

State Council for Educator Effectiveness – Draft Recommendations

Measuring Student Growth for Use in Teacher Evaluations

Draft March 10, 2011

Colorado’s definition of teacher “effectiveness” is premised upon the demonstrated ability to improve student academic growth. In developing its recommendations for how to measure student growth for use in individual teacher evaluations, the Council closely examined the limitations of currently available assessments to provide fair, valid, and reliable measures of student academic growth for use in the evaluations of all teachers. Members also paid close attention to the risk of unintended negative incentives that could result from not being thoughtful in developing an approach to using student growth measures (discouraging collaborative work among teachers, for example).

The following recommendations reflect the Council’s effort to balance the current technical limitations and concern with unintended consequences against the value that effective teachers have demonstrated positive impact upon student academic growth. Council members strongly agreed that these recommendations reflect the best response to current conditions and that they should be periodically revised in response to developments in available assessments and experience with how best to incentivize positive teacher performance. Specifically, the Council has agreed on the following broad principles which are reflected in their recommendations:

- An individual teacher’s performance on Quality Standard VI, Responsibility for Student Growth, should be evaluated based upon two primary sets of measures:
 - Measures reflecting the academic growth of the teacher’s students in the content area delivered by the teacher;
 - Measures reflecting the academic growth of students attributable to all educators who are responsible, directly or indirectly, for ensuring that such students attain mastery of the Colorado Academic standards. Attribution of student growth may be shared across all educators in a district, school, grade level, content area or other professional learning community. Every teacher’s overall student growth measure shall include a measure of student growth which is common and shared by teachers across classrooms.
- Two main issues are involved in gauging the technical merits of a student growth measure:
 - Validity – the extent to which an assessment/measure closely reflects the academic content and skills included in a teacher’s instruction;
 - Reliability – the extent to which an assessment/measure consistently measures student performance across a wide variety of occurrences and students.
- Measures of student growth should strive to provide both valid and reliable assessments of student growth for use in educator performance evaluations. The assessments that are currently available at the national, state, district and school level each have varying degrees and combinations of validity and reliability which need to be considered carefully with respect

to different categories of personnel, particularly when choosing student growth element measures for individual teachers.

- Districts should aim to ensure that student growth measures and ratings are comparable among teachers, knowing that it will be difficult to achieve given the technical realities of measuring and attributing student growth to individual teachers. Districts must thoughtfully weigh considerations about what can be compared and what can't.
- As states, districts and schools continue to refine current assessments and develop new ones, CDE guidelines developed as a result of these recommendations will need to be revised.
- Districts will need to use the data they already collect to make calculations of student growth or progress. Technical guidelines should be provided by CDE in this area to ensure that a minimum standard of rigor is being met. CDE should continually monitor teacher student growth ratings, and compare them with other teacher performance measures and school-wide growth indicators in order to confirm findings and ensure that technical rigor is being met.

Teacher Quality Standard VI/Student Growth measures recommendation language

NB: Current law requires that the SAC provides guidance/advice to inform decisions that a school principal is making about student growth tools. SCEE should propose a policy change so that this is a DAC function instead of a SAC function.

1. All districts shall develop evaluation systems which measure teacher performance against Standard VI using multiple measures.

The following recommendations relate to the process of choosing and applying assessments for the purpose of calculating a teacher's individual student growth [score].

Definition: "District" includes district administrators and building leaders

2. For the purpose of calculating a teacher's individual student growth score, districts in collaboration with principals and teachers including representatives of the local teachers association or federation, if one exists, shall categorize personnel into appropriate categories based on the availability of state summative assessment data [Personnel Categories Matrix].
3. Districts, in collaboration with principals and teachers including representatives of the local teachers association or federation, if one exists, shall choose or develop appropriate measures of student growth (as defined in Section 8 below) to be used in the evaluation of each personnel category. Districts shall consider the following issues in selecting assessment measures:
 - a. The results of discussions with teachers in the district about which of the available measures will best match their instructional responsibilities;
 - b. The technical quality of the analytic methods available.
4. Districts shall develop a process for identifying and handling the calculation of individual student growth scores for educators teaching two or more subjects, where there are multiple sources of student growth information.
5. Districts shall develop a process for assigning teachers to the role of "teacher of record" versus "contributing professional" for the purpose of state data collection. A teacher need not be required to be identified as a "teacher of record" for a particular student in order for that student's academic growth data to be used in a teacher's performance evaluation.
6. Districts shall clearly articulate to each educator the category or categories of personnel into which they fall, and how the growth of the students they teach will be measured for the purpose of informing their performance evaluation.
7. For the purposes of evaluating the technical validity and reliability of measures of student growth or learning, [NB for rationale language: using only "growth" means that this statement should technically not allow districts to use things like SGOs since SGOs will not allow for a calculation of "student growth" in the way that state assessments or other level 1 or 2 measures would allow. For some categories of personnel, calculations of student learning against standards are feasible while calculations of "growth" in a

