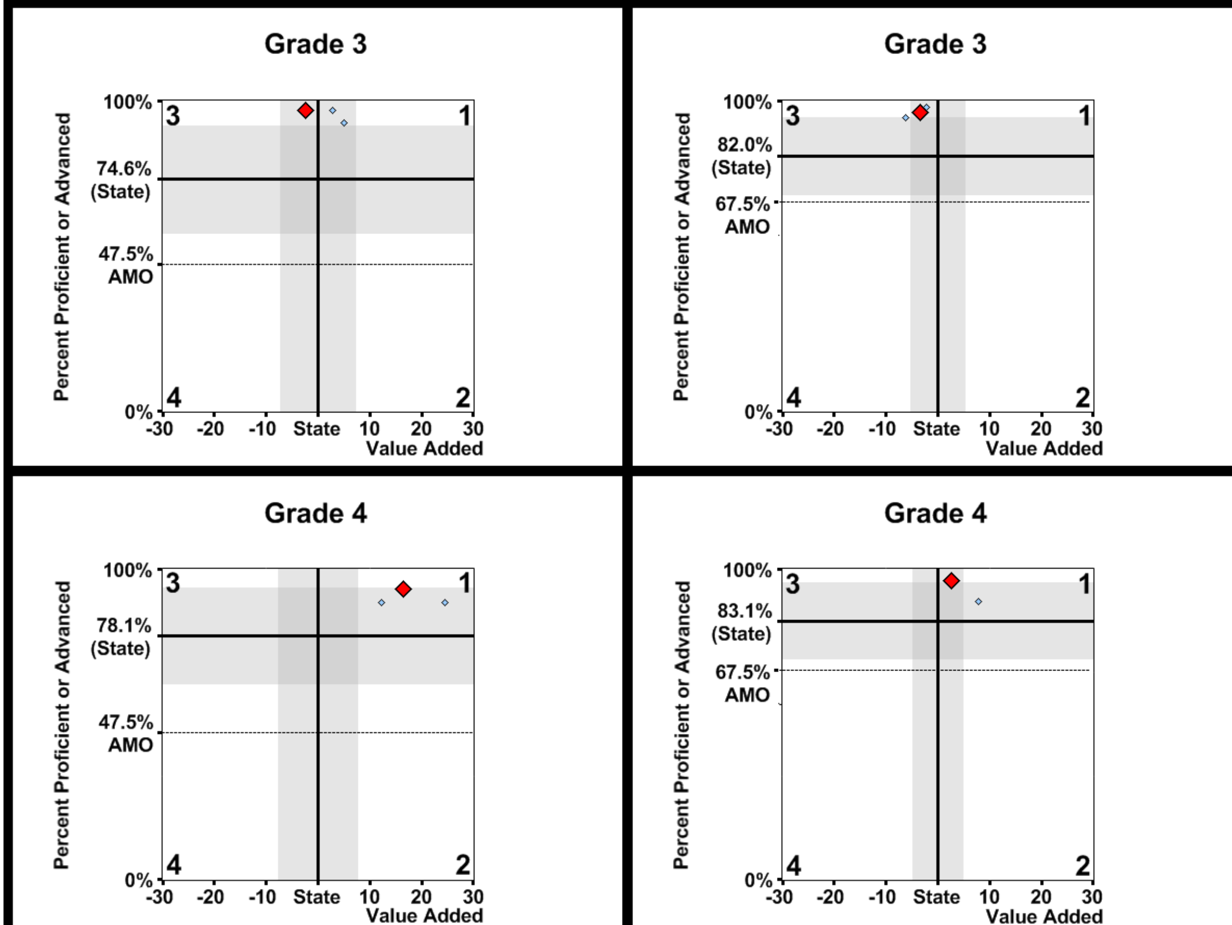
Grade 4



Quadrant 1 - High Value Added, High Attainment
Quadrant 2 - High Value Added, Low Attainment
Quadrant 3 - Low Value Added, High Attainment
Quadrant 4 - Low Value Added, Low Attainment

Prairie / State Averages
Value-Added and Attainment Quadrants

Mathematics **2006-07** **Reading**



Quadrant 1 - High Value Added, High Attainment
Quadrant 2 - High Value Added, Low Attainment
Quadrant 3 - Low Value Added, High Attainment
Quadrant 4 - Low Value Added, Low Attainment

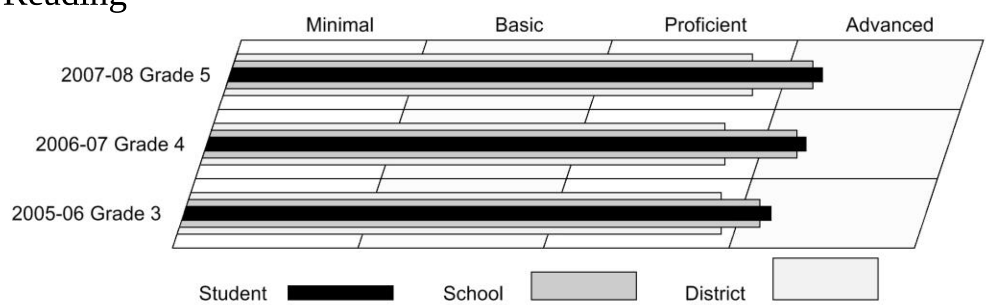
Appendix C.4 Individual Growth

Individual Student Report

Name : Ima Sample
Current Grade : 5th
Current School : Shorewood Hills
Current District : Madison Metropolitan

