COMMON ASSIGNMENT STUDY

Using Common Assignments to Support Educator Effectiveness in Colorado

Scott Marion, Center for Assessment[1]

January 24, 2015

Executive Summary

Student performance results have become a central feature of reformed educator effectiveness systems. This is not to say that measures of teacher practice (e.g., classroom observations) are not a major focus as well, but classroom observations of teacher practices has been a part or, in fact, the only part of most previous educator evaluation designs. Incorporating the results of Measures of Student Learning (MSL) into effectiveness determinations, while a valuable aspect of effectiveness, has been technically challenging on many fronts, from the quality of the assessments to the type of analytic methods employed to transform those data into indicators in effectiveness systems. The challenge exists for teachers from both "tested" and "non-tested" subjects and grades (NTSG), but perhaps more so for the latter. A "tested" subject and grade is one with a state test score and a state test score from at least one prior grade (generally, the immediate prior grade) to allow for the use of complex statistical models (e.g., value-added models, student growth percentiles) to analyze the change in students' scores. The Common Assignment Study (CAS) likely holds promise for helping to improve the quality of MSL for use in the effectiveness determinations for teachers in both tested and non-tested subjects and grades.

The Common Assignment Study (CAS) is a research and development project in which teachers collaborate across districts and states to develop and implement high-quality curricular units to ensure student learning of meaningful disciplinary content and skills. Initiated in the 2013-14 school year, the Common Assignment Study is a three-year effort being led by the Colorado Education Initiative (CEI) and The Fund for Transforming Education in Kentucky (The Fund) with support from the Bill & Melinda Gates Foundation. In response to the demand for resources to provide instruction that captures both the rigor and content of the Kentucky Common Core

[1] Thanks to Andy Brownstein, Amy Spicer, Ash Vasudeva, Christy Schneider, Jeri Thompson, and other members of the CAS team for helpful edits and suggestions.

Standards and the Colorado Academic Standards, the CAS was conceived to facilitate multi-district and multi-state collaboration among teachers within participating districts and states during the transition to new college and career-ready standards.

This brief describes how CAS units and assessments may be well-suited to provide much needed, high-quality learning and assessment information to contribute to educator effectiveness determinations. The CAS units are already designed around critical disciplinary concepts and skills (the "big ideas" of the discipline) and being able to link two or more CAS units throughout a school year via a common big idea will allow educators and others to determine how much students have progressed on such key concepts and skills. Those familiar with educator effectiveness methods for NTSG should be recognizing the similarity between CAS units and Student Learning Objectives (SLO). In fact, this brief presents an argument for using the structure of SLOs to provide the evaluative framework for using CAS units as part of educator effectiveness determinations.

Using CAS units eliminates several threats to the use of student performance information in educator effectiveness determinations. First, CAS units provide high quality assessment information in ways that are fully integrated with instruction and the daily work of teaching and learning. In other words, "extra" assessments are not needed to support effectiveness determinations when using CAS. Second, many NTSG approaches rely on a limited amount of assessment information (e.g., pretest and posttest). The SLO model in Colorado allows for and even encourages the use of more information than just those two end points. Each CAS unit contains multiple assessment tasks, so using multiple CAS units to support effectiveness determinations would fit into the SLO model seamlessly. Third, CAS is built on a premise that collaboration is critical for producing high quality units and that the CAS process is an effective vehicle for supporting collaboration. A recent report from the Colorado Education Initiative substantiates the high value placed on collaboration both within and across districts in the

development of MSL and the districts' full educator effectiveness systems[2]. This collaborative model can strengthen MSLs by encouraging teachers (and leaders) to work together to create stronger MSLs than individuals can create on their own. Finally, the CAS units provide a natural bridge between the evaluation of educator practices and student performance results in ways that can support more holistic determinations of effectiveness. In short, CAS units are designed to support high quality curriculum and instruction related to the Common Core State Standards, which is a considerable benefit in its own right, but the CAS units and processes can provide rich and credible information to support educator effectiveness systems.

