Learning Progressions Project: Documentation of Pilot Work and Lessons Learned in the 2013-2014 School Year

Derek C. Briggs

Elena Diaz-Bilello

Fred Peck

Jessica Alzen

Rajendra Chattergoon

Abby McClelland

Center for Assessment, Design, Research and Evaluation

University of Colorado

National Center for the Improvement of Educational Assessment

October 2014

Table of Contents

Chapter I: Introduction 5
Motivation for a Pilot Project to Improve SLOs 5
Goals and Purpose of the 2013-14 SLO Pilot Project 9
Overview of this Report 11
References 13
Chapter II: A Learning Progression Model for SLOs 14
The Assessment Triangle as a Conceptual Framework for SLO Development 14
Unique Features of the Learning Progression Approach to SLO Development and Scoring 15
Using a Learning Progression to Define an SLO (Cognition Vertex of Assessment Triangle) 16
Choosing or Writing Items for Student Assessment (Observation Vertex of Assessment Triangle) 20
Scoring and Interpreting Student Work (Interpretation Vertex of Assessment Triangle) 21
Rating Teachers on the Basis of Student Growth 25
Chapter III: Pilot Implementation of the Learning Progression SLO Approach 30
Overview 30
Blue Elementary School 33
Spruce Middle School 34
Green Beechway High School 35
Chapter IV: Results by Math Pilot Site: Blue Elementary School 38
Using a Learning Progression to Define an SLO at Blue 38
Course-Specific Trajectories 41
Choosing or Writing Items for Student Assessment at Blue 42
Initial Task Development and Score Moderation 42
Development of Tasks for End of Course Assessment 52
Scoring and Interpreting Student Work 53
Rating Teachers on the Basis of Student Growth 54
Concerns about Placing Students into Retroactive Baseline Levels 57
Teacher Reactions and Feedback 58
Concerns about Learning Progression 58
Scoring Student Growth 59
Opportunities for Collaboration and Vertical Alignment 60

Successes and Challenges for SLO Model Implementation at Blue . Error! Bookmark not defined.
Chapter V: Results by Math Pilot Site - Spruce Middle School 61
Using a Learning Progression to Define an SLO at Spruce 61
Choosing or Writing Items for Student Assessment at Spruce 65
Scoring and Interpreting Student Work at Spruce 70
Rating Teachers on the Basis of Student Growth at Spruce 72
Teacher Reactions and Feedback 76
Chapter VI: Results by Math Pilot Site - Green Beechway High School 80
Using a Learning Progression to Define an SLO at GB 80
Choosing or Writing Items for Student Assessment at GB 85
Scoring and Interpreting Student Work at GB 90
Rating teachers on the basis of student growth at GB 95
Teacher Reactions and Feedback 96
Teachers interest and concerns with LPs 97
Scope and time commitment of the project. 100
Benefits of collaboration 100
How the pilot impacted the teachers' classroom practice 100
Chapter VII: The Visual Arts Pilot 104
Introduction 104
Chapter VIII: The Original Visual Arts Pilot 107
Professional Development Structure 107
Participant Description and Context 109
Teacher Background and Perspectives on SLOs and SGOs 109
Data Teams Context and Professional Development 110
Assessment Practices 111
Developing Visual Arts Learning Progressions for SLOs 113
Process for Developing the Multi-Grade Progressions 113
Process for Developing the Course-Specific Trajectories at GB 116
Refining Selected Tasks and Rubrics 119
Establishing Baseline Performance Data 121
Defining Performance Expectations on Tasks to Determine Location on Progressions 122
Scoring Student Growth 124
Categorizing Teachers on the SLO 125