- technical sense are currently not feasible. Having said this, growth may be calculable within an SGO framework if the measures can support it] the following three categories shall apply. CDE shall develop formal definitions of these data levels and use them to classify popular assessments or assessment approaches, indicating the strengths and potential issues involved with using them to measure student-level outcomes and calculate growth:
- a. Category A (state criterion-referenced/standards-based) data: CDE-certified student-level assessment data (e.g. CSAP) that is of a technical quality (standardized, external and objective) that allows student growth to be calculated for personnel in specific grades and subjects using the CGM, and justifies its use as a major portion of the educator’s student growth score effectiveness evaluation ;
 - b. Category B (externally-created assessments) data: student level assessment data collected from assessment tools that are comparable across classrooms with demonstrated rigor which meet CDE guidelines for technical quality. (e.g. NAEP, NRT, SAT/ACT). These measures may or may not allow for the calculation of student “growth”;
 - c. Category C (school or district-created) data: student level assessment data that are collected at the district, school or individual classroom level, and do not meet the higher technical requirements of Category A and Category B data but which do comply with minimal technical guidelines developed by CDE. These measures may be highly valid as measures of student progress/learning against standards, but will not technically allow for the calculation of student “growth.”
8. In choosing appropriate measures of student growth, districts shall use at least one of the measures outlined in section 9 below, applied in a manner that is consistent with CDE guidance on evaluating the technical rigor of particular approaches as outlined in Section 7.
- a. Districts shall consider both the validity and reliability of data when determining proper student growth measures and their weight. Districts are strongly encouraged to emphasize the validity of the measures they use while maximizing reliability to the extent possible. Districts shall be transparent about what measures of student growth will be included within a teacher’s evaluation.
 - b. Measures of academic growth of students designed to improve educators’ knowledge and skills (and ultimately, effectiveness), should be validated for that purpose. CDE shall facilitate the development of such student growth measures.
 - c. Districts are strongly encouraged to use the data identified in 7(a)-(c) as the predominant measure of a teacher’s individual performance against Quality Standard VI.
9. Districts shall measure teacher performance of Quality Standard VI using guidance developed by CDE that informs the selection of reliable and valid available measurement methods. Currently, the most reliable available measurement methods are as follows:
- a. For a given category of personnel if there is a state summative assessment available and:

- i. there is a state summative assessment available in the same subject for the prior grade then districts shall use the median student growth percentiles calculated by CDE as part of the Colorado Growth Model; these shall comprise a portion of a teacher’s individual student growth score;
 - ii. there is a state summative assessment available in a related subject for the prior grade then districts may calculate conditional status;
 - iii. there is another valid covariate, as defined in CDE technical guidelines, then districts may calculate conditional status;
 - iv. there is no other assessment data or covariate available then districts should consider using student growth objectives or other goal setting approach.
 - 1. CDE shall develop guidance on the use of student growth objectives, and shall develop high-quality exemplars.
 - b. For a given category of personnel if there is no state summative assessment available but there is a high-quality [end of course], norm-referenced, or interim assessment and:
 - i. there is a high-quality predictive test then districts should calculate growth or value-added results in the most technically defensible manner possible;
 - 1. CDE shall develop guidance on the technical requirements for appropriate use of pre-tests for calculating student growth and shall develop high-quality exemplars.
 - ii. there is no high-quality predictive test then districts should consider using student growth objectives or other goal setting approach consistent with CDE guidelines.
 - c. For a given category of personnel if there is no state summative assessment available, and no high-quality end of course assessment, norm-referenced, or interim assessment available then:
 - i. Districts should consider using student growth objectives or another goal setting approach, which is consistent with CDE-developed technical guidelines for the development of SGOs, and choosing an appropriate measure of student growth as per CDE guidelines.
 - ii. Districts may use shared attribution of available and appropriate assessments as a greater proportion of such educators’ overall student growth score than outlined in Section 12(a) below.
10. Regarding the choice of assessments for all categories of educators, districts shall meet the technical requirements and considerations laid out in CDE guidelines for measuring student growth for teacher evaluations. [Elena’s documents in modified form].