[2] See Colorado Education Initiative (2014). *Colorado MSL Systems: Patterns and Progress 2013-2014.* http://www.coloradoedinitiative.org/wp-content/uploads/2014/12/MSL-Report-FINAL.pdf

Introduction

State and district education leaders are rapidly reforming their educator effectiveness systems to include a greater emphasis on the results from student performance measures. In Colorado, this has been required since the passage of SB 191. Incorporating the results of Measures of Student Learning (MSL) into effectiveness determinations has been technically challenging on many fronts, from the quality of the assessments to the type of analytic methods employed to transform those data into indicators in effectiveness systems. However, if used appropriately and thoughtfully, the results of student test scores can provide useful information to enhance the quality of educator effectiveness determinations. However, only 25-30% or so of the teaching force nationally is in grades and subjects with at least two years of state test data — what is referred to as the "tested subjects and grades." One of the greatest challenges in designing and implementing reforms in the area of educator effectiveness determinations is documenting the contributions of the other 70% or so of educators in the "non-tested subjects and grades (NTSG)." There are several assessment and analytic approaches for incorporating measures of student learning into the evaluations of teachers from NTSG, but in reality, most approaches suffer from significant weaknesses (see Appendix A for an overview of commonly used analytic methods).

One of the major shortcomings of current educator evaluation and effectiveness systems is that the two major components—teaching observations and student performance—are treated as separate entities which risks introducing more incoherence into the system than necessary. We argue that evidence from the Common Assignment Study can help bridge this gap. This brief provides an overview of how the Common Assignment Study may offer some promise for strengthening teacher evaluations.

The Common Assignment Study

The Common Assignment Study (CAS) is a research and development project in which teachers collaborate across districts and states to develop and implement high-quality curricular units to ensure student learning of meaningful disciplinary content and skills. Initiated in the 2013-14

school year, the Common Assignment Study is a three-year effort being led by the Colorado Education Initiative (CEI) and The Fund for Transforming Education in Kentucky (The Fund) with support from the Bill & Melinda Gates Foundation. Technical and implementation support for the project is being provided by multiple organizations including Stanford Center for Assessment, Leadership & Equity (SCALE), Westat, The Center for Assessment, and Research for Action.

In response to the demand for resources to provide instruction that captures both the rigor and content of the Kentucky Common Core Standards and the Colorado Academic Standards, the CAS was conceived to facilitate multi-district, multi-state collaboration among teachers within participating districts and states during the transition to new college and career-ready standards. The goals of the project include the following:

- **Unit Design and Use:** Explore how to successfully design, revise, and implement high quality Common Assignment units in participating districts and states with a focus on the use of student work on common assignments to inform instructional conversations;
- **Resources and Support:** Identify the resources, support and professional learning opportunities educators need to effectively design and implement the units in order to support professional learning; and,
- **Examining Evidence of Student Learning:** Determine how the evidence of student learning yielded through the common assignment units may contribute to more robust teacher evaluation systems, in both tested and non-tested grades and subjects.

A major focus of CAS project is to address some of these weaknesses and evaluate the feasibility of using the evidence of student learning generated through the common units as part of educator effectiveness determinations. The project will explore, among other goals, the extent to which student work samples, other classroom-based artifacts, and end-of-unit assessments provide information related to educator effectiveness to complement, supplement or provide alternatives

to traditional test-based measures of student performance. Some of the key questions and topics for study under this strand of the project are:

1. What are the most defensible approaches for analyzing data drawn from the common units in order to make inferences about teacher contributions to student learning?
2. How do inferences about student learning drawn from the common unit assessments relate to external or other measures of student learning, such as state tests?
3. What are the opportunities and challenges associated with inferences about student learning over time (from multiple units) compared with single or even dual data collection windows?
4. Does drawing evidence of student learning from sources of evidence that are fully integrated into the teaching and learning process ameliorate some of the potential unintended negative effects associated with accountability uses of assessment information?

Based on the CAS project, as well as considerable experience developing and implementing educator effectiveness systems around the country, this brief presents an overview of how common assignment information can be used to support educator effectiveness determinations in Colorado. Common assignment information can be used to support the "student growth" component of the effectiveness determination, but also can provide important evidence related to teaching practices. The brief first discuss some concerns with current approaches for integrated student performance data into educator effectiveness methods and then illustrates how the use of common assignment information may ameliorate some of these challenges and support more credible effectiveness evaluations for Colorado educators.

Issues and Challenges with Current Educator Effectiveness Approaches

As noted above, the NTSG issue is an obvious challenge that could be minimized through the use of common assignments. Less obvious, though, is the benefit that common units may offer to evaluating teacher practices and connecting those effectiveness evaluations to student learning.

