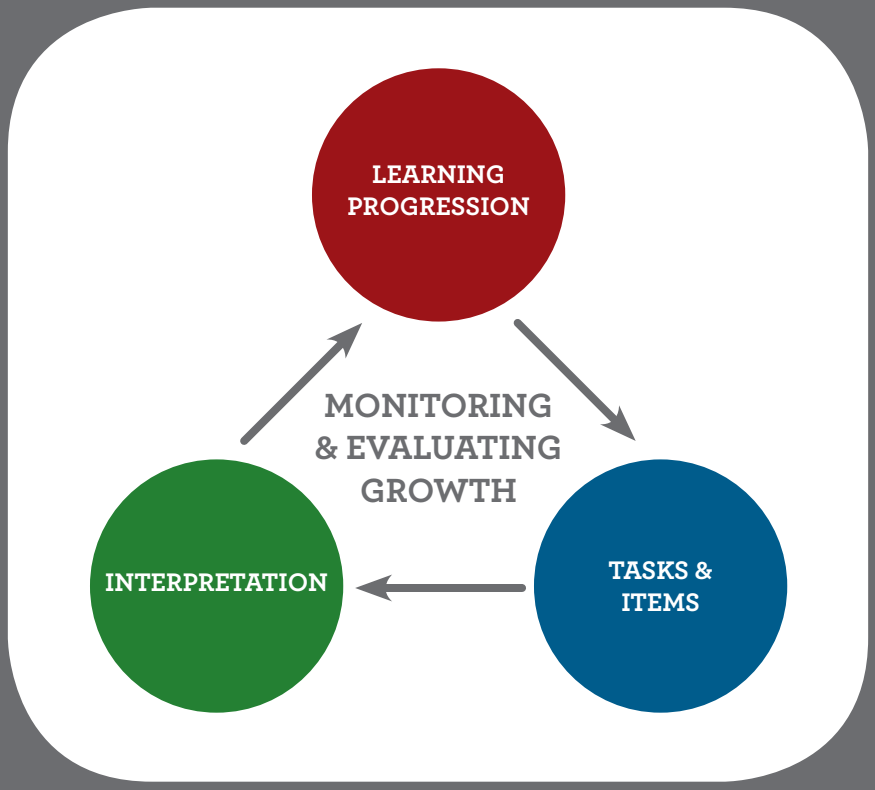


Using a Learning Progression Framework to Assess and Evaluate Student Growth

April 2015

CADRE

CENTER FOR ASSESSMENT DESIGN
RESEARCH AND EVALUATION (CADRE)
UNIVERSITY OF COLORADO, BOULDER

Center for Assessment

NATIONAL CENTER FOR THE IMPROVEMENT
OF EDUCATIONAL ASSESSMENT (NCIEA)
DOVER, NEW HAMPSHIRE

Derek C. Briggs
Elena Diaz-Bilello
Fred Peck
Jessica Alzen
Rajendra Chattergoon
Raymond Johnson

Foreword

The complex challenges of measuring student growth over the course of a school year have become familiar to anyone working to improve educator evaluation systems. Recent legislation in many states has postponed or scaled back the use of student growth or outcomes within accountability frameworks, suggesting that they are not ready to attach these measures to high-stakes decisions regarding educator effectiveness. Put simply, we don't have it all figured out yet.

While we struggle to measure student growth in a valid and reliable way for the tested subjects, such as math and literacy, we face even more significant challenges in the non-tested subjects, such as the arts and physical education. In the absence of state-mandated tests for the latter, many states and districts have begun to implement Student Learning Objectives (SLOs), a process of developing subject-specific learning goals and measuring the extent to which each student has mastered them. There is no single way to design and implement SLOs, no sure-fire way for SLOs to solve the problems of measuring student growth—hence the variety of models in use across the country. The authors of this paper, however, present what might be called an alternative model of SLOs, a *Learning Progression Framework* (LPF), which aims not simply to fill a gap in accountability systems, but also to offer promising approaches to the improvement of day-to-day classroom practices.

As the authors of this paper demonstrate, while some SLO models appear more promising than others, their designs share unfortunate limitations: 1) they offer little more than spurious claims about student growth, 2) they can incentivize the setting of different expectations for different students, arguably creating a *de facto* system of unequal access to grade-level content, and 3) despite their best intentions, they typically serve the needs of accountability for outcomes without offering a means to improve everyday instruction.

In light of these limitations, the authors demonstrate how the LPF model acknowledges the reality that intended outcomes cannot, on their own, improve teaching and learning. In the Denver Public Schools, we are implementing the LPF model based on a commitment to the idea that improvement of teaching and learning will depend on the way teachers meet students' individual needs *along the path* toward standards mastery. At the risk of indicating the obvious: formative assessment and appropriate adjustments to instruction play an indispensable role in any attempt to improve student learning. However, while standards define end-of-course outcomes for student learning, they haven't helped us much in figuring out what should happen along the way. Without guidance along the way, how can a teacher assess, with confidence, whether students are on track to master the content and skills of a given course? This is where the LPF offers much promise.

In Denver, we have been working over the past two years to develop and implement an SLO model based on learning progressions. At its core, the model is meant to be authentic to everyday cycles of formative assessment and instructional shifts, even as we intend to use it for summative accountability purposes. We know that this approach will take some years to refine, as it relies on long-term professional learning, effective structures and systems for teacher collaboration, distributed leadership, and a careful balance of teacher autonomy and quality assurance for the larger system. And yet, we have seen some early successes worth celebrating, which include: teachers' deeper knowledge of the standards in their subjects, the development of shared goals and more consistent collaboration among teams of teachers, and more intentional uses of assessment to support instructional planning and student learning.

A deeper understanding of assessment is critical at a time when standardized and summative tests are becoming increasingly prevalent and controversial. The LPF strives to put educational assessment where it belongs: close to the curriculum, the instruction, and the student. When we think in terms of learning progressions, we are reminded that assessment need not consist of—and with few exceptions should not

consist of—an isolated event or a traditional test. Rather, assessment is integrated naturally within the fluidity of instruction; as such, it includes not only tests, but also mid-lesson checks for understanding, rigorous tasks assigned to students throughout the school year, and systematic observations of student performances—to name only a few. The key point is that good instruction entails good assessment as one of its necessary conditions.

As we work to refine our SLO model in Denver, we communicate explicitly that our use of SLOs is a choice, and that we would promote their use in classrooms even if there were no legislation requiring student growth measures. This is because SLOs, when implemented within the Learning Progression Framework, capture the essence of high-quality instruction, pushing us to ask fundamental questions like the following: *What is most important for my students to know and be able to do? What do my students know right now, and how do I know that? What should I do to meet my students' needs?* If we can answer these questions with increasingly greater confidence, we will position ourselves to develop growth measures that meet the needs of accountability and, most importantly, of students in their day-to-day learning.

Michael I. Cohen, Ed.D.
Assessment Support Manager
Department of Assessment, Research & Evaluation
Denver Public Schools
Denver, CO

I. Overview

The metaphor of growth is central to conversations about student learning. As students advance from grade to grade, one naturally expects that what students know and understand about the world and about themselves is becoming more sophisticated and mature. Of course, there are many factors that could influence this process, such as family resources, peers, community, culture and a student's school environment. Over the past decade, empirical evidence has reaffirmed what most parents and children have believed intuitively for decades: the quality of teaching a student experiences can have a significant impact on growth (Aaronson, Barrow & Sander, 2007; Chetty, Friedman & Rockoff, 2014; Gordon, Kane & Staiger, 2006; Kane & Staiger, 2008).

But how does one know both what and how much students have learned over a delimited period of time in some content area? Although a well-designed assessment may help to answer the question of what students seem to know and understand, no single test can support inferences about how much this has changed over time. Furthermore, even when students are assessed periodically, inferences about growth may remain elusive if the content on the assessments is also changing. Finally, if these assessments are not well aligned to what is taught as part of a school's curriculum, the picture of student growth being presented can easily be distorted. In other words, although the metaphor of growth is conceptually appealing, finding a way to characterize growth numerically in a manner that leads to valid and reliable inferences is extremely challenging.

This challenge is becoming all the more apparent as a rapidly growing number of states and school districts in the United States seek to incorporate evidence of student growth into formal evaluations of teachers under the auspices of educational accountability (c.f., Dougherty & Jacobs, 2013). Although a great deal of research and debate has surrounded the use of statistical models for this purpose, only about one third of classroom teachers teach students for whom state-administered standardized tests are available as inputs into these statistical models (Hall, Gagnon, Schneider, Marion, Thompson, 2014). Hence, for a majority of teachers, other evidence is needed to support inferences about student growth. In more than 30 states, this evidence is being gathered through the development and evaluation of student growth through "Student Learning Objectives" or SLOs (Lacireno-Paquet, Morgan & Mello, 2014).

SLOs typically involve a process in which teachers establish measureable achievement goals for their students, assess students at the outset of an instructional period and then establish targets for student growth over the duration that period. A central impetus for this report is the belief that SLOs will only be able to support sound inferences about student growth if they have been designed in a way that gets educators attuned to the right motivating questions.

- What do I want my students to learn?
- What do my students know and understand when they arrive in my classroom?
- How is their knowledge and thinking changing over time?
- What can I, and other teachers at the school, do to help them learn?
- What evidence do I have that my students have demonstrated adequate growth?