The following recommendations relate to the process of choosing and applying assessments for the purpose of determining the appropriate manner of attributing student growth among teachers.

- 11. Every district shall determine the method to be used for attributing student growth for students to multiple educators or on a school-wide basis that best supports district and school improvement priorities, school design and mission, collaboration among educators, and available sources of student growth data.**
- 12. Schools are highly encouraged to include measures of student growth that are attributable to multiple teachers, whether on a school-wide basis or across grades or subjects. For the purpose of including such measures into an individual teacher’s performance evaluation, districts in collaboration with teachers including representatives of the local teachers association or federation, if one exists, shall categorize personnel into appropriate categories based on the grade level of the students with whom they work and the availability of Category A assessment data. At a minimum, these categories shall include the following broad categories based on the current types of assessments available for students, though districts may add more as they deem necessary:**
 - a. teachers responsible for early childhood and primary students;**
 - b. teachers responsible for students in intermediate grades (4-8);**
 - c. teachers responsible for high school students.**
- 13. Districts, in collaboration with teachers including representatives of the local teachers association or federation, if one exists, shall choose appropriate Category A or Category B measures of student growth as outlined in Section 10 to be used for purposes of shared attribution for each personnel category, or groups of teachers within each personnel category.**
 - a. Districts shall consider both the validity and reliability of data when determining the proper student growth measures and their weight. For the purposes of determining student growth measures that will be jointly attributed across teams of teachers or school-wide districts are strongly encouraged to emphasize the validity of the measures they use while maximizing reliability.**
- 14. Districts shall develop processes for identifying and handling shared attribution of individual student growth scores for educators falling into more than one personnel category, where there are multiple appropriate sources of student growth information. Districts shall ensure that these teachers’ evaluations include student growth measures for all subjects for which they are responsible. Responsibility shall be determined at the district level taking into account a teacher’s state “teacher of record” status.**
- 15. Districts shall clearly articulate to each educator the category or categories of personnel into which they fall and how the growth of the students they serve will be measured and attributed to them for the purpose of informing their performance evaluation.**

The following recommendations charge CDE with developing guidelines related to the above recommendations.

- 16. The SCEE recommends that CDE develop guidelines that at a minimum address and require that:**

- a. districts consider the match of available assessments to the instructional responsibilities of personnel categories, both in terms of content and attribution of student learning;
 - b. districts involve teachers in the district, including representatives of the local teachers association or federation, if one exists, in choosing or developing appropriate measures of student growth that match teachers' instructional responsibilities;
 - c. state-wide assessments, where available and aligned to instructional responsibilities, be used in the evaluation of student growth;
 - d. for subjects with annual state assessments available in two consecutive grades, districts shall use results from the Colorado Growth Model for evaluating student growth;
 - e. if available and feasible, districts include at least one additional measure of student growth common to personnel teaching in the same or similar content area (even when state tests are available) in order to create more valid and reliable measure of a teachers' performance of Quality Standard VI;
 - f. the content tested shall align to the Colorado Academic Standards;
 - g. districts shall incorporate some shared attribution of student growth as part of each individual teacher's student growth calculation.
17. CDE shall also develop technical guidelines regarding the development and use of various measurement instruments for evaluating student growth, which shall be updated as research and best practices evolve. Tools to be addressed within these guidelines include but are not limited to:
- a. The development and use of teacher-, school- or district-developed assessments;
 - b. The use of commercially available interim and summative assessments;
 - c. The development and use of student growth objectives;
 - d. The development and use of other goal-setting approaches; and
 - e. Piloting of new and innovative practices.
18. CDE shall develop and/or provide examples of the following:
- a. Approaches to categorizing personnel for the purposes of measuring individual student growth;
 - b. Approaches to categorizing personnel for the purposes of joint attribution of student growth;
 - c. Exemplars of student growth approaches for all categories of personnel and for the major categories of assessment data available.

Once school and district leaders have determined what tools will be used to include student performance information in educator evaluations (see prior recommendations), they will need to calculate student “growth” or increases in achievement, and determine how to attribute student test scores to teachers. Simply having two scores for each student (e.g., a predictive test and a post-test) does not automatically imply a method for evaluating these scores, or calculating student growth. There are many methodological choices that must be considered when determining how to most validly analyze and incorporate student performance information into educator evaluations. The following draft language is based on several small working group conversations around the issue of measuring and calculating student growth in both tested and non-tested subjects and grades.

