

# Measures & Metrics

---

We are committed to designing a fair and accurate accountability system that measures the growth and attainment of all students to ensure that every Wisconsin child has the opportunity to graduate ready to succeed in college or a career.

In order to fairly and accurately measure the readiness of our students, we must use multiple measures (state tests and other performance indicators) and multiple metrics (growth and attainment). Using multiple measures and metrics allows for a more comprehensive picture of school and district performance, provides richer data to stakeholders, and increases confidence in the results. That is, the use of multiple measures is a systems approach to analyzing student performance. Using a systems approach allows us to identify the areas in which schools and districts struggle the most, and will allow this focus to drive improvements.

## Measures

*Instruments that measure performance objectives*

### State Assessments

Accountability systems have traditionally relied on summative assessment data from the state test for accountability calculations. Currently, the state assessment (WKCE/WAA-SwD) is given in grades 3-8 and 10. The WKCE is a summative assessment that fulfills NCLB requirements. As the focus of NCLB testing has been on reading and math, and doesn't cover all grades, the resulting data is limited to these grades and content areas. In Wisconsin, we have used the WKCE plus attendance (K-8) and graduation rates (9-12) as our accountability measures.

While it is still administered, the WKCE is being phased out. The successor state test will come from the SMARTER Balanced Assessment Consortium, and will cover similar grades and content areas. There will be summative assessment in ELA and mathematics for grades 3-8 and 11. The new ELA assessment will include a writing component. These assessments are meant to determine if students are ready for credit-bearing college coursework. The SMARTER assessment system will include performance assessment events as well as interim assessments that benchmark progress within the school year. The SMARTER consortium is a systems approach to assessment that will permit students to demonstrate their ability through multiple assessment opportunities in the course of the school year.<sup>1</sup>

### Other Measures

Since state assessments cannot measure all that we value in education—early childhood literacy, 21<sup>st</sup> century skills, technological literacy, science competency, etc—it is important that the accountability system include other measures of student performance beyond ELA and mathematics tests. Other accountability measures could potentially include:

- ACT/SAT participation and performance rates
- Advanced Placement (AP) participation and performance rates

---

<sup>1</sup> The shift from grade 10 WKCE reading test to grade 11 ELA test will be an issue in the short-term, as the SMARTER assessment will be operational in the 2014-15 school year.

- Attendance rates
- Course completion rates
- Credit accumulation
- Graduation rates
- International Baccalaureate (IB) participation and performance rates
- Occupational certifications
- Post-secondary enrollment
- STEM coursework participation and performance rates
- School climate surveys (student, family, teachers)
- Suspension rates

## Metrics

*Calculations of performance on measures*

### Attainment

Attainment calculations look at whether students attained a specified level of performance at one point in time, also called performance snapshot. Under NCLB, attainment was calculated from the number of students deemed proficient on the state reading and math tests. Attainment reports show the status of each school and district in getting their students to a set performance level on these assessments each year.

### Growth

Growth calculations track individual student progress on an assessment over time. Rather than looking at the status of the school as a whole, or the attainment levels at one point in time, growth calculations look at individual student trends. While a student may not be proficient in mathematics, we may see on a growth report that the student has improved achievement over the course of the year. With growth reporting, we may also see a student who has already attained proficiency has not grown, has a flat performance trend or declining performance even though they may still be in the range of proficiency. As such, growth metrics can tell us **more** about proficiency status than attainment alone.

A best practice in achievement reporting is to report status and growth, as both metrics provide valuable information, but singularly, do not provide a complete profile of achievement. Several techniques are available to calculate growth, some more complicated than others. But all approaches are premised on providing a more robust picture of student performance.

**STUDENT GROWTH PERCENTILES (SGP)** The statistical process behind SGP, developed by Damian Betebenner at the National Center for the Improvement of Educational Assessment, compares individual student progress against their academic peers, specifically those with the same test scores in prior years. An individual student is compared to all other students in the state who had the same score last year, to see whether this year's score represents growth that is lower, the same, or higher than their peers. DPI began providing SGP reports for individual students, schools and districts in August 2011. The reports can be parsed by student subgroup (race, economic status, students with disabilities, gender, etc), grade level, school and district.

**VALUE ADDED MODELS (VAM)** The statistical process behind value-added calculations attempts to account for differences attributable to a particular school, district, teacher, or program on student performance. In other words, the method attempts to control for

student variables such as race, socio-economic status, and English Language Learner status. Value-added reports show whether and how much value a program has added to student performance, while controlling for student characteristics found within the program. There are different approaches to value-added modeling; variance between VAMs is often found in which variables are being controlled. The Value-Added Research Center (VARC) at UW-Madison's Wisconsin Center for Education Research (WCER) has developed a value-added model and has a data-sharing agreement with DPI to calculate value-added data. VARC also partners with various districts and CESAs across the state.

Wisconsin could use SGP, VAM, or both in an accountability system. (As they answer different questions, both are going to be included in the new state reporting tool.) It is also possible for growth to be included in the accountability system but that the growth metric remains unspecified. As the procedures for modeling statistical growth improve—alongside advances in assessments—DPI will continue to evaluate the techniques available and identify those most appropriate for each purpose.

### Gap Reduction

Achievement gaps refer to the difference between the performance of groups of students, typically among different race/ethnicities and socioeconomic statuses. For example, if 65% of African American students graduate from high school in four years, and 95% of white students graduate in four years, the black-white achievement gap is 30 points on this measure (graduation rate). If the system performed equitably there would be no difference between these two groups. The decrease in performance gaps between student subgroups—or Gap Reduction—is the shrinking of the gap between school years. With this metric, a statistically significant decrease in one or more achievement gaps could be considered as one metric in an accountability system.

Gap reduction can be constructed in many ways. For example:

- If the attainment gap is 20 points between white students (70% scored proficient or advanced) and Hispanic students (50% scored proficient or advanced) in 2011-12 but this particular achievement gap shrank to 6 points in 2012-13, this school would be making progress in reducing the gap. Significance tests could be conducted to determine if the gap Reduction is statistically significant.
- This school's gap reduction of 6 points could be compared to similar schools or statewide to identify whether their gap reduction is accelerated, average or below average.
- Similarly, this analysis could be done on growth gaps, which may be beneficial in identifying schools that have high student growth rates, but may not yet have high proficiency rates.

As is the case with other metrics, use of Gap Reduction by itself is not sufficient. Using a mix of the three options is possible (e.g. schools are held accountable for results on two out of three metrics, whichever two are the best results).