Inferences about student growth that account for such questions need not only learning objectives but a framework that structures objectives into a progression of student learning. In this report we introduce a learning progression framework (LPF) that applies innovative thinking about educational assessment to better support valid and reliable inferences about student growth. A principal benefit of this framework is that it directly anticipates the questions posed above. In addition, it can support a process that is much more encompassing than the de-facto use of SLOs for teacher evaluation. Indeed, in this report we show that SLOs can be cast as a special case within an LPF.

Importantly, an LPF has three features that are always present irrespective of the content domain. First, a critical condition for operationalizing the framework is to have teachers work collaboratively to identify a Learning Progression (LP) within and, ideally, across grades or courses. Collaboration in this matter requires that teachers clearly establish what it means to say that a student has shown "adequate" growth in a

criterion-referenced sense. Second, an LPF emphasizes growth toward a common target for all students. Students are not differentiated with respect to what they should know and be able to do by the end of course but only with respect to differences in their preparedness at the start of the course. Third, the nature of teacher collaboration is explicitly oriented toward the analysis of student work. This work might range from multiple-choice answers on a standardized test, to responses on an open-ended essay, or to videos of students carrying out a classroom project. Teachers use this evidence to document the variability in how students respond to assessment activities, make distinctions among students with respect to the sophistication of their responses, and come up with teaching strategies that can best support the students with less sophisticated responses and further challenge the students with more sophisticated responses. An LPF focuses attention on student reasoning and the sometimes significant space between "not getting it" and "getting it." This means that teachers who think in terms of learning progressions are often just as interested in the process a student uses to solve a task as they are in whether an answer to the task is correct. This also means that there is a focused interest in understanding the space between the two points of "getting it" and "not getting it" as a means for locating and understanding student misconceptions and strengths in solving a task.

In this report we will argue that learning progressions are a valuable framework for characterizing and elucidating the objectives, targets, or goals of instruction. The terms objectives, targets and goals are all convenient synonyms educators use—often interchangeably—to help organize the classroom activities and assignments. In this sense, there is nothing new about the concept of a student learning objective; these have existed since the first teacher-student relationship was formed. The only thing novel about a (capitalized) "Student Learning Objective" is its formalization and standardization as a process that is intended to meet two different needs: (1) to help teachers monitor student growth for formative purposes, and (2) to help the public monitor this growth for accountability purposes. There is a tension between these two needs, as the high-stakes nature of accountability consequences has the potential to undermine the use of SLOs for teaching and learning (Campbell, 1976). Our theory of action is that if an LPF has been put in place first using the collaborative processes we describe in what follows, any subsequent SLO that is generated from this framework will have a greater chance of securing teacher buy-in as something that is authentic to what they value in the classroom and something that they can control.

In the next section we provide specific examples of SLOs used as part of teacher evaluation systems in two states. We do so in order to highlight common threats to the validity of SLOs. We then present the LPF as a possible solution to these validity threats.

II. Student Learning Objectives

Due in large part to federal requirements that mandate evidence of growth in student achievement to be included in educational accountability systems, a total of 30 states who received Race to the Top (RTTT) funding and/or submitted an Elementary and Secondary Education Act waiver now use SLOs as a key component of their educator evaluation systems (Lacireno-Paquet et al., 2014). Originally viewed as a solution by RTTT states and districts for measuring student growth for non-tested subjects because only a minority of teachers teach in subjects for which state-administered standardized tests are available, SLOs in many places now apply to both tested- and non-tested subjects and grades Hall, Gagnon, Marion, Schneider & Thompson 2014). Common elements of the SLO process are:

- Specification of a goal as part of an "objective statement" that defines the content students are expected to learn.
- Specification of the interval of instruction over which the learning is expected to occur (e.g., semester or year-long).
- Identification of assessments to be given to students at the beginning and end of the instructional interval.
- Specification of distinct growth targets for individual students or for groups of students.
- An evaluation of each teacher based on the proportion of the teacher's students who have met their growth targets by the end of the instructional period.

The implementation of these common SLO elements varies across states and districts depending upon the degree of centralization and comparability desired in the set of learning goals used across all teachers, the set of data sources used to support the process, and the methodology used to specify student growth targets and to compute teacher ratings (Lachlan-Hache, Cushing & Bivona, 2012a).

To illustrate key differences found in the implementation of common features to the SLO process, we turn to examples from two states, Georgia and Rhode Island. We consider these two states because they used contrasting approaches and policies in designing the SLO process. In Georgia, school districts, also known as local education agencies (LEAs), dictate which assessments should be used by teachers and how student growth targets should be set. The state provides guidelines on how teacher ratings are assigned based on the extent to which growth targets are met by students. In Rhode Island, the state allows teachers to select their own assessments to evaluate students on their learning objective and to set the growth target expectations for their students. Although Rhode Island provides guidance on two approaches for scoring each SLO, school districts have the right to modify either approach or devise something new. The state defines how SLOs are used to form a final teacher rating to fulfill accountability requirements. Although the two approaches differ in terms of the constraints they place upon teacher enactment of SLOs, the SLO process in both states involves all the same elements. After presenting the SLO process for each state separately, we call attention to what we view as problematic aspects of these common elements.

GEORGIA

In Georgia, only teachers in non-tested subject areas are eligible to submit an SLO. For any given SLO, options for learning objective statements, assessments, and growth targets are established by grade and content area by staff and/or working committees formed in each school district. Evidence of student growth is required to come from the administration of a "pre-test" at the beginning of the school year and "post-test" near the end of the year. In most cases the same test is given twice, but this is not a mandated requirement. The state gives school districts considerable discretion in choosing assessments that would be eligible for use as a pre-test and post-test. These may range from commercially developed tests to locally developed performance tasks.

The Georgia Department of Education's SLO guidance manual (2014) provides two publicly available examples of approaches that could be used to set growth targets, one that establishes individualized student targets, and another that establishes rubric-based student targets. In the first approach, each school district picks expected growth targets in every course and grade. This is illustrated in Figure 1 for an SLO in reading comprehension set for a hypothetical classroom of grade 3 students. Here each student's target is defined relative to their "potential growth" which is the difference between 100 and the student's pre-test score expressed in a percentage-correct metric. In this example, a student will have demonstrated expected growth if the post-test score is 35% of potential growth, and high growth if the post-test score is 60% of potential growth. In the second approach, illustrated in Figure 2, a growth target is defined as movement of one performance level along a rubric, with high growth defined as movement across more than one performance level. In principle, a teacher would use the first approach with assessments that lend themselves to a set of multiple-choice or short open-ended test items that can be readily scored as correct or incorrect. Teachers would use the second approach with assessments that consist of performance-based tasks or a portfolio of student work that is scored holistically.

In either of these two approaches, a student's performance from pre- to post-test is subsequently placed into one of three categories: did not meet growth target, met growth target, exceeded growth target. The frequency counts of students in each category are then tabulated for the teacher of record and converted into a percentage out of total. This forms the basis for scoring teachers according to an "SLO Attainment Rubric," an example of which is provided in Figure 3. These rubrics are used to categorize teachers into one of four effectiveness levels on the basis of the percentage of their students who meet or exceed their growth targets. The choice of thresholds that determine teachers' placements in each category is set by the state and remains the same, irrespective of a teacher's school district, grade or subject.

FIGURE 1: SAMPLE SLO USING INDIVIDUALIZED TARGETS

SLO Statement Example:

From August 2014 to April 2015, 100% of third grade reading students will improve their knowledge of vocabulary and comprehension skills as measured by the Mountain County Schools Third Grade Reading SLO Assessment. Students will increase from their pre-assessment scores to these post-assessment scores as follows:

The minimum expectation for individual student growth is based on the formula which requires each student to grow by increasing his/her score by 35% of his/her potential growth.

Pre-Assessment Score + [(100 – Pre-Assessment Score) * Expected Growth] = Target

Example using 40 on a Pre-Assessment:

40 + [(100 - 40) *.35]

40 + [(60) * .35]

40 + [21] = 61

A score of 61 is the expected growth target for the post-assessment.

Students increasing their score by at least 60% of their potential growth would be demonstrating high growth. A score of 76 or above is the high growth target.

Source: Georgia Department of Education, 2014 [see 2014-2015 SLO Manual at http://www.gadoe.org/School-Improvement/Teacher-and-Leader-Effectiveness/Pages/SLO-Resources-and-Tools.aspx]

FIGURE 2: SAMPLE SLO USING RUBRIC-BASED TARGET

Sample SLO for Grade 6 Intermediate Chorus:

From August 2014 to April 2015, 100% of sixth grade Intermediate Chorus students will improve their sight reading and noting music skills as measured by the Mountain County Schools Intermediate Chorus Performance Task. Students will increase from their pre-assessment scores to these post-assessment scores as follows:

Level 1 will increase to Level 2

Level 2 will increase to Level 3

Level 3 will increase to Level 4

Level 4 will maintain.

Students who increase one level above their expected growth targets would be demonstrating high growth.

Students with a pre-assessment score of Level 4 or an expected growth target of Level 4 may complete a developmentally appropriate project or assignment based on the SLO assessment's content.