The following recommendations address the issue of districts converting data gathered through student performance measures into calculations of student growth, and charge CDE with developing guidelines related to technical requirements for such calculations.

19. Districts shall ensure that:

- a. Student growth measures or assessments are technically adequate, as per CDE guidelines, to support the intended analyses and inferences;
- b. Analyses are designed in a manner that warrants reasonably consistent inferences. Issues to be considered shall include but are not limited to:
 - i. ensuring a large enough number of students to warrant reasonably consistent inferences;
 - ii. ensuring that measures are of sufficient levels of reliability.
- c. The approach or model makes design choices explicit and transparent (e.g. in a value-added model the student- or school-level factors which are statistically-controlled for should be clearly identified), and has technical documentation sufficient for an outside observer to judge the technical quality of the approach (i.e. a value-added system must provide adequate information about the model) [N.B. for rationale: this is especially important for techniques such as SGP, VAM, etc.]; and
- d. The model, or those implementing the model, should produce and present results in a manner that can be understood and used by educators.

20. Districts using student growth objectives to evaluate teacher performance shall ensure that each of the following requirements is met:

- a. A set of procedures for establishing and evaluating goals is established at the district level. These procedures shall include district-wide policies and practices for using SGOs, as well as guidance for SGO development in specific content areas.
- b. Goals are established at the individual and/or the aggregate classroom level ensuring that:
 - i. Individual student goals are ambitious and standards-based;
- c. Aggregate goals are either standards-based or normative, and focus on ensuring that all students progress. Goals are based on data such as prior assessment/grades history and reflect meaningful (e.g., college readiness) and measurable targets.

- d. If multiple goals are established for each student, at least one of the goals must be a long-term goal (e.g., a semester or year).
- e. Goals are set by teachers in consultation with professional learning communities, a committee of peers, and/or principals, and monitored by colleagues.
- f. Progress toward and attainment of goals is determined by measures that are aligned with the learning targets and technically appropriate to determine whether students have actually met the goals.
- g. The assessments used to measure student goals are reviewed by a committee of peers and administrators to judge their adequacy for evaluating student progress towards the goals.

21. Districts using any other measures of student growth including student growth percentiles, value-added models, growth models or conditional status models to evaluate teacher performance shall ensure that the following requirements are met:

- a. Assessments are technically adequate, as per CDE guidelines, to support the intended analyses and inferences;
- b. Analyses are designed in a manner that warrants the conclusions which are being drawn from the results. Issues to be considered shall include but are not limited to:
 - i. ensuring a large enough number of students to warrant reasonably consistent inferences;
 - ii. ensuring that measures are of sufficient levels of reliability.
- c. The approach or model makes design choices explicit and transparent (e.g. in a value-added model transparency about student- or school-level factors which are statistically-controlled for), and has technical documentation sufficient for an outside observer to judge the technical quality of the approach (i.e a value-added system must provide adequate information about the model) **[N.B. for rationale: this is especially important for techniques such as SGP, VAM, etc.];** and
- d. The model, or those implementing the model, should produce and present results in a manner that can be understood and used by educators to improve instruction.
- e. CDE shall develop additional technical guidelines to assist districts in developing specific approaches to analyzing student growth data and converting initial scores into student growth ratings. Guidelines will specify the technical requirements by analytic method, and at a minimum provide districts with guidance sufficient to ensure that:
 - i. Predictive and post test scores are from assessments in the same subject with adequate student-level correlations or in subjects that are conceptually-related;
 - ii. Analyses are conducted on adequately large student sample sizes, which will vary by technique and shall be specified in CDE guidelines;
 - iii. Assessments are technically adequate to support the intended analyses. Technical guidelines shall be adequate to ensure that:
 - 1. Both predictive and post assessments meet minimum reliability thresholds (e.g. are tied to the curriculum);

2. Both assessments are aligned to the same content domain in conceptually coherent ways (e.g. CSAP scores from 8th grade could be used as a predictive score for a 9th grade algebra class; and 8th grade ELA CSAP grade probably should not);
 3. Each assessment uses a scale that has adequate interval properties; and
 4. Each test has sufficient “stretch” or variability in the scores to account for a wide range of student abilities.
22. In making decisions concerning the attribution of school-wide growth on state summative assessments to individual teachers, districts shall ensure that decisions about attribution reflect school-wide student outcome goals and school design, and employ a shared decision-making approach for determining levels and types of attribution.