

Incorporating Student Performance Data into Educator Evaluations:

An Overview of the Work of the Student Growth Workgroup

April 24, 2012

The Charge

The Student Growth Workgroup of the Utah Educator Effectiveness Project has been wrestling with producing recommendations for the most valid approaches to meet the requirement of incorporate student performance results in educator evaluations. This is one of the most complex issues in all of educational accountability, especially for those educators in the non-tested subjects and grades (NTSG). The purpose of this document is to highlight the major work accomplished by the Student Growth Workgroup and discuss key aspects of the upcoming pilot process in terms of the student performance component.

Some Definitions

The Workgroup recognizes that we have been dealing with and using many terms that might not be familiar to a broad group of stakeholders. Therefore, we define some of these terms below.

Tested and Non-Tested Subjects and Grades

- ✓ **Tested** subjects and grades are those with a state-administered assessment in that course as well as a state assessment in the same content area in the prior grade. These are generally math and language arts in at least grades 4-8 (up through 10 for language arts) and in science for grades 5-8.
- ✓ **Non-tested** subjects and grades are those that do not have both a state-required “pre-test/prior score” and “post-test” in the same content area that can be incorporated into their teacher evaluation. In most states, teachers in NTSG comprise approximately 70% of the workforce.

Assessment, Analytic Methods, and Evaluation

While these are somewhat familiar terms, the Workgroup has found that there is often confusion in distinguishing among these terms and wants to ensure that there is a shared understanding among Summit participants.

- ✓ **Assessment:** Is the process of collecting data about some set of knowledge, skills, and/or behaviors that can range from highly structured, formal standardize tests to “in-the-moment” observations of students working on group projects. Assessment is typically classified as summative, interim, and formative, but only summative and interim assessments should be considered for accountability uses. Formative assessments are generally used to inform instruction and practice in order to improve.
- ✓ **Analysis:** Is the process of converting assessment data into some sort of indicator or other relevant piece of information. In terms of testing, this might mean turning raw scores

(number right) into some type of scale score. In terms of educator evaluation and other forms of accountability, analysis refers to the process and techniques (often statistical) of turning assessment results into an indicator that can be used in accountability determinations. Such analytic techniques include Student Growth Percentiles (SGP), Value-added Models (VAM), and Student Learning Objectives (SLO). These techniques will be described and defined in this document.

- ✓ **Evaluation:** Is the process making reasoned judgments using decision rules to classify personnel, programs, and/or policies in terms of the effectiveness or some other attribute. Critically, both the assessment and analytic approaches must be valid to produce valid accountability (evaluation) inferences.

Basic Components of the System

The Workgroup discussed, at length, the various assessment and analytic options that can be put together in a State Model educator evaluation system. We highlight just a few of these here. Building off of the distinction between tested and non-tested subjects and grades, several of the assessments and analytic approaches apply exclusively to tested grades and subjects, while others are available for all subjects and grades.

Assessments

There are multiple forms of assessment information that may be used in educator evaluation systems. We classify them into four broad categories.

- ✓ **State assessments:** These are the high quality assessments that are designed to measure the Utah Core Standards in language arts (grades 3-10), mathematics (grades 3-7, plus end-of-course tests through Algebra II), and science (grades 4-8, plus four potential end-of-course high school tests). These tests define the “tested” grades and subjects except for the first grade tested for each subject (since there is no prior score for computing growth).
- ✓ **Commercially available** norm-referenced, interim, and other assessments: There are many legitimate reasons for incorporating commercially produced assessments in educator evaluation systems. In spite of significant concerns about the quality of such assessments, most are scaled and equated using defensible psychometric approaches so that scores across multiple test forms can be compared validly. However, there are numerous concerns with rushing to use an existing commercial product. An important consideration in any accountability system is that people should only be held accountable for those things that they can control. Most current commercial products are often weakly aligned to what educators are being asked to teach. This difference between the curriculum content and the standards (expectations) and the assessments will turn into a chasm as schools more fully implement the Utah Core Standards unless these commercial tests are revised considerably. Nevertheless, some commercially available assessments may fulfill important assessment needs in developing educator accountability systems.
- ✓ **Common district assessments:** These assessments are designed by groups of content area teachers, specialists, and district leadership and are intended to provide rich representations of district curriculum, state content standards, and Core Standards. They

are generally higher quality than most classroom assessments and administered at important check points in the course, such as at the end of important units and the end of the course. While many common assessments rely on selected response (e.g., multiple-choice) items only, many others include more open-response items and more complex tasks than what is found on commercially available tests.