Source: Georgia Department of Education, 2014 [see 2014-2015 SLO Manual at http://www.gadoe.org/School-Improvement/Teacher-and-Leader-Effectiveness/Pages/SLO-Resources-and-Tools.aspx]

FIGURE 3: SAMPLE "SLO ATTAINMENT RUBRIC" FOR EVALUATING TEACHERS

LEVEL IV *In addition to meeting the requirements for Level III*	**LEVEL III** *Level III is the expected level of performance*	**LEVEL II**	**LEVEL I**
The work of the teacher results in exceptional student growth. ≥90% of students demonstrated expected/ high growth and ≥30% high growth on the SLO.	The work of the teacher results in appropriate student growth. 65-89% of students demonstrated expected/ high growth on the SLO. OR ≥90% of students demonstrated expected/ high growth and <30% high growth on the SLO. OR 65-89% of students demonstrated expected/ high growth and ≥30% high growth on the SLO.	The work of the teacher does not result in appropriate student growth. 50-64% of students demonstrated expected/ high growth on the SLO.	The work of the teacher results in minimal student growth. < 50% of students demonstrated expected/ high growth on the SLO.

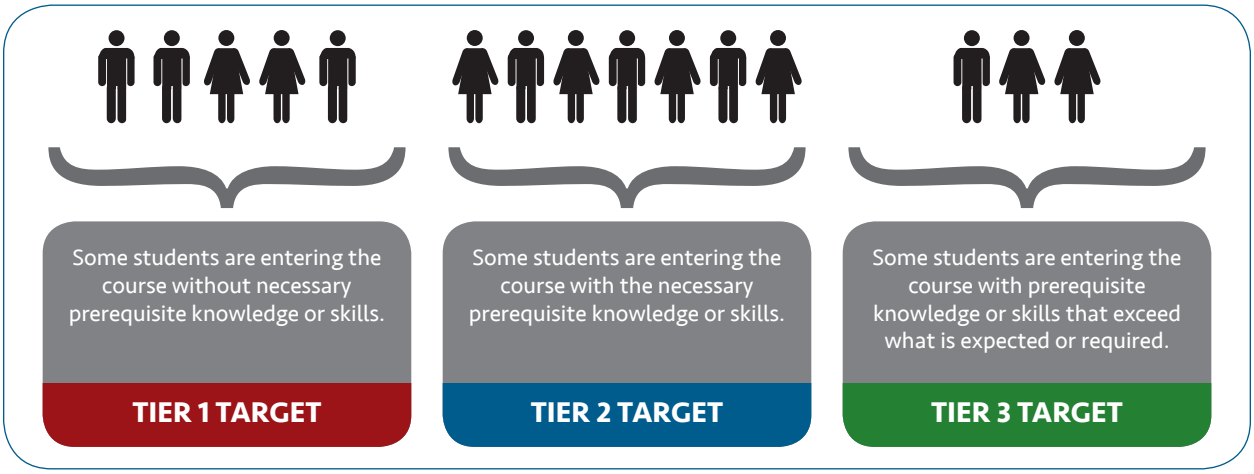
Source: Georgia Department of Education, 2014 [see 2014-2015 SLO Manual at http://www.gadoe.org/School-Improvement/Teacher-and-Leader-Effectiveness/Pages/SLO-Resources-and-Tools.aspx]

RHODE ISLAND

In contrast to Georgia, Rhode Island's SLO system provides considerable latitude to teachers in specifying the learning goal, selecting student assessments, and setting targets for individual or groups of students. Considerable latitude is also given to school districts to define the standards for scoring each SLO. The state encourages all teachers in both non-tested and tested content areas to develop learning goals that reflect critical knowledge and skills that are needed to be successful in the next grade level. Collaboration among teachers is an expected aspect of the SLO process in Rhode Island. The state encourages teachers to work in teams to select important learning goals and select or develop tasks to assess students both over the course of the academic year and at its culmination.

Teachers are expected to consult a variety of data sources that can vary from teacher-made assessments to commercially developed products to first make a "baseline" assessment of how students perform at the beginning of the course and to later make an end of the course assessment of their students. Teachers typically use this baseline information to set targets for students in one of three ways. The first approach is to set a singular target for the whole group; the second is to assign students to one or more distinct baseline groups and create targets specific to each group; and the third is to create targets that are individualized to each student. Figure 4 presents an illustration of how different baseline groups may be defined under the second approach, and Figure 5 presents an example of each approach.

FIGURE 4: AN EXAMPLE OF DEFINING BASELINE GROUPS IN RHODE ISLAND



Source: Rhode Island Department of Education, 2014 [see Measures of Student Learning - Teacher at http://www.ride.ri.gov/TeachersAdministrators/EducatorEvaluation/GuidebooksForms.aspx]

In the above example there are 3 baseline groups set by the teacher and defined according to the descriptors provided. For each of these groups, different targets would be set by the individual teacher for each group in consultation with an administrator and, ideally, in collaboration with other teachers. Column 2 in Figure 5 presents an example of tiered targets set for different baseline groups. In the case where all students perform similarly, only one target is set for one group of students (see column 1 of Figure 5). In the third target setting approach, individual targets are assigned to individual students on SLO assessments selected. Column 3 in Figure 5 presents one example of how individual targets are set for each student using just one data source.

FIGURE 5: SAMPLE "SLO ATTAINMENT RUBRIC" FOR EVALUATING TEACHERS

WHOLE GROUP TARGETS	TIERED TARGETS	INDIVIDUAL TARGETS
One target for all students This works best when: • Baseline data show that all students perform similarly • The course content requires a certain level of mastery from all students in order to pass/ advance (e.g., a C&T course in Plumbing) • It is necessary for all students to work together (e.g., orchestra, theater, dance).	*Different targets for groups of similar students* This allows for projecting achievement for students who are at, above, or below grade level.	*Individualized targets for each student* This can work well in Special Education settings and/or when class sizes are small.
EXAMPLES: 100% of students will pass the State Cosmetology Exam.	**EXAMPLES:** The 18 students who scored a 2 on the baseline writing prompt will score a 4 or higher on the final monthly writing prompts. The 6 students who scored a 3 on the baseline writing prompt will score a 4 or higher on the final monthly writing prompt.	**EXAMPLES:** Students will meet individual targets on Fountas & Pinell guided reading levels. Student 1 will reach a level O Student 2 will reach a level N Student 3 will reach a level M Student 4 will reach a level K Student 5 will reach a level N Student 6 will reach a level L

Source: Table adapted from Rhode Island Department of Education, 2014 [see Measures of Student Learning - Teacher at http://www.ride.ri.gov/TeachersAdministrators/EducatorEvaluation/GuidebooksForms.aspx]

Similar to Georgia, in Rhode Island students are placed into three categories as a function of whether they have not met, met, or exceeded their growth targets. Teachers are then placed into four categories that hinge upon the proportions of students in each category. It is left to the discretion of school districts to establish the relevant thresholds (i.e., lowest category if less than 70% of students meet target, second category if between 70 and 90% meet target, third category if at least 90% of students meet target, and top category if at least 90% meet target and 25% exceed the target).

SUMMARY

The SLO processes in Georgia and Rhode Island are similar in the sense that they involve the same elements but differ in the degree to which teachers are involved in choosing learning goals, assessments, and targets for student performance. Georgia is an example of a "top-down" SLO process in which decisions about learning goals, assessments, and growth targets are all made by school districts and the state. Rhode Island is an example of a "bottom-up" SLO process in which teachers are expected to collaborate in the writing of learning goals, choosing and/or developing assessment tasks, and setting growth targets. In both states, thresholds for scoring teachers on the basis of the proportion of their students who meet or exceed SLO targets are established by LEAs.

Regardless of the degree of centralized control exercised upon the SLO process, states and districts face common challenges that are potential threats to the validity of the SLO process in the long-term. Below, we elaborate on three specific areas that we see as potential threats to the current SLO processes implemented in different RTTT states.

COMMON THREATS TO THE VALIDITY OF SLOS

1. Murky definitions of "growth"

Because SLOs are supposed to represent growth achieved by students between the beginning and the end of an instructional period of interest, all states have devised various approaches to quantify and to classify the level of growth achieved by each student or groups of students. However, for many places using a pre-test and post-test approach to compare baseline and end-of-course location of students, including Georgia and Rhode Island as described above, the answer to the question of "how much" students have learned remains unclear (see Chapman, 2014 and Marion, DePascale, Domaleski, Gong, & Diaz-Bilello, 2012). In cases where different tests are used, the practice of subtracting pre- from post-test scores will lead to questionable results when the two tests differ in the content represented and/or the difficulty of that content (Marion, et al., 2012, Diaz-Bilello, 2011). In cases where the same test is used twice, the pre-test may provide no useful information if the content of the course is sufficiently novel (Buckley, 2015) and also can create a strong incentive to narrow the curricular focus to the material that will be tested. Perhaps most importantly, rules for establishing the adequacy of growth from pre- to post-test are typically arbitrary (e.g., Georgia's formula given in Figure 1), since a focus on proportional increases in gain scores may or may not actually reflect qualitatively rich differences in learning. Ultimately, we see relatively few examples in which qualitatively rich explanations for what it means to show a year's worth of growth have been thoughtfully established.