Reflections on the Original Pilot 128
Challenges 128
Successes 129
Chapter IX: The Teacher Leader Workshop Pilot 132
Professional Development Structure 132
Participant Description and Context 134
Teacher Background 134
Developing the Unit-Specific Learning Progressions 134
Process of Refining Unit Tasks 137
Process of Refining Rubrics 137
Reflections from the Teacher Leader Workshops 138
Challenges 138
Successes 139
Chapter X: Final Reflections on Goals Achieved in Pilots, Recommendations and Next Steps 143
Initial Pilot Goals ..... **Error! Bookmark not defined.**
Recommendations ..... **Error! Bookmark not defined.**
Short-Term Considerations ..... **Error! Bookmark not defined.**
Scale-Up Considerations ..... **Error! Bookmark not defined.**
Next Steps 156
Mathematics 156
Visual Arts 157
Appendix A 159
Appendix B. Assessment statistics at Spruce 160
Appendix C. Assessment statistics at GB 163
Appendix D. The across-grade LP at GB 166
Appendix E. The course-specific trajectories at GB 169
Appendix F. Multi-grade visual arts learning progression 172
Appendix G. Course-specific visual arts learning progressions 177
Appendix H. SLO assessment quality check tool 181
Appendix I. Visual arts tasks and rubrics with feedback from the original pilot 185
Appendix J. Sample rubric from teacher leader pilot 191

Chapter I: Introduction

Motivation for a Pilot Project to Improve SLOs

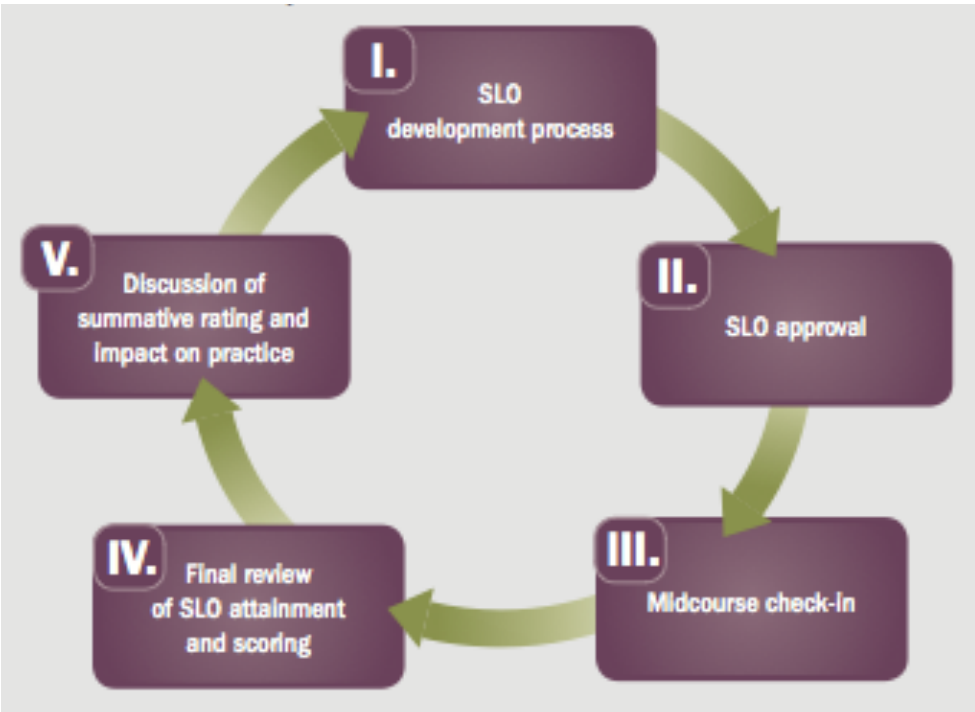
Student Learning Objectives (SLOs) represent content and grade or course specific measurable learning objectives that can be used to document student learning over a defined period of time. They provide a means for teachers to establish learning goals for individual or groups of students, monitor students' progress toward these goals, and then evaluate the degree to which students achieve these goals using relevant measures (see Slotnick, et al., 2004; Goe & Holdheide, 2011; Marion & Buckley, 2011). The active involvement of the teacher throughout the process is a key advantage of the SLO approach over traditional test-centered approaches to accountability. It is designed to reflect and incentivize good teaching practices such as setting clear learning targets, differentiating instruction for students, monitoring students' progress toward these targets, using student learning data to adjust instruction, and evaluating the extent to which students have met the targets. SLOs can constitute an instructional improvement process, driven by teachers in all grades and subjects. More specifically, SLOs must include the following key components:

- The learning goal(s);
- The assessments used to evaluate the learning goals; and
- The targets for student learning growth and teacher performance.