NTSG Challenges

We want to avoid some of the current misconceptions about the evaluation of NTSG, where unfortunately many states/districts are using analytic approaches that cannot support the intended inferences about educator effectiveness. This is most often played out in the ways that learning goals are operationalized that determine the targets for student "growth," which are aggregated to targets for teacher "growth." For example, one of the most common and egregious approaches involves simply subtracting a pretest score from a posttest score, when both scores actually represent different scales and perhaps different learning targets. Many of these concerns were discussed in a previous paper (e.g., Marion et al., 2012)[3] about measurement problems with most of the analytic approaches currently used with NTSG. Another problem with some of these more simplistic approaches is that many they rely on only two sources of data (pretest/posttest). In other words, these systems generally preclude the use of the kinds of multiple measures that recognize the value of each measure. Fortunately, Colorado's approach for employing Measures of Student Learning in its State Model Evaluation System avoids many of the common problems associated approaches noted above in the way that the state encourages the use of MSLs most tailored to local needs.

Another, perhaps even more important, challenge to the successful implementation of educator effectiveness evaluation systems for teachers in NTSG is the disconnect between the evaluations and the actual day-to-day work of teaching and learning. Many see these additional tests administered as part of the effectiveness evaluation as "dropped from the sky," not much differently from the way in which they see state tests. Even Student Learning Objectives (SLO) are not immune from these problems, depending on how they are implemented. For example, teachers and students in one state with strong state control over the SLO system have been heard saying things like, "We took the 'SLO' test today," which clearly indicates that is seen as distinct from the regular teaching and learning system. In other words, the tests being used in this

[3] For more details on these issues, see Marion et al. (2012). http://www.nciea.org/publication_PDFs/Measurement%20Considerations%20for%20NTSG_052212.pdf

particular system are developed at the district level and teachers are required to administer the tests at specific times whether they fit the particular curriculum at that specific time. On the other hand, Colorado has taken a more holistic approach for teachers in NTSG and uses existing assessments and other data rather than adding new tests into the mix.

A Role for CAS

Using assessment results drawn from common units offers the promise of ameliorating both of these concerns. The common units, by design, include multiple measures of student performance, hopefully at successive levels of understanding. Second, the rich performance assessments in the common units are fully integrated into the regular work of teaching and learning, and are not seen as extra, accountability assessments. Certainly, we still need to attend to the potential corrupting influences of accountability on the results from assessments included in the common units, but CAS offers promise that these more robust measures of teachers' contributions to student learning will feel more trustworthy and there would be less temptation for corruption. Third, CAS was based on the premise, which has been borne out by initial evaluations, that collaboration among teachers was an important vehicle for increasing the quality and usefulness of units. Finally, common units and LDC modules offer a way to bridge the false dichotomy in teacher evaluation between teaching practices and student learning.

Incorporating CAS Results into Educator Effectiveness Systems

If evidence of student learning could be collected from multiple units — for example, one from the fall and one from the spring — we would still have to make sense of the data. Ideally, in the future, three or four units may be administered throughout the year. Before getting into the specifics, let's back up and review the essential components of educator effectiveness evaluation determinations.

- ✓ **Assessment/Data**: These data include assessments administered to students, but could also include data from student survey and classroom observations as is being done with TPGES. It is helpful to think broadly about data and to recognize that while high quality data are necessary for valid educator evaluations, they are far from sufficient. In other

words, a very high quality pretest could be administered in the fall and a very high quality posttest could be administered in the spring, but the inferences in terms of educator effectiveness from these two tests might be poorly substantiated for reasons we discuss below.