2. Different targets are set for different students

In many states—including Georgia and Rhode Island but also New York, Hawaii, Ohio, and others—different targets and expectations are set for students located at varying baseline starting points. Although the intention behind setting these targets is to ensure that realistic expectations for students with varying degrees of preparedness are set, several challenges have emerged with this approach. In some places (e.g., Hawaii, Rhode Island, North Carolina and Ohio), teachers are asked to estimate how many students will reach expected performance levels in the future. That is, teachers are asked to define expected targets for their students at the beginning of the school year or course, and the percentage of students meeting the expected targets is used to derive teacher ratings for the SLO. Although this approach was developed with the intention of deferring to a teacher's professional judgment to determine where students should be located on the learning objective, it is premised on the assumption that teachers can accurately predict each student's performance. It also sets up a potentially perverse incentive for teachers to lower standards for lower achieving students. The resulting outcome of this approach is that although these targets are supposed to be informed by baseline data, the expectations set at the beginning of the year can take on a more "arbitrary"

feel with some teachers feeling resentful of those who set consistently low targets for students and others feeling frustrated at never reaching the predicted targets set for their students (Briggs, Diaz-Bilello, Maul, Turner & Biblios, 2014).

Moreover, setting differentiated expectations may send a troubling message to the field that lower expectations are an acceptable standard for lower performing students. For example in Figure 2, communicating the expectation that all students in Level 1 only need to move to Level 2 appears to directly contradict the efforts of civil rights groups that want to ensure that all students, particularly historically underserved populations, are given opportunities to learn and access higher standards. Although we agree that reasonable standards need to be set for students with different levels of preparedness to take on the learning goal or objective, labeling movements short of the objective as "met" may not set the right expectations or signal for these students from an equity standpoint.

3. Accountability concerns trump instructional improvement

Fundamental to any theory of action behind the use of high-stakes accountability is the belief that systemic improvement will come from changes to instructional practice, and much of the rhetoric describing the SLO process places considerable emphasis on this for formative purpose. For example, at the Georgia Department of Education website[1], SLOs are characterized as both a tool to increase student learning and provide evidence that will factor into its educator evaluation system ("The primary purpose of SLOs is to improve student learning at the classroom level. An equally important purpose of SLOs is to provide evidence of each teacher's instructional impact on student learning."). Similarly, in Rhode Island, the Department of Education has produced a video that emphasizes the ambition that SLOs should be a useful tool that stimulates classroom practice and teacher collaboration[2].

However, in many states, including Georgia and Rhode Island, "pre" and "post" assessments become the sole focus of the SLO process. This encourages teachers to set narrow goals that can be easily assessed and to attend only to post-assessment results. Ultimately, the SLO process becomes a compliance activity in which teachers are motivated by accountability concerns, rather than the SLO process being a driver of instructional change (Proctor, Walters, Reichardt, Goldhaber, & Walch, 2011).

Given these three threats to the current state of SLOs, what can be done? Below we introduce a novel approach with the potential to address these threats that we call the Learning Progressions Framework (LPF). The LPF we introduce in the following section was developed as part of the Learning Progressions Project. This pilot project involved a two-year collaboration with three Denver Public Schools (one elementary, one middle, one high) and roughly 40 teachers in the content areas of math and the visual arts. We chose these subject areas and grade levels purposefully so that we could get a better sense of the different issues that emerge when teachers are being asked to implement SLOs across grades within a common subject and within grades across different subjects. The LPF we present next is what has emerged from this pilot project. In this report we focus on the LPF itself and less on the qualitative findings from our implementation of the LPF to help teachers implement SLOs at our pilot sites. Findings from the implementation work are captured under a separate report (see Briggs et al., 2015).

1 http://www.gadoe.org/School-Improvement/Teacher-and-Leader-Effectiveness/Pages/Student-Learning-Objectives.aspx

2 Video located at: http://media.ride.ri.gov/PD/Eval/ImplementingSLOsRI.mp4

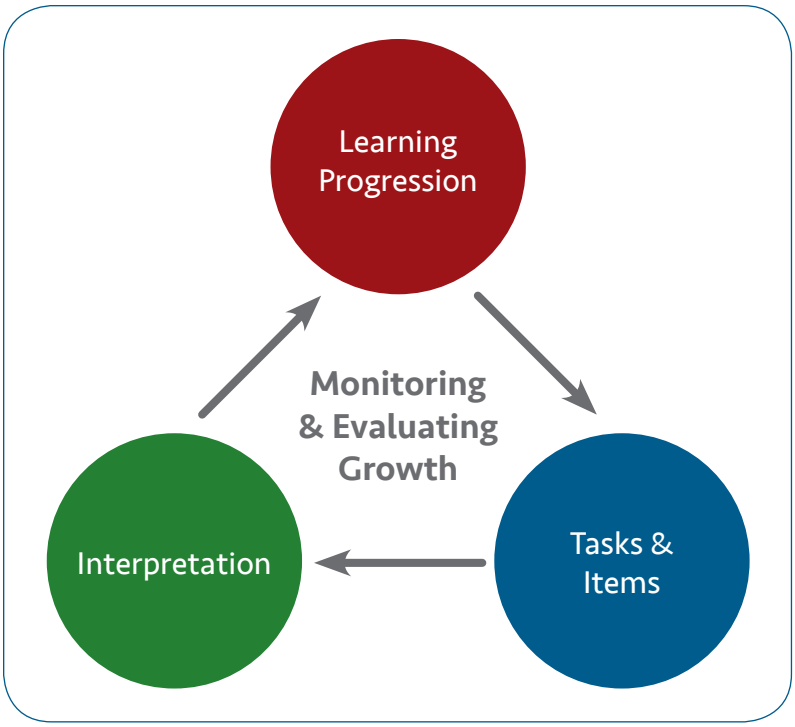
III. The Learning Progression Framework

THE ASSESSMENT TRIANGLE

The LPF was directly inspired by the idea of the "assessment triangle" presented in the National Research Council report "Knowing What Students Know" (Pellegrino, Chudowsky & Glaser, 2001). In this report, an argument was advanced that any high-quality assessment system will always have, at least implicitly, three elements: a theory of student cognition or of how students learn (e.g., a qualitatively rich conception of how students develop knowledge), a method for collecting evidence about what students know and can do, and a method for making inferences from this evidence (e.g., a sample of student answers to assessment tasks) to that which is desired (the ability to generalize knowledge and skills in applied settings). In our pilot project in Denver Public Schools, we worked with teachers to develop learning progressions (LPs) that served as a starting point for building a theory of student cognition. With this in mind we see the LPs as central to our approach for assessing student growth.

Figure 6 presents a visual representation of the LPF. In the simplest sense, an LP (see top of Figure 6) represents a continuum along which a student is expected to show progress over some designated period of time. The progression is defined with respect to some combination of knowledge and skills a student is expected to master within a given content domain by taking a particular course (or courses) in school. In the science and math education communities, a learning progression (or learning trajectory) has been defined as empirically grounded and testable hypotheses about how students' understanding of core concepts within a subject domain grows and become more sophisticated over time with appropriate instruction (c.f., Corcoran, Mosher, & Rogat, 2009). As applied to any general academic context, an LP must be operationalized in terms of a "big-picture" learning target (or targets) identified by teachers at the beginning of the academic year/semester. These targets can be identified with respect to state or district grade-level content standards. Students are expected to start at one position on the LP and then, as they are exposed to instruction, move to a different position. Items, tasks or activities (bottom right of Figure 6) must be purposefully designed to elicit information about what a student appears to know and be able to do with respect to a given LP of interest. Finally, the information elicited from items, tasks and activities has to be converted into a numeric score that can be used to make a judgment about a student's growth and final position on the learning progression (bottom left of Figure 6). When these three elements are taken together, they form a principled and systematic basis for monitoring and evaluating student growth.

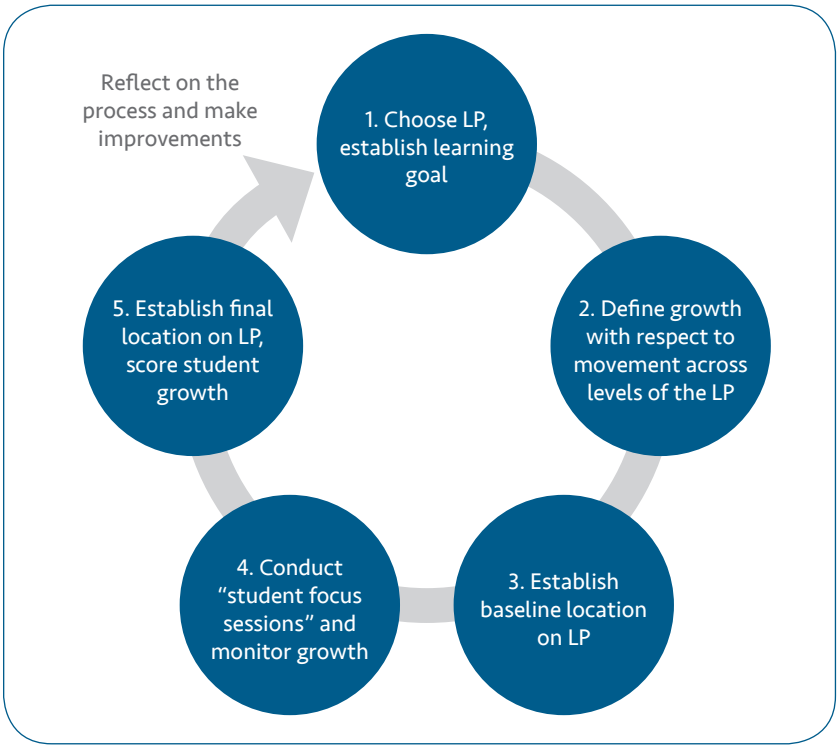
FIGURE 6: A LEARNING PROGRESSION FRAMEWORK



OVERVIEW OF THE LEARNING PROGRESSION FRAMEWORK FOR ASSESSING STUDENT GROWTH

As shown in Figure 7, implementing the LPF is a cyclical process We illustrate these steps in what follows by drawing upon aspects of the LPF that were implemented in our work with three pilot schools in the domain of mathematics.