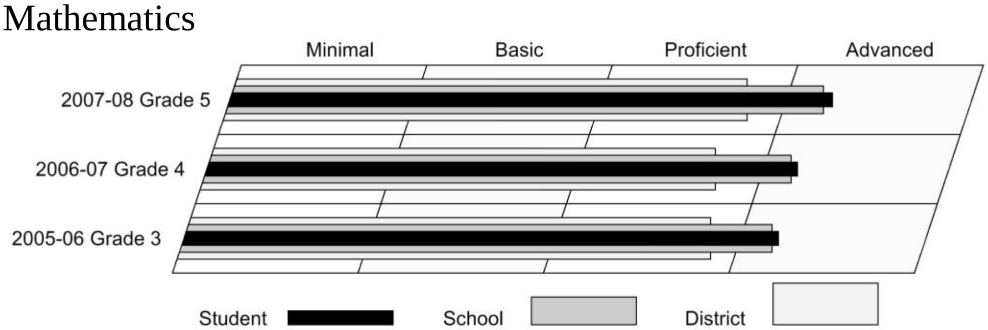
The reading and mathematics reports below provide graphical and tabular representations of the student's growth compared against school and district growth averages and against state proficiency levels.

Reading



Reading	WCKE Grade-Level Scores				WCKE Proficiency Level Cut-Scores			
Year	Grade	Student	School Average	District Average	Minimal Performance	Basic	Proficient	Advanced
2007-08	5	532	523.3	481.6	290-400	401-443	444-496	497-690
2006-07	4	520	511.3	473.8	280-395	396-439	440-488	489-650
2005-06	3	491	481.6	457.6	270-393	394-429	430-465	466-640

Mathematics



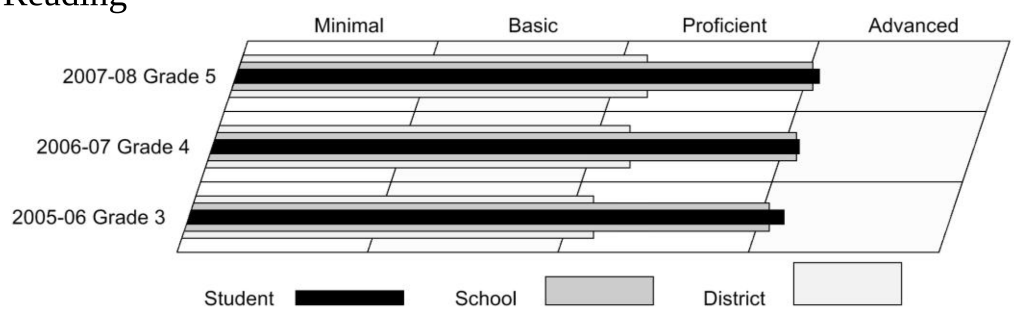
Mathematics	WCKE Grade-Level Scores				WCKE Proficiency Level Cut-Scores			
Year	Grade	Student	School Average	District Average	Minimal Performance	Basic	Proficient	Advanced
2007-08	5	545	539.2	491.5	270-444	445-462	463-504	505-680
2006-07	4	511	504.7	466.9	240-420	421-437	438-483	484-650
2005-06	3	481	474.4	440.1	220-391	392-406	407-451	452-630

Individual Student Report

Name	:	Ima Sample
Current Grade	:	5th
Current School	:	Meir
Current District	:	Milwaukee Public School

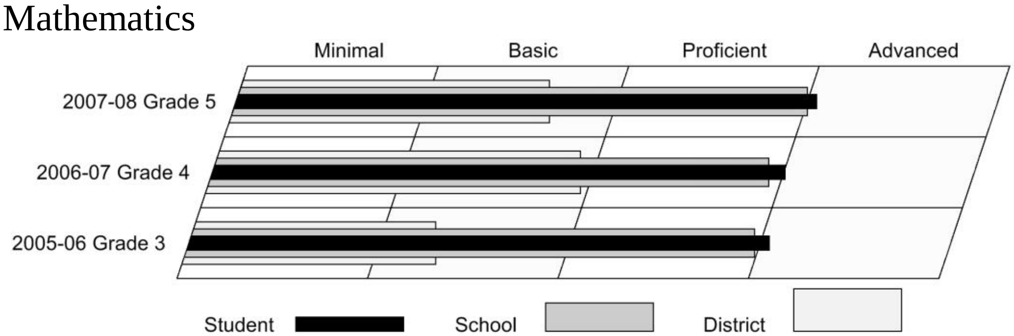
The reading and mathematics reports below provide graphical and tabular representations of the student's growth compared against school and district growth averages and against state proficiency levels.

Reading



Reading	WCKE Grade-Level Scores				WCKE Proficiency Level Cut-Scores			
Year	Grade	Student	School Average	District Average	Minimal Performance	Basic	Proficient	Advanced
2007-08	5	512	504.7	452.3	290-400	401-443	444-496	497-690
2006-07	4	507	502.8	449.8	280-395	396-439	440-488	489-650
2005-06	3	485	477.4	436.5	270-393	394-429	430-465	466-640

Mathematics



Mathematics	WCKE Grade-Level Scores				WCKE Proficiency Level Cut-Scores			
Year	Grade	Student	School Average	District Average	Minimal Performance	Basic	Proficient	Advanced
2007-08	5	511	504.7	456	270-444	445-462	463-504	505-680
2006-07	4	485	476.3	435.6	240-420	421-437	438-483	484-650
2005-06	3	460	449.6	398.1	220-391	392-406	407-451	452-630