- ✓ **Analytics**: There is a famous Seinfeld episode that takes place at a car rental counter where Jerry tries to explain the difference between "taking the reservation," which the clerk keeps emphasizing that they have done, and "holding the reservation," which the car rental company clearly has not done because there is no a car available. The analytics of accountability systems is the "holding of the reservation," because it turns the assessment information into something useful. In other words, it is relatively easy to collect data (taking the reservation), but it is quite difficult to analyze the data in ways to support defensible inferences (holding the reservation). This includes the methods by which we turn raw data into accountability indicators. In terms of state tests, this may mean applying student growth percentile (SGP) methods to translate test scores across multiple years into SGPs for students and aggregate SGPs for educators. Similarly, an observation rubric is a tool for turning the raw material of classroom interactions into numeric quantities that are then usually summed or averaged. Given the many choices of methods for turning raw data into indicators (e.g., noticeably different observation rubrics), it is easy to see how the choice of analytic method can influence the inference about educator quality. The analytic challenge is generally greater for NTSG than for state tests or classroom observations, in part because of the small sample sizes and limited analytic capacity among school and district personnel.
- ✓ **Attribution**: Perhaps the least considered and most important aspect of using data for evaluation decisions is deciding which teachers deserve the "credit" or "blame" for which students. This is one of those issues that seems so simple to those outside of education, but the devil is truly in the details. Besides the overwhelming issue of data quality, there is a critical need to understand how schools really operate. Many elementary students, for example, are nominally assigned to a single teacher, but in reality are moved around in grade-level teams for some forms of differentiated instruction or other activities. Some

have tried to address this issue by measuring "dosage" which is the assignment of weights to the student score based on the proportion of time the child spent with each teacher. Attributing student learning results to teachers is more complex than simply counting seat team and such dosage approaches would require strong evidence to support such strong assumptions. Therefore, we have advocated certain types of "shared attribution" approaches in which all students in a given team of teachers might be "pooled" for the purposes of attributing the scores from student assessments — if that fits the theory of action and structures for the school. This can occur for both tested and non-tested subjects and grades.

Student Learning Objectives as a Framework

The previous discussion should demonstrate that we cannot just throw assessment data at the wall and hope that something useful for educator evaluation sticks. We need a framework for contextualizing and analyzing the evidence of student learning that emerges from the common assignments. Student Learning Objectives (SLO) provides such a framework in Colorado.

SLO have gained popularity as a means of attributing student performance results to educators in new forms of teacher evaluation systems for all teachers, but especially those in NTSG. SLOs can serve as a framework for incorporating the results from common assignments into educator evaluation systems. SLOs are content- and grade/course-specific measurable learning objectives that can be used to document student learning over a defined period of time, often based on a SMART goal-setting approach. SLOs are 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.

Clear and Meaningful Learning Goal

SLOs should reflect the relevant content standards, skills, and the associated curriculum. They should describe what students will be able to do at the end of the course/grade, or at least over a

reasonably long period of instruction (e.g., a semester). The learning goal(s) generally will be established by a collection of teachers in the same grade and/or subject area and overseen by the district or school leadership. The learning goal for an SLO should reflect an **"enduring skill" in the discipline** and may encompass several key content standards. Multiple teachers are often working with their students on the same learning goals and the results from the same SLO implemented by multiple teachers may be shared (i.e., shared attribution).

The following is an example of a learning goal from a Common Assignment Study middle school social studies:

> *Students will independently use primary and secondary sources to form evidence-based interpretation of historical events and ideas significant to westward expansion in the antebellum era. In doing so, students will consider and discuss multiple causes for historical events with a focus on how and why the US acquired Western lands. Further, students will understand how the use of authentic primary source documents enriches our understanding of historical events and ideas.*

As can be seen, the learning goal represents key ideas in the study of history, but is still somewhat general. The addition of the assessments and student targets help make the SLO more specific in terms of performance expectations[4].

<u>Assessments used to evaluate students' achievement of the learning goals</u>

The assessments used to evaluate the degree to which students have achieved the learning goals should be of high quality; that is, they should be designed to provide evidence aligned to the specific learning goal. First, if the learning goals take the rich form described above—as they should if they are to support high quality instruction and deeper learning— it is unlikely that they will be measured well with just a single assessment. Multiple assessments will likely be required, and performance or other authentic assessments must be part of the assessment system

[4] Much more information about SLOs (SLOs) can be found in the Center for Assessment's SLO Toolkit at www.nciea.org.

designed to evaluate the learning goal. An example of a multiple tasks from the middle CAS history unit tied to the learning goal presented earlier follows:

> *Task 1: Considering the historical perspectives on the Mexican-American War, was President Polk justified when he declared war in 1846? After reading primary and secondary sources given in class, write an essay in which you address the question and argue whether Polk was justified in his decision to declare war with Mexico. Support your position with evidence from the text(s). Be sure to (acknowledge; refute) competing views.*
>
> *Task 2: Part A. Explain in detail how the U.S. acquired one territory through war and one territory through diplomacy*
>
> *Task 2: Part B. Choose two groups of people that were affected by westward expansion and explain how the expansion affected those groups.*
>
> *The responses to these tasks will be scored with a multi-dimensional rubric focused on historical analysis, argumentative writing, use of sources, and quality of writing and presentation.*