FIGURE 7: IMPLEMENTING A LEARNING PROGRESSION FRAMEWORK (LPF)



STEP 1: CHOOSE AN LP AND ESTABLISH A LEARNING OBJECTIVE RELATIVE TO THIS LP.

The identification of an LP and the learning objective it supports starts with a deceptively simple question: "What are the most important things you expect your students to know and be able to do by the end of the school year?" This usually results in a wide variety of answers that vary in grain size. The choice, whether it is made by a single teacher or a team of teachers and/or curriculum specialists, should be motivated by the following questions:

- What goals do you have for your students?
- How are these goals related to your state or district's content standards and the scope and sequence of your curriculum?
- How do these goals relate to what students were learning in previous grades/courses? How do they relate to what students will learn in future grades/courses?
- Which of these objectives can be realistically evaluated within an instructional interval defined by a semester or a school year?
- What do success criteria look like for students on these learning objectives (i.e., how will you know it when you see it?)

Importantly, a learning objective should be written such that it is more than a list of facts that students should know. Instead, a learning objective should specify a blending of both knowledge and skills—not just what students should know but what they should be able to do. As a concrete example, the Common Core State Standards for mathematics (www.corestandards.org/Math/Content/) are written with respect to standards, clusters and domains. A learning objective specified at the grain size of a single standard (e.g., "Students will be able to round whole numbers to the nearest 10 or 100") would be too small; an objective specified at the grain size of a domain (e.g., "number sense") would probably be too big. An optimal grain size in this context would be mathematics content specified in terms of a cluster of standards that is also crossed with a subset of mathematical practices (e.g., "Students will be able to use place value understanding and properties of operations to solve problems that involve multi-digit arithmetic.").

The Ideal Approach: An Across-Grade Learning Progression

The ideal is to have teachers working collaboratively in teams to establish an LP that tracks the same big picture concept over *multiple* grades. For some concepts in math, science and English Language Arts, preexisting LPs can be found in research literature (Anderson et al., 2012; Daro, Mosher, & Cochran, 2011; for a partial list of published learning progressions in math, see Appendix A in Daro et al., 2011; and in science see Liu, Rogat, Bertling, 2013, Rogat, et al., 2011, Masters & Forster, 1996). However, even when a research-based LP is adopted, it will still need to be customized to a district's local context. For example, we worked with middle school mathematics teachers to adapt a published LP on equi-partitioning and proportional reasoning in mathematics (Confrey, 2012) so that it would better align with Denver Public Schools' grade-level expectations as defined by district curriculum guides and state standards.

When there is not a pre-existing research-based LP (as will commonly be the case), a new one would need to be created. These created LPs will not match the research rigor or depth of research-based LPs. However, they are still valuable because they are rooted in state standards and teachers' experience, and as they are put in use, they are likely to become more refined and detailed over time. Following from our previous example, the CCSS for mathematics provides an organization of standards across grades in domains such as Operations & Algebraic Thinking (grades K-5), Geometry (grades K-8), and Ratios & Proportional Relationships (grades 6-7). A cluster of standards within any one of these domains, crossed with standards for mathematical practice (i.e., modeling, reasoning) could be chosen as the starting point for elementary or middle school learning progressions across grades, in which grade-specific learning objectives would correspond to substantive statements about what students should know and be able to do by the end of an instructional period. Of course, standards documents for mathematics lend themselves most readily to the formation of an LP that crosses grades because the content tends to have a hierarchical structure. In the following section we discuss how the LPF can still be applied in a context in which progressions of content across grades are more tenuous.

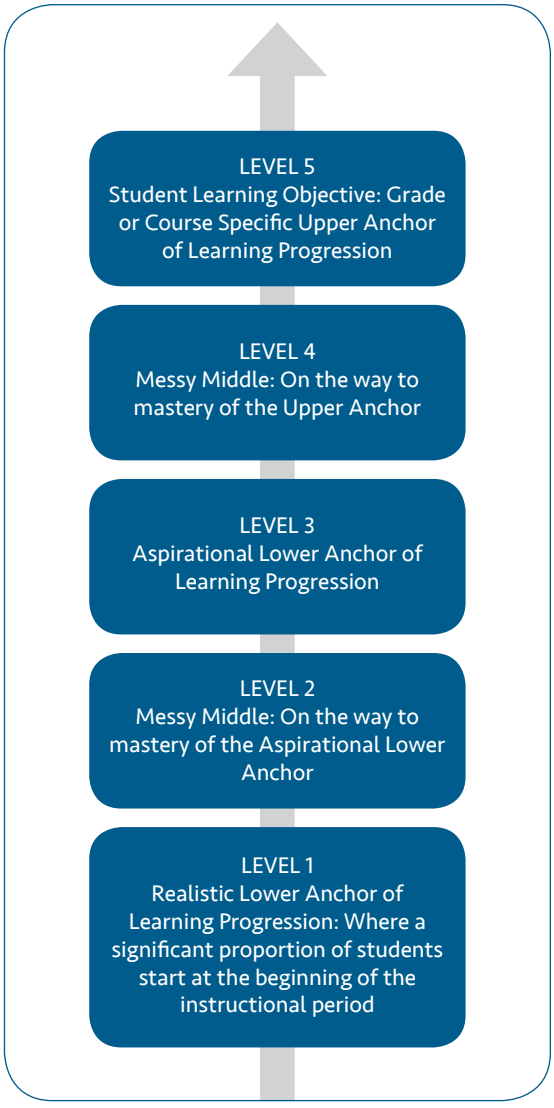
A full example of an across grade LP is depicted in Figure 8. For this LP, grade K-5 elementary school teachers at one of our pilot sites chose the big-picture topic of place value (within the domain of number & operations in base ten in the CCSS) as a focal area that represented a critical foundation students would need in order to learn math at a higher level upon entry to middle school.

FIGURE 8. AN ACROSS-GRADE LEARNING PROGRESSION FOR PLACE VALUE

In Figure 8, each grade level descriptor could be the basis for a different teacher's end of instructional period learning objective. Notice that in Figure 8, the levels 5 and 7, which fall in between the grade K and 1 targets and grade 1 and 2 targets, are at present undefined. These levels represent students that have not yet fully mastered the place value concepts in the upper grade target, but their thinking about place value is clearly more sophisticated than the lower grade target. In the LP science literature, Gotwals (2012) refers to this in-between state as "the messy middle" because it represents a stage in which students may vary considerably with regard to the ideas central to the LP. For example, in the context of place value a student in first grade might be able to correctly compare two digit numbers using symbols by memorizing certain rules without being able to decompose the numbers into combinations of tens and ones. This attention to subtle differences in how students are thinking and reasoning is a hallmark of an LP approach and offers the potential of providing teachers with a useful basis for providing students with feedback and individualized instruction. When a learning progression has been initially specified by stitching together grade-level standards as was done in this example, the messy middle that is most critical in order to characterize course-specific trajectories will usually be undefined. However, if they are given time and support to work in teams, teachers can begin to fill this in as they become attuned to patterns in how students develop their understandings. In our model, this is done through "student focus sessions," which are elaborated in Step 4.

When an LP includes levels that span more than two grades, attention is focused on the changes in student understandings that would be expected to take place across grade levels as students are exposed to instruction targeted to certain core concepts. For a given teacher, it is still a priority to decide how to properly characterize the student growth expected within their grade or course. But for groups of teachers in collaboration, more can be understood about student growth across grades. So we draw an important distinction between an across-grade LP, and a course-specific trajectory. In some instances, if the LP itself is written to capture growth within a single course, the LP and course-specific trajectory will be the same. But usually, teachers will choose some subset of an across-grade LP as the course-specific trajectory. Every course-specific trajectory makes it possible to characterize, in a criterion-referenced manner, where it is that students are expected to start at the beginning of an instructional period, where they are expected to end at the culmination of the instructional period, and how much they have grown in between.

FIGURE 9: A GENERIC COURSE-SPECIFIC LP



An LP Specific to a Single Grade or Course

We now attempt to generalize the LPF sketched out to this point such that it could be employed by any teacher for any subject area, with or without an LP that spans more than two grades. To do this we need to introduce the following terms: the aspirational upper anchor (i.e., the learning target), the aspirational lower anchor, and the realistic lower anchor. The aspirational upper anchor for a student is defined by the content standards that motivate the learning objective. Hence, so long as there are standards or goals that a school, district or state has established for each subject and grade, there should always be a basis for operationalizing a meaningful aspirational target. The aspirational lower anchor represents the set of knowledge and skills that defines the anticipated level of a student's preparedness for the course. Often this can be defined by the anticipated academic preparation in prior grades (or pre-school in the case of Kindergarten). A realistic lower anchor might represent a student who was a full year behind in their preparation. This can be defined by taking stock of the knowledge and skills evident of students who arrive in a course with the least amount of preparation.