It can be argued that the process of establishing, monitoring and evaluating student learning objectives is something that is a natural expectation for any teacher. In this sense, SLOs are nothing new. The motivation behind the recent surge in interest around SLOs is educational accountability policies that call for the evaluation of teachers on the basis, at least in part, on evidence of growth in student achievement. For teachers who teach students in grades 4 through 8, it is at least possible to find such evidence from the results of state-administered standardized assessments in mathematics and English Language Arts. However, this only accounts for about 25% of all public school teachers, and laws mandating formal teacher evaluations that incorporate evidence of student growth are typically written to apply to all teachers. One response to this problem is to attempt to formalize what teachers do implicitly (set, monitor and evaluate student learning objectives) as part of a district of statewide SLO *process*.

In 2001, Denver Public Schools became the first large urban school district to formalize an SLO process as an element of its professional compensation program (ProComp). In DPS, SLOs have been referred to as "student growth objectives" (SGOs). To maintain a consistent terminology in this report, we will use the term SLO throughout to refer to the new process being described. The term "SGO" will always refer to the approach taken under ProComp. In general, most SLO processes have a similar process, easily visualized in Figure 1.1, which depicts a cycle for each teacher that begins with the development of an SLO, moves to approval of the SLO, then a midcourse check-in with regard to student progress, followed by a final review of SLO attainment at the end of the instructional period and then a period of reflection that should lead to improvements to both teaching practice and the SLO process in a subsequent instructional period.

Figure 1.1 The SLO Evaluation Cycle



Source: American Institutes for Research

During the ProComp pilot years, professional development support teams were deployed at 16 pilot sites to work with teachers on establishing SGOs. A key aspect of this support was to help teachers develop meaningful learning goals and set appropriate but rigorous growth targets to evaluate student growth achieved during the school year. The support teams also provided training to the pilot school principals to help them evaluate the quality of the learning objectives

developed by teachers and to ensure that meaningful feedback and resources were provided to teachers during the mid-conferencing and end of year conference meetings. In addition to the supports provided to teachers to help develop the SGOs, teachers worked in teams to help peers identify effective instructional strategies to move students toward achieving the specified targets. As documented by the evaluators of the ProComp pilot, the Community Technical Assistance Center (CTAC), the collaborations used by teachers to support the SGO work at the pilot sites helped to foster a strong sense of community across teachers (Slotnik et al., 2004).