Targets for both student performance and aggregate targets for educator performance

Student Targets. The student target is the expected outcome at the end of the instructional period. Those proposing SLOs should ensure that the student targets are both ambitious and realistic, which is quite a challenging design task, especially during the early years of SLO implementation. Several researchers (e.g., Marion, et al., 2012; Lachlan-Haché, Cushing, & Bivona, 2012[5]) have suggested that teachers set targets using available baseline data to help contextualize the learning targets for individuals or groups of students. Many SLO approaches use baseline data as a basis for both growth and achievement/proficiency goals for students. A description of a set of targets for the learning goal described above follows:

> *These targets are based on the rubric scores for the CAS Unit tasks described above, as well as the three (3) formal performance tasks administered in this*

[5] Lachlan-Haché, L., Cushing, E., & Bivona, L. (2012). Student Learning Objectives as Measures of Educator Effectiveness: The Basics. Washington, DC: American Institutes for Research.

course over the entire school year. This is not an average of scores, but rather an evaluation of students' demonstrated consistent improvement over time in the criteria of the rubric.

- *At least 75% students will advance one performance level from the beginning of the year to the end of year. Students maintaining the highest level of performance from the beginning to the end of year will count as having met this requirement.*
- *Approximately 60% of the students will score at the "proficient" level on the majority of the rubrics' criteria across the multiple summative tasks.*

This example presents one way of establishing SLO targets. Setting ambitious and reasonable targets for SLOs is one of the most challenging aspects of SLO design and implementation. This has been the focus of several recent national meetings as states wrestle with how to approach the issue in fair and valid ways.[6]

Teacher Targets. Teacher targets specify how the student aggregate scores (results) will be used to determine the degree to which the teacher has met the SLO targets and whether these results will be employed directly or transformed into an indicator for use in accountability determinations. In some cases, these targets are set by the state, but more typically, they are determined by the district in conjunction with school leaders. Ideally, school leaders will tailor the targets in consultation with teachers to account for specific classroom contexts. Typically, teacher targets and the corresponding performance rating are classified into three or four levels. For example, a teacher may be classified as "not meeting the SLO" if less than 50% of the students reach their target, "meeting" if 51-85% of the students reach their target, and "exceeding" if more than 85% of the students reach their target. Obviously, the appropriateness of these targets is contingent upon the learning goal, assessments, and student targets. It will take several years of data collection and analysis to evaluate the appropriateness of these targets.

[6] This topic could fill an entire paper, but for now see Marion, et al. (2012).

Incorporating CAS into an SLO Framework

By now, those familiar with CAS should recognize the obvious connections between SLOs and CAS. These connections are explored in more detail in this section. If the multiple common units administered through the school year were tied to common "enduring skills" of the discipline, we could begin to describe student progress referenced to actual changes in knowledge and skills rather than just scale-based (i.e., points) descriptions of growth. The common assignments could provide the measurement information at key points along the trajectory implied by this learning goal. Many district SLO approaches, especially those being implemented in several Denver Public Schools, expects ongoing data collection to evaluate the goals and the CAS could contribute by providing high quality data to support effectiveness determinations. Given the current state of affairs in teacher evaluation, this could be one of the most important contributions of this project. In the following section, I relate the CAS approach to an SLO framework.

Learning Goal

Ideally, the learning goal(s) should be identified ahead of time so that the units are designed to support the learning and assessment of this goal. In the real world, this will likely be an iterative process between high-priority learning goals and available units. To provide credible evidence of changes in student learning over time, the multiple common units in a school year must be tied to the same big idea of the discipline. Of course, the question is "What idea?" The Literacy Design Collaborative (LDC) module incorporated into each unit is designed to focus on various aspects of argumentative writing across the curriculum. There is no question that this is very important and argumentative writing is critical vehicle for communicating understanding in most disciplines, but the key (and specific) concepts and skills of science, social studies, or other content areas should be represented in the learning goals. More recent work in LDC has focused on ensuring that the disciplinary content is well-represented in the LDC task and scoring rubrics. Therefore, argumentative writing could represent one of the cross-unit learning goals while a more discipline-based idea (e.g., understanding how geographic and economic factors influence human settlement) could represent an additional cross-unit strand.