The levels of this generic LP are illustrated in Figure 9. If a teacher or team of content specialists can identify a learning objective, an aspirational lower anchor, and a realistic lower anchor, it is possible

to create an LP to conceptualize student growth. Note that the levels of the LP could be tailored to topics in any specific course and the characteristics of students enrolled in it. If, for example it was not possible to articulate two levels below the aspirational lower anchor, the LP could be reduced to four levels rather than five with the lowest representing any students that come into a course underprepared. Or, if students tend to come into the course with a background that exceeds the aspirational lower anchor, it would be possible to add a level above the aspirational upper anchor. The aspirational target for all students—the upper anchor the LP—would remain the same, but by avoiding a ceiling effect the amount of growth it would be possible for any given student to demonstrate would change.

The critical features of this step are threefold. First, it requires a criterion-referenced explication of what it means for students to demonstrate growth during an instructional period. Second, it establishes a common target for all students in the course: the aspirational upper anchor of the LP. Instead of differentiating students in terms of their target, students are differentiated in terms of their preparedness and the level of additional instructional support that is likely required to catch them up. Third, it places attention on the fact that there are levels of student understanding and reasoning that fall in between the LP anchors ("the messy middles").

In the remaining steps of the LPF that follow, we will assume that a teacher has established a course or grade-specific LP as depicted in Figure 9. Note that even for teachers working collaboratively in vertical teams who may have developed an across-grade LP with 15 levels as depicted in Figure 8, it is quite likely that student growth would only encompass a subset of these levels as in Figure 9. Hence whether or not an across-grade LP is available, the steps of the LPF that follow would remain the same.

STEP 2. DEFINE GROWTH WITH RESPECT TO MOVEMENT ACROSS LEVELS OF THE LP.

Growth is defined as the movement of students across levels from the beginning of the instructional period to the end, and can be roughly quantified by subtracting a student's starting level from the student's ending level. In order to determine a meaningful reference unit, we define "one year's growth" based on movement from level 3 to level 5 in Figure 9. Recall from Figure 9 that these levels have a well-defined criterion-referenced meaning as the beginning and ending locations of students who follow the path envisioned in standards documents. Because this is a two-level move, we define "one year's growth" as movement of two levels anywhere within the trajectory.

Any amount of growth can be defined relative to this unit as follows: *negative or no growth* is defined as a student who remains at the same level or moves backward, *minimal growth* is defined as movement of only one level, *aspirational growth* is defined as movement of two levels (i.e., "one year's growth"), and *exceeded aspirational growth* is defined as movement of more than two levels. The value of this approach to measuring growth is that growth can be captured anywhere along the trajectory, so that a student who moves from level 1 to level 3 is conceptualized as demonstrating a full year's growth, even though, from a status perspective, the student is not "at grade level."

STEP 3: USE THE LP, ALONG WITH AN ASSESSMENT CHECK TOOL, TO IDENTIFY PREEXISTING AND/OR CREATE NEW ASSESSMENT DATA TO ESTABLISH THE LOCATION OF STUDENTS WITHIN LP LEVELS AT THE BEGINNING OF THE INSTRUCTIONAL PERIOD.

Choosing or Writing Assessment Items

The next step is to either pick or develop items (i.e., tasks, activities) that can be administered over the course of the instructional period with an eye toward eliciting information about what students appear to know and be able to do with respect to the big picture idea captured by the LP. A key consideration in relating student performance on these assessments to LP levels is the quality of the items.

- Are the items well-aligned to the LP?
- Do they allow for varied means for students to express understanding?
- Are they written at the appropriate level of cognitive demand?
- Are the rubrics written so as to minimize inter-rater variance?

One challenge when finding or developing items for the purpose of determining a student's location on the underlying LP is that no single item, unless it is a rather involved performance task, is likely to provide information about all possible levels of an LP. As an example relevant to the place value LP from Figure 8, consider a student who is given a task in which he/she is asked to compare and explain the difference between multiple three digit numbers. Such information might help to establish whether a student is at level 10 relative to level 9 but not whether the student is at level 8 or even 7. In other words, when taking an LP approach, assessment items need to be written purposefully so that, collectively, they target multiple levels of the LP. Even within a level, multiple items would need to be written that give students more than one opportunity to demonstrate their mastery of the underlying concept. For many LPs, this will mean items need to be written with an eye toward not just answering an item correctly but also the process used to answer the item correctly. To go back to the place value example, a student may be able to identify which of two three-digit numbers is larger without fully understanding how to compose and decompose a three-digit numbers into hundreds, tens, and ones. The need for a "bank" of items that can distinguish students at multiple LP levels points to another advantage of working in multi-grade vertical teams. Namely, if teachers at each grade level are able to focus on finding or writing items that would distinguish mastery of the aspirational learning target from the level below, an assessment could be assembled relatively easily by pulling from the full bank of items across grades.

The above discussion has mainly focused on assessment *items*. When multiple items are assembled into an *assessment*, one can ask questions about the quality of the overall assessment.

- Does the set of items cover all of the relevant levels in the LP?
- What is their cognitive complexity?
- Are there multiple items per level?
- Is the assessment fair and unbiased?
- Do student scores generalize over other types of parallel items that could have been administered, or other teachers that could have done the scoring (in the context of constructed response items)?

To help teachers evaluate items and assessments vis-à-vis the considerations above, we collaborated with teachers to develop rubrics for evaluating items and assessments. These rubrics, which we call "assessment check tools" are available on the CADRE website (http://www.colorado.edu/education/cadre).

Establishing Baseline Locations of Students on the LP

It is common for teachers to convert student performance on an assessment into grades or into "proficiency bands" by summing up the total points earned, expressing these points as a percent of total, and then assigning demarcations at 90, 80, 70, etc. to distinguish an A from a B from a C, or "Advanced" from "Proficient" to "Partially Proficient." These sorts of demarcations based on percentage of points earned are not necessarily meaningful in the context of the levels that have been defined for an LP. If the assessment tasks happened to be very easy (or very hard), then students might all receive A and Bs (or Fs and Ds) even if they were located somewhere in the middle of the LP. Without closer inspection, one may conflate the ability of the students with the difficulty of the assessment. This is why many teachers "curve" the scores from their exams.

A very carefully considered "mapping" needs to occur any time an assessment is administered for the purpose of estimating a student's level on a LP. The mapping must convert student scores or performance on assessment items or performance task(s) into a location on the LP. This process is important because it removes much of the arbitrariness associated with student scores and instead provides each score with a meaning that is directly aligned to the descriptions present in the LP. In addition, aligning each assessment to the LP eliminates the need to give a common pre-post assessment. Instead, assessments can be designed to provide maximum information about a students' presumed location (for example, relative to Figure 9 this would be levels 1-3 for baseline and levels 4-5 for final).

The key point is that absent of formal mapping process, differences in raw test scores do not necessarily communicate the same thing as differences in LP levels. In some cases, such as the administration of a performance task, the conversion might be one to one—that is, the performance task might be scored

according to levels of the LP. But in other cases, when an assessment consists of many tasks and/or items, scores from 1 to 20, or 10 to 100, etc., an approach needs to be developed to convert these scores into LP levels.

One practical way to accomplish this is to follow the two step process below (for a more detailed example, see pp. 50-51 of Chapter 6 in the full report by Briggs et al, 2015):

1. Start by dividing student scores into bins, such that there is the same number of bins as there are levels in the LP (e.g., if the LP has five levels, divide student scores in quintiles; if an LP has four levels, divide student scores in quartiles, etc.). This serves as a basis for grouping students who are likely to be relatively similar in their overall performance, and serves as a tentative mapping from the assessment scores to the LP. The top score group maps to the top LP level, the lowest score group maps to the lowest LP level, etc.

2. The next step is to examine the kinds of task and item responses (i.e., answers to test questions) from students in each score group to determine if the responses are in fact representative of the mapped LP level. For the students in the top performing group, one question to be asked is whether the lowest score in this bin is indicative of someone at the top level of the LP. And so on for each bin. If, for example, all students in a teacher's class have high scores, it would be possible that students in the top three score bins all get mapped to the top level of the LP. In contrast, if all students have low scores, the top three score bins might only be mapped to middle level of the LP.

There are two occasions when the mapping process described above will need to be used to place students into LP levels. At the outset of the instructional period, in order to establish each student's baseline location, and at the end, to establish the amount of growth they have demonstrated relative to this baseline location. This is necessary in order to compute, for each student, a growth score in terms of the difference in LP levels from baseline to the end of instructional period. In each case, assessments should serve as only one piece of a body of evidence that a teacher uses to place students. A practical approach for this is for teachers to consider other evidence first (e.g., a student's score on the previous year's state test for baseline or the teacher's observations of the student throughout the school year for final), and use this evidence to tentatively place students into levels. Then, the student's performance on the baseline or final assessment can be used to *refine* this tentative location within one level. Larger adjustments would require significant evidence from the assessment. For example, the teacher's observations of a student throughout the year suggests that the student's final location is level 4, then the final assessment might be used to refine this location into levels 3, 4, or 5. Substantial evidence would be needed from the final assessment to place the student into level 1 or 2. It should not require multiple sources of evidence to show a student begins the year at the aspirational lower anchor. To ease the setting of benchmarks, it should be assumed students are at the aspirational lower anchor and evidence should be considered to judge whether that assumption is sound.