- ✓ **School or classroom-based assessments:** While the other three assessment forms described above can potentially be used effectively in educator evaluation systems, there will still be many courses for which no common or commercial assessment is available. Additionally, there may be important aspects of courses that are not appropriately measured by the three forms of tests mentioned above such as performing a scientific investigation or conducting a research project. In cases such as these, school-based or individual classroom-based assessments will be used for educator evaluation purposes.

Analytic Approaches

As mentioned above, assessment is only the first step in determining the degree to which educators have contributed to helping students learn. The following four general methods are ones that may be used by the state (SGP) or districts to analyze assessment data for use in educator evaluation systems.

- ✓ **SGP:** The Student Growth Percentile is a regression based measure of growth that works by evaluating current achievement of students compared to students with the same score history. This provides a familiar basis to interpret performance – the percentile, which indicates the relative change in test scores given the student’s starting point and academic peer group. This can be used to gauge whether or not the student’s growth was atypically high or low. SGPs will be used in Utah, calculated by USOE, to evaluate the change in performance on the state CRTs. SGPs have the advantage of being “scale agnostic,” which means that they may help provide a smoother transition across assessment changes than other methods.
- ✓ **VAM:** Value-added model scores are calculated as the difference between the average gain on test scores made by a group of students from the prior year to the current year, and centered on the average gain within the district. The difference between the actual and predicted scores is considered the “value add.” These models require pre- and post-test data, with a scale that is amendable to gain scores, and if available, multiple years of prior student achievement may be incorporated into the model. While SGPs are generally calculated only on statewide data, VAM can be used with smaller samples at the district level, especially for the larger districts, to analyze commercial assessments or even common assessments.
- ✓ **Gain scores:** Gain scores describe when the pretest score is subtracted from a posttest score. This may happen in spring-to-spring designs for non-tested subjects and grades (NTSG), but may also be used with fall-to-spring designs. There are two fatal flaws with using simple gain scores in educator evaluation systems. The first is that judgments of “gain” are usually based on non-equated (non-comparable) test scores and the second is that gain scores do not take into account differences in where students start, which leads to a potentially very unfair accountability system when measuring growth and learning.
- ✓ **SLO:** Student Learning Objectives (SLOs) are content- and grade/course-specific measurable learning objectives that can be used to document student learning over a

defined period of time. To boil SLOs down, they provides a means for educators to establish learning goals for individual or groups of students, monitor students' progress toward these goals, and then evaluate the degree to which students achieve these goals. Any of the assessments described above can be used to evaluate SLOs and, in fact, other analytic approaches described here can be part of SLOs. The active involvement of the teacher throughout the process is a key advantage of the SLO approach over traditional test-centered approaches to accountability. It is designed to reflect and incentivize good teaching practices such as setting clear learning targets, differentiating instruction for students, monitoring students' progress toward these targets, and evaluating the extent to which students have met the targets.

Recommendations for a State Model

Given the limitations, both technical and the lack of available data, the Student Growth Workgroup has recommended that Student Learning Objectives serve as the analytic framework for all NTSG and should be encouraged in tested subjects and grades as well. Student Growth Percentiles will be used to analyze the results of CRTs and districts will be expected to incorporate the results of these SGP analyses in the evaluations of educators in the appropriate grades and subjects. Additionally, the Student Growth Workgroup recommends that a portion of each educator's evaluation should be based on "shared attribution" whereby either SGP and/or SLO results may be shared among an appropriate group of educators such as a grade-level or content-area team. The Student Growth Workgroup is currently developing guidance for each of these aspects of the student growth component of the educator evaluation system.

The Proposed Structure of the Pilot

The Student Growth Workgroup is developing a model system that will serve as the basis for the systems to be piloted. However, the Workgroup recognizes that in this rapidly emerging field, we do not know enough now to suggest narrowly defining the structure of the pilot. The Student Growth Workgroup recommends using the pilot to learn about what might work best in a variety of districts. However, the Workgroup wants to ensure that the pilot is rigorously structured so that we are best able to learn from these efforts. The Workgroup will propose a State Model, but will indicate certain decision points where pilot districts may choose slight alternatives. For example, the State Model may require that all teachers create and use SLOs in the educator evaluation system, but certain districts may decide to modify this model and require SLOs only in NTSG, while relying on SGP and shared attribution in the tested grades and subjects. This is just one example. In subsequent documents, the Student Growth Workgroup will identify a limited number of choice points for the pilots.