Assessments

A major advantage of using assessments from common assignment units for educator evaluation is that it avoids the typical impoverished assessment evidence common to many NTSG approaches. Unit assessments are designed as a system to yield high quality information about students' understanding of the big idea as well as supporting knowledge and skills. Further, the assessments across multiple units will be intentionally connected —not formally linked (psychometrically), but conceptually connected via the assessment targets and rubrics. This will require a focused effort to design and implement high quality assessments and rubrics that can support inferences related to the enduring skills. Importantly, these units are closer to the actual teaching and learning than is typically the case with assessments like end-of-course tests and as evidence accumulated through the CAS research that hopefully shows the relationship between the CAS assessments and these more distal assessments, states like Colorado will feel confident about the validity of the interpretations from the CAS assessments.

Targets

As noted above, establishing targets for students is one of the most challenging aspects of SLO design and implementation. Establishing reasonable baselines to contextualize the end-of-year assessment results is a critical component of the SLO process. Depending on when the fall unit is implemented, the assessments from the fall unit could serve as the baseline for the SLO. However, if baseline information is available from earlier in the school year than implantation of the fall unit, assessment information from the fall unit could serve as an interim check on the goal. The assessment information from the spring unit would be used to evaluate the degree to which students have met the targets associated with the learning goal. Using multiple assessments would be an improvement over most SLOs, but SLOs based on common assignments could be improved even further if the multiple units were tied to practice-based learning progressions. The progressions would then allow teachers to develop a better understanding of how students move toward deeper understanding of the big idea. The progressions can serve as explicit anchors for the unit assessments and other activities.

Documenting student progress along these learning progressions can then serve as evidence for teacher effectiveness determinations.

Importantly, we should avoid simple reductionist approaches for determining and evaluating targets. The "rough conditioning" approach described previously (Marion, et al., 2012) offers a sensible alternative to establishing arbitrary numeric goals based on some distance between a pretest score and 100% (or some other percentage correct). The evidence from common assignments will allow us to more accurately document students' learning trajectories in a discipline. It is beyond the scope of this brief and perhaps too decontextualized to offer very specific recommendations for establishing meaningful student targets based on CAS units, but I strongly recommend capturing qualitative changes in student understanding through a domain-specific rubric. Such rubric scores can then be used to aggregate student-level results to the classroom or school levels.

Common Assignments as Evidence of Teaching Practices

Essentially all teacher evaluation and effectiveness systems being implemented in the United States count evidence of student learning and evidence of teacher practices as separate elements in overall evaluation scores. To be fair, there are sound arguments for such approaches. Many believe that observations and other indicators of teacher practice measure the process of teaching, while assessment data provides evidence of the "product" of teaching. Further, nobody doubts that essentially all traditional evaluations have focused exclusively on process. Therefore, the theory of action for counting student learning results as a significant part of teacher evaluations is intended to demonstrate that student learning must be valued.

However, by treating the two major components as separate aspects of teacher evaluation, there is a risk of introducing more incoherence into the system than necessary. For example, most of the major tools for measuring teacher practice (e.g., Danielson's Framework for Teaching)[7]

[7] See: http://www.danielsongroup.org/article.aspx?page=frameworkforteaching

require evidence of how well teachers plan and execute instructional activities, but it would also make sense to include evidence related to what students learned from this instruction. This sounds obvious, but is rarely done. The evidence of student learning generally comes from assessments only peripherally related to classroom observations.

This disconnect is expected (but not a goal) for the "tested grades and subjects" because of the broad brush of the state tests[8]. On the other hand, we have the leeway with NTSG to design assessments closer to the specific activities in the course or classroom. For example, if the school was using common assignments to generate assessment information for SLOs, it would make sense to link classroom observations to the times when the teacher and students are working on the common units. This would provide the observer a way to triangulate multiple forms of evidence in order to draw more defensible inferences. This is a relatively easy step schools could take when implementing common assignments, but additional steps can be followed to bring even more coherence to educator evaluation systems.