STEP 4: MONITOR GROWTH AND ADJUST INSTRUCTION BY CONDUCTING "STUDENT FOCUS SESSIONS" OVER THE COURSE OF THE INSTRUCTIONAL PERIOD.

Student responses to assessment items can be used to not only locate students on a learning progression, but to facilitate changes to instruction. To accomplish this, we developed a process in which teachers meet together to discuss student responses with the goals of (a) improving assessment items, and (b) understanding student reasoning and using this understanding to design responsive classroom activities. This process takes place during a "student focus session." We have developed a guidebook for student focus sessions, which is available on the CADRE website (http://www.colorado.edu/education/cadre). Below, we briefly summarize the process.

Student focus sessions are oriented around a small number of strategically-chosen student responses to assessment tasks that have been written to align to the LP. They have two phases, which reflect the two goals discussed above. In the first phase, all participants score the same subset of student work. They then discuss any variation in their scores, and they come to a consensus score. They discuss ideas to modify the task and rubric to minimize score discrepancies in the future. In phase two, participants examine the consensus scores and the student work to generate a better sense for the strengths and weaknesses in individual students as well as groups of students. They then discuss next steps for this student based on their analysis of the student's reasoning, and their goals for the student. Even though the focus is on a single student, the students' response is likely representative of a set of students in the class. Thus, by deeply understanding this single student and designing responsive classroom activities, the teachers are really understanding and designing for a group of students.

Our initial analysis of student focus session suggests that they are productive spaces for improving tasks and for understanding student reasoning. We are currently exploring two further conjectures related to student focus sessions. First, we conjecture that student focus sessions will, over time, help teachers make principled changes to their LPs, including fleshing out the "messy middles." Second, we conjecture that using student focus sessions as part of a LP-based SLO process improves the chance that the process will be perceived as more than a compliance-based activity.

Note that although student focus sessions are presented here as a discrete step, they would actually continue throughout the year, as teachers create and administer new assessment items to monitor growth.

STEP 5: MAKE CONNECTIONS FROM CULMINATING ASSESSMENT AND BODY OF EVIDENCE TO THE LOCATION OF STUDENTS AT END OF INSTRUCTIONAL PERIOD. SCORE STUDENT GROWTH AND REFLECT UPON THE SLO.

Using a similar process as that described in Step 3 above, teachers administer an assessment at the end of the instructional period and use this evidence to establish a final position of students on the LP. Student growth is then calculated by subtracting the baseline LP level from the final LP level. Inferences about growth at an aggregate level can then be made by examining the distribution of growth scores and calculating summary statistics.

At this stage teachers using the same LP should reflect upon the process and consider ways that it can be improved in the following year.

- Can the levels of the LP be fleshed out in greater detail?
- Are their insights to be shared about what student reasoning looks like in the previously undefined messy middle?
- In retrospect, were some assessment items used to monitor student performance as part of student focus sessions better than others?
- To what extent do teachers now have a bank of assessment tasks that could be used for formative purposes in the next year?
- How could instructional activities connected to the LP be changed for the better in the next year?
- What lessons learned about students in the context of this LP might generalize to other topics for instruction?

It is important to appreciate that the most time-consuming aspect of using the LPF as a basis for SLOs is the initial fixed investment in setting up the infrastructure. But once this has been done, the focus should shift to making gradual improvements. Finally, it is in the stage where teachers are considering the implications for the next year where the value of an across-grade LP is especially evident because information about a student's LP level in the lower grade/course would become that much more relevant to the student's parents and the upper grade/course teacher.

CONCEPTUAL ADVANTAGES OF THE LPF

In the LP-based framework for student growth, teachers set meaningful, standards-based learning objectives for students, create learning progressions based on those objectives, and then evaluate students, monitor progress, and assess growth relative to the learning progression. Key advantages of the framework include:

- A priority is placed on thinking about *growth* across multiple grades, rather than student status or level of mastery.
- The same target is set for all students.
- Because growth can happen anywhere on the LP, teachers are not penalized for having students who enter their class underprepared.
- Clear, criterion-referenced definitions are provided for what it means to have "one year of growth" and what it means to demonstrate minimal growth.

• Teacher collaboration around the discussion of student work (formalized as part of ongoing "student focus sessions") are central to what goes on in between the beginning and end of an instructional period.

These advantages are especially pronounced when compared to the design of the vast majority of SLO systems in place in many districts and states. As we described in Section 2, these implementations often face three threats to validity. Table 1 summarizes how the LPF may address each of these three threats.

TABLE 1. SUMMARY OF THREATS ADDRESSED BY AN LP

THREAT	HOW AN LPF CAN ADDRESS THIS THREAT
1. Murky definition of growth	Clear, criterion-referenced definitions are provided for what it means to have "one year of growth" and what it means to demonstrate minimal growth. Additionally the different levels specified within an LP can help distinguish and clarify the set of knowledge and skills that a student is expected to acquire in order to reach the learning objective.
2. Different targets are set for different students	Different targets are not set for different students. Because growth can happen anywhere on the LP, awarding the amount of growth achieved by each student toward reaching the learning objective is of interest rather than determining whether students meet pre-defined and differentiated targets either set by a teacher or set by a central authority.
3. Accountability concerns trump instructional improvement	Student focus sessions are central to the LP approach. This fosters teacher collaboration around the careful study of student reasoning, and helps teachers improve assessment items, understand student reasoning, and design learning activities based on student reasoning. By engaging in this collaborative process with other teachers, teachers serve as peer-review regulators for this process by checking on the quality of products used to evaluate student learning and by evaluating the evidence presented by teachers to support claims about how much growth was achieved by their students.

As indicated in the above table, although we are optimistic that an LP framework may address these common threats found in SLO systems implemented in many places, we also recognize that this approach would require significant staffing supports and resources. In the next section, we discuss some of the limitations of this approach to illuminate the high level of commitment required to undertake this effort even as a small pilot with a few participating schools.

IV. Limitations

Although our initial pilot work developing and implementing the LPF as a basis for SLOs suggests to us that the approach has great potential, we do not wish to give the impression that it represents a panacea. To begin with, writing SLOs from an LPF would require a greater fixed investment in time and resources to get things off the ground, especially if an attempt was being made to implement SLOs using an across grade LP. Developing an across-grade LP takes significant work and effort to do with a group of teachers. In math, despite being able to continue using the same across grade LPs developed at each school site in year 1, teachers spent a significant amount of time aligning tasks to the LPs and critically reviewing student work to examine how student reasoning was evolving at various time points during the year.

The difficulty of implementing LPs is further compounded in subject areas such as in the electives where no common curriculum may be in place or where a common understanding of standards needs to be established first with teachers participating in the LP development work. In our work with elective teachers, we quickly realized that an across grade LP would prove to be difficult to define without a common curriculum or clear understanding of standards in place, and so we limited the LP development work this year to construct smaller within grade LPs linked to grade-specific standards. However, even developing the narrower set of within grade LPs took a considerable amount of time and effort for elective teachers to develop. For many elective teachers we worked with, this was the first time they were being asked to explicitly define and examine what students with varying levels of sophistication know and can do in their respective content areas.

Moreover, is it unlikely that the processes supporting the LPF, which center on cultivating formative practices to guide assessment and instructional practices, are readily scalable over a short period of time or can be

simply adopted by a school district or state. Instead, this work will require significant professional development and collaborative time among teachers to:

- Begin the work of adapting existing LPs or in many cases with the majority of subject areas, move forward with the work of constructing within grade or across grade LPs.
- Move through an iterative process of developing and refining tasks aligned with the LP.
- Move through an iterative process of defining and adjusting instructional steps.
- Make adjustments to the LPs based on reviews of student work.

Most of the teachers we have worked with appreciated the opportunity to collaborate with their peers and discuss student work relative to an LP. Further, in the case of the elective areas, participating teachers and curriculum coordinators noted that they appreciated getting the opportunity to focus on their own content areas and to better understand the academic expectations and instructional targets to be set for students across grades. But within the context of this district, it is unclear whether this work could be sustained without the extensive facilitation and supports that our research team has provided over the two-year period at participating schools and with the group of teacher leaders and curriculum coordinators in the elective areas. Districts that are considering this approach should be prepared to allocate substantial professional development resources to the effort. We conjecture that the investment will pay off in that the approach is likely to be viewed by participating teachers as an authentic and valuable process apart from its role in teacher evaluation (c.f., Briggs, 2013). Hence even if the mandate for formal teacher evaluations involving evidence of student growth were to shift, SLO approaches based on an LPF would have the potential to remain as an infrastructure and process that supports good pedagogical practice and professional development. But at this point, our conjecture is limited to our experiences working with teachers on the LP project over the past two years.