The common units provide opportunities for generating and collecting data related to many aspects of teaching practice. In fact, the Colorado Department of Education has already made similar linkages between MSL and the Colorado State Model Evaluation System. For example, the common units require considerable planning and thinking about instruction to ensure that students are provided with meaningful learning opportunities. It is clear that common units could easily provide evidence for the specific elements depicted below from Quality Standards III and IV in the Colorado State Model. Among other artifacts, the CAS unit planning templates, the student work generated from the assessments, and the teachers' actions as a result of analyzing the student work will serve as meaningful evidence for these elements. I am using the Colorado State Model as an example, but similar strands are found in other tools for capturing evidence of teaching practice (e.g., Danielson, Marzano, Marshall).

[8] This is not to say that value-added (VAM) or student growth percentile (SGP) models for large scale assessment data cannot provide useful contributions to teacher evaluation decisions especially when the results of such analyses of student growth are shared among multiple educators.

COMMON ASSIGNMENT STUDY

QUALITY STANDARD III Teachers plan and deliver effective instruction and create an environment that facilitates learning for their students.				
ELEMENT B: Teachers plan and consistently deliver instruction that draws on results of student assessments, is aligned to academic standards and advances students' level of content knowledge and skills.				
Basic	**Partial Proficient**	**Proficient (Meets State Standard)**	**Accomplished**	**Exemplary**
THE TEACHER: ☐ Uses assessment results to guide adjustments to instruction. ○ ***Has specific student outcomes in mind for each lesson.***	. . . and **THE TEACHER:** ○ ***Aligns instruction with academic standards and student assessment results.*** ○ ***Monitors instruction against student performance and makes real-time adjustments.*** ○ ***Assesses required skills.***	. . . and **THE TEACHER:** ○ ***Encourages students to take academic risks.*** ○ ***Makes sure students meet learning objectives while increasing mastery levels.***	. . . and **STUDENTS:** ○ ***Monitor their level of engagement.*** ○ ***Confer with the teacher to achieve learning objectives.***	. . . and **STUDENTS:** ○ ***Initiate activities to address their learning strengths and next steps.***
ELEMENT H: Teachers use appropriate methods to assess what each student has learned, including formal and informal assessments and use results to plan further instruction.				
THE TEACHER: ○ ***Involves students in monitoring their learning.*** ○ ***Assesses learning outcomes appropriately.***	. . . and **THE TEACHER:** ☐ Implements appropriate strategies for assigning grades. ☐ Evaluates student performance based on multiple measures. ☐ Includes documentation of student progress toward mastery of state content standards in assessment plans.	. . . and **THE TEACHER** ***provides actionable, timely, specific and individualized feedback about the quality of student work to:*** ☐ Families and significant adults. ☐ Other professionals who work with students. ○ ***Teaches students to use feedback to improve their learning.***	. . . and **STUDENTS:** ○ ***Self-assess on a variety of skills and concepts.*** ○ ***Articulate their personal strengths and needs based on self-assessment.*** ○ ***Effectively use formal and informal feedback to monitor their learning.***	. . . and **STUDENTS** *assume ownership for:* ○ ***Monitoring their progress.*** ○ ***Setting learning goals.*** ○ ***Applying teacher feedback to improve performance and accelerate their learning.***

QUALITY STANDARD IV Teachers reflect on their practice.				
ELEMENT A: Teachers demonstrate that they analyze student learning, development and growth and apply what they learn to improve their practice.				
Basic	**Partial Proficient**	**Proficient (Meets State Standard)**	**Accomplished**	**Exemplary**
THE TEACHER: ☐ Collects and analyzes student data to inform instruction. Uses data to: ☐ Support student learning.	. . . and **THE TEACHER:** ☐ Collects multiple examples of student work to determine student progress over time.	. . . and **THE TEACHER** applies knowledge of how students learn and their prior knowledge to the development of: ☐ Instructional strategies.	. . . and **THE TEACHER** develops student learning plans based on: ☐ Multiple examples of student work. ☐ Information gathered from students, families and colleagues.	. . . and **THE TEACHER:** ☐ Monitors and evaluates personal behavioral changes to determine what works for students.

More specific to assessment literacy, almost all of the major tools for evaluating practice include rubrics for teachers' assessment literacy and the use of assessment results to improve learning. For example, the Rubric for Evaluation Colorado Teachers includes criteria for "analyzing student work from assessments" as part of Quality Standard IV, Element A as characterized by the descriptions seen in the rubric above.

This is just one excerpt, but it is clear that the Colorado Model draws attention to appropriate design and use of assessments for instruction and learning in many aspects of evaluating teaching practice. The CAS unit planning template and other planning tools, along with actual assessments and associated student work, can provide rich evidence to support this and related strands.

<u>The Role of Collaboration and Examining Student Work</u>

The CAS project has emphasized the critical role that close examination of student work can play in the development and evaluation of units as well as for helping teachers gain insights into

students' strengths and needs. We argue that student work can serve as a credible bridge between the evaluation of teachers' practice and characterization of student performance and growth. If we cannot see it in the student work, it isn't there! In the case of educator effectiveness, student work is the perfect bridge between teacher practices and student performance results in that the student work can help shed light on the degree to which specific teacher practices were effective at improving student learning.

As noted earlier, CAS is based on the premise that collaboration among engaged educators can lead to higher quality instructional, curriculum, and assessment materials than would be the case if teachers were working individually. While there is certainly value in having individual teachers examine student work from their classrooms, the value increases exponentially when groups of teachers collaboratively examine student work from common and even unique (to each classroom) tasks. Generally, teachers are able to gain deeper insights into student learning and their own teaching practices through collaborative inquiry into student work than each could likely realize on their own.

While the MSL process used in Colorado does not require that teachers work with others when establishing and evaluating their MSLs, a recent report from the Colorado Education Initiative[9] suggests that this is an important aspect of MSL development and implementation in CO. We support the notion that common and collaboratively established MSLs and SLOs, specifically, for educators teaching the same courses can strengthen the SLO process and improve the effectiveness of those educators involved in the collaborative work. Additionally, Quality Standard V (Teachers Demonstrate Leadership) supports teachers working with colleagues to improve the profession and each teacher's growth. This is another example of how the CAS process provides a connection between the student performance and teacher practice aspects of educator effectiveness determinations.

[9] Colorado Education Initiative (2014). *Colorado MSL Systems: Patterns and Progress 2013-2014*. http://www.coloradoedinitiative.org/wp-content/uploads/2014/12/MSL-Report-FINAL.pdf

Discussion

As illustrated throughout this paper, the common units at the core of CAS offer considerable opportunities for providing evidence in support of both student learning outcomes and teaching practices. An advantage in Colorado is that there is considerable flexibility afforded to local educational leaders for designing assessments and methods for documenting changes in student achievement over time. Designing local educator evaluations is very difficult work. Educational leaders and others are often being asked to do things without the necessary expertise and experience to do so. The work of CAS can help support these district leaders and educators as they strive to build coherent systems that are both fully integrated into the real work of teaching and learning and serve as a vehicle for deepening this work. As project leaders, we will have to work with policy leaders in both states to understand the potential of CAS to support meaningful educator evaluations. Closer to the action, project leaders will need to work closely with district leaders responsible for educator evaluation so they have an understanding of how to credibly include CAS information in local evaluation decisions.

COMMON ASSIGNMENT STUDY

Appendix A: Summary of Major Analytic Approaches Used for Educator Evaluation[10]

Analytic Approaches for Estimating Growth	Necessary Conditions	Procedures	Growth Interpretation
Growth Models	Pre-and post-test measures are available within the subject area of interest and exist on a common vertical scale	Calculate difference between pre and post-test performance on common scale	Gain (or Loss) in student performance between two points in time
Conditional Status Models (e.g., VAM-Models; SGP)	Pre-test data on one or more assessment(s) in or related to (i.e., correlated with) the subject are of interest and subject-specific post-test data are available	Condition on pre-test data (and potentially other covariates) as a means of evaluating post-test performance for a given student.	Difference between expected performance and observed performance given prior performance in the same or a related subject area.
Goal Setting/ SLO Process	Process by which teachers use existing student performance data (of a variety of types) to establish learning goals for students in their class, and then evaluate student performance relative to those goals		Degree to which a student or group of students attained one or more specified learning goals
Shared Attribution	Any of a variety of techniques that involves the attribution of a common estimate of student growth, achievement or teacher impact on student learning — based on aggregation at the group, school or district level — to one or more teachers		Depends on nature of shared attribution approach

[10] From: Hall, E., Gagnon, D., Thompson, J., Schneider, C., & Marion, S. (2014). State Practices Related to the Use of Student Achievement Measures in the Evaluation of Teachers in Non-Tested Subjects and Grades. www.nciea.org.