Another limitation involves psychometric concerns in using the LPF as a basis for measuring student growth. In the general approach we have outlined, student growth is represented by transitions across discrete LP levels. The reliability of this classification depends in large part upon the quality of the underlying assessment tasks. It is quite likely that student-level growth scores would have low levels of reliability, though the problem would be lessened when aggregated to the teacher level. In the long-term, if the approach were to be adopted by all teachers in the same subject/grade combination in a school district, then it might be possible to develop common assessments that afford more reliable student-level growth measures. But in the short-run, the psychometric quality of student-level growth scores would need to be given considerable scrutiny if these were to be factored into teacher evaluations. Of course, the same issue regarding the quality of assessment tasks used applies to the current SLOs that are being implemented. That is, SLOs based on the LPF would surely be no worse than SLOs based on the status quo, and in many cases the LPF likely results in higher-quality assessments. As the statistician John Tukey once wrote "Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise" (Tukey, 1962, p. 13).

Overall, the biggest limitation for implementing an LPF as the basis of the SLO process is the dedicated commitment required for a school and district to have teachers participate in regularly scheduled student focus sessions to review student work relative to the developed LP and to deepen their knowledge of assessment as they engage in the process of aligning and developing high quality tasks to support inferences about student learning. Lachlan-Haché, Bivona, & Cushing (2012b) and the Reform Network (2011) also recognized that significant resources and time are required to sustain the SLO process focused on data inquiry. However, using an LPF as the basis for SLOs would require even more time and resources due to the work required to not just define a learning objective (as required by the SLOs) but to define and embed those objectives as part of a larger constellation of instructional and assessment targets defined either across grades or situated within a grade. If districts or school do not schedule time for teachers to collaborate and review student work in this manner as part of the regular PD schedule (in our pilot work we found that 1-2 hours per week are needed to sustain student focus sessions), then the SLO approach we have sketched out in this paper will be short-lived.

As year 2 of the LP project draws to a close for this 2014-2015 year, we are compiling an evaluation report that summarizes the lessons we are learning from this ongoing work. This report will be released during the summer of 2015. We anticipate that the results from the evaluation that focuses on the larger district's SLO process and the LP project will be used to determine how best to structure supports for the next school year and to define what aspects of the LP project work are scalable and applicable district-wide for the 2015-2016 year.

In addition, the district has begun to implement two aspects of our pilot work. First, the district is organizing groups of teachers to begin mapping out developmental trajectories tied to their learning objectives as a way to evaluate what novice students should know and can do relative to students exhibiting "expert" characteristics. Second, the district is organizing teachers across content areas to begin discussing and examining student work as a way to evaluate and refine the trajectories they've developed to evaluate student learning. Although the district's work in these two areas is at a nascent stage and the student focus sessions have not extended into evaluating the quality of tasks used by teachers, we applaud the district for moving in the direction of integrating these types of activities into their current SLO process. Even if the use of an LPF as an organizing framework for SLOs in the district remains a distant goal, the decision to organize collaborative reviews of student work and to begin developing shared understandings across teachers for what different types of students know and can do will hopefully contribute to the instructional relevance of the SLO process for teachers both in the short- and the long-term.

References

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public highschools. *Journal of Labor Economics*,*25*(1), 95-135.

Anderson, C. W., Cobb, P., Barton, A. C., Confrey, J., Penuel, W. R., & Schauble, L. (2012). *Learning progressions footprint conference: Final report* (Vol. 28). East Lansing, MI.

Briggs, D. C. (2013). Teacher evaluation as Trojan horse: the case for teacher-developed assessments. *Measurement: Interdisciplinary Research and Perspectives*, Vol 11(1-2), 24-29.

Briggs, D., Diaz-Bilello, E., Maul, A., Turner, M., & Bibilos, C. (2014). Denver ProComp Evaluation Report: 2010-2012. Retrieved from: http://www.colorado.edu/education/cadre/publications

Briggs, D., Diaz-Bilello, E., Peck, F., Alzen, J., Chattergoon, R., McClelland, A. (2015). Tier 3 Student Learning Objective Pilot: Documentation of Pilot Work and Lessons Learned in the 2013-2014 School Year. Retrieved from: http://www.colorado.edu/education/cadre/publications

Buckley, K. H. (2015). *An evaluation of the interpretability of Student Learning Objectives results in one state based on student assessments and growth targets*. (Unpublished doctoral dissertation). Harvard University, Cambridge, MA.

Chapman, L.H. (2014). The Marketing of Student Learning Objectives (SLOs): 1999-2014. Unpublished manuscript. Retrieved from: http://vamboozled.com/laura-chapman-slos-continued/

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2013). *Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates* (No. w19423). National Bureau of Economic Research.

Campbell, D. T. (1976). Assessing the Impact of Planned Social Change. Occasional Paper Series, No. 8.

Confrey, J. (2012). Better measurement of higher cognitive processes through learning trajectories and diagnostic assessments in mathematics: The challenge in adolescence. In V. F. Reyna, S. B. Chapman, M. R. Dougherty, & J. Confrey (Eds.), *The adolescent brain: Learning, reasoning, and decision making* (pp. 155–182). Washington DC: American Psychological Association.

Corcoran, T., Mosher, F. A., & Rogat, A. (2009). Learning Progressions in Science: An Evidence-Based Approach to Reform. CPRE Research Report# RR-63. *Consortium for Policy Research in Education*.

Daro, P., Mosher, F. A., & Corcoran, T. (2011). *Learning trajectories in mathematics: A foundation for standards, curriculum, assessment, and instruction*.

Diaz-Bilello, E. (2011). A validity study of interim assessments in an urban school district. Unpublished doctoral dissertation. University of Colorado, Boulder

Doherty, K. M., & Jacobs, S. (2013). State of the states 2013: Connect the dots: Using evaluations of teacher effectiveness to inform policy and practice. *National Council on Teacher Quality. http://www. nctq. org/ dmsView/State_of_the_States_2013_Using_Teacher_Evaluations_NCTQ_Report*.

Georgia Department of Education, Office of School Improvement, Teacher and Leader Keys Effectiveness Division. (2014). *Student learning objectives: Operations manual.* Retrieved from: http://www.gadoe.org/School-Improvement/Teacher-and-Leader-Effectiveness/Pages/SLO-Resources-and-Tools.aspx

Gordon, R. J., Kane, T. J., & Staiger, D. (2006). *Identifying effective teachers using performance on the job*. Washington, DC: Brookings Institution.

Gotwals, A. W. (2012). Learning progressions for multiple purposes. In *Learning progressions in science* (pp. 461-472). SensePublishers.

Hall, E., Gagnon, D., Schneider, M.C., Marion, S., Thompson, J., (2014). State Practices Related to the Use of Student Achievement Measures in the Evaluation of Teachers in Non-Tested Subjects and Grades. Unpublished manuscript. Retrieved from: http://www.nciea.org/publication_PDFs/Gates%20NTGS_Hall%20082614.pdf

Lachlan-Haché, L., Cushing, E., & Bivona, L. (2012a). Student learning objectives as a measure of educator effectiveness: The basics. Washington, DC: American Institutes for Research.

Lachlan-Haché, L., Cushing, E., & Bivona, L. (2012b). Student learning objectives: Core elements for sustainability. Washington, DC: American Institutes for Research.

Lacireno-Paquet, N., Morgan, C., & Mello, D. (2014). How states use student learning objectives in teacher evaluation systems: a review of state websites (REL 2014–013). Washington DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands. Retrieved from http://ies.ed.gov/ncee/edlabs.

Kane, T. J., & Staiger, D. O. (2008). *Estimating teacher impacts on student achievement: An experimental evaluation* (No. w14607). National Bureau of Economic Research.

Liu, L., Rogat, A., Bertling, A. (2013). A CBAL Science Model of Cognition: Developing a Competency Model and Learning Progressions to Support Assessment Development. Educational Testing Service: Princeton, NJ.

Marion, S., DePascale, C., Domaleski, C., Gong, B. & Diaz-Bilello, E. (2012, May). Considerations for analyzing educators contributions to student learning in non-tested subjects and grades with a focus on student learning objectives. Dover, NH: Center for Assessment. Retrieved from http://www.nciea.org/publications-2/

Masters, G., & Forster, M. (1996). Progress Maps: Assessment resource kit. Camberwell, Victoria, AU: Australian Council for Educational Research.

Pellegrino, J.W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington DC: National Academy Press.

Proctor, D., Walters, B., Reichardt, R., Goldhaber, D., & Walch, J. (2011). Making a Difference in Education Reform: ProComp Evaluation Report 2006-2010. The Evaluation Center, University of Colorado Denver: Denver, CO.

Rhode Island Department of Education (2014). *Measures of student learning.* Retrieved from: http://www.ride.ri.gov/TeachersAdministrators/EducatorEvaluation/GuidebooksForms.aspx

Rogat, A., Anderson, C., Foster, J., Goldberg, F., Hicks, J., Kanter, D., Krajcik, J., Lehrer, R., Reiser, B., & Wiser, M. (2011). Developing Learning Progressions in Support of the New Science Standards. Consortium for Policy Research in Education, Teachers College, Columbia University: New York, NY.

The Reform Support Network. (2011). Targeting growth: using student learning objectives as a measure of educator effectiveness. Retrieved from: http://msde.state.md.us/tpe/TargetingGrowth_Using_SLO_MEE.pdf

Tukey, J. W. (1962). The future of data analysis. *The Annals of Mathematical Statistics*, 1-67.