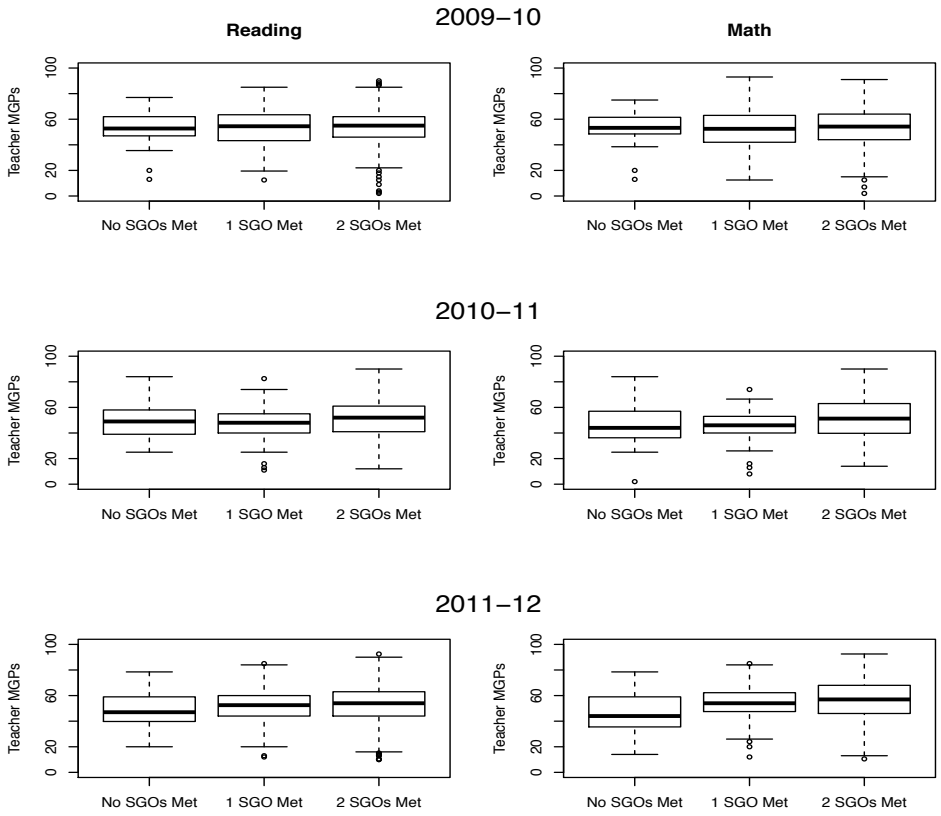
In 2004-2005, Denver voters supported the implementation of ProComp district wide and passed a mill-levy to fund the program. Although voters funded the merit pay system, the original supports implemented for the pilot sites to assist teachers with SGO development were not maintained or scaled up districtwide. Goe and Holdheide (2011) stress that SLOs require continuous professional development and sustained training over time to build capacity, especially for novice teachers, to develop meaningful learning goals and targets. Ten years later, although many DPS teachers still see some potential in the SGO process, SGOs are not viewed as a critical tool for supporting the work that teachers do in the classroom to evaluate student learning progress, but rather as a ProComp element accompanied by weak standards and expectations for student growth (Briggs, Diaz-Bilello, Maul, Turner, & Bibilos, 2014). In short, the process appears to have gone from one viewed by teachers in pilot schools as a valuable and meaningful approach with the potential to support both instructional and learner goals to one that is fundamentally a compliance-based task tied to the receipt of a monetary bonus.

The results from recent evaluations indicate that teacher growth ratings based on SGOs, at least as presently constituted in DPS, would have at best questionable validity as an indicator of quality instruction. Among all DPS teachers that submitted results from two SGOs over a three year period from 2009-10 to 2011-12, 95%, 95% and 93% of these teachers were rated as having met at least one of the two submitted SGOs. In any given year, roughly 75%[1] of DPS teachers meet both of their submitted SGOs. Yet for those teachers who submit SGOs and teach students who take state administered tests in math and English Language Arts, there is only a weak association between SGO attainment and average student growth. This can be seen in Figure 1.2, in which the distributions of mean student growth percentiles in math and reading for

[1] The commonly reported figure of roughly 80% is based upon the percentage of ProComp teachers who meet both SLOs. When factoring in non-ProComp teachers, the figure is roughly 75 to 77% each year.

each DPS teacher are compared as a function of SGO attainment. With the possible exception of math in 2011 and 2012, there are no significant differences between teachers who attain 0, 1 or 2 SGOs in terms of the performance of their students on the Colorado Growth Model. The rank correlations between SGO attainment and MGPs is never higher than 0.12. This negligible association raises questions about the validity of SGO results.

Figure 1.2. Reading and Math Mean Growth Percentiles by SLO Attainment.



Perhaps not surprisingly, there is considerable variability in the quality and rigor of the SGOs that DPS teachers write and submit for approval. Although school administrators who approve SGOs are given a standard rubric to apply in order to ensure that minimal quality criteria have been met, there is some evidence that these criteria are either not well-understood or not consistently applied in the way they were intended. For example, in 2012, out of 1,328 SGOs that were rated using a DPS rubric, 75% were rated as Acceptable or Excellent. As a validation exercise, two content experts from the National Center for the Improvement of Educational

Assessment (NCIEA) used the same rubric to rate a random sample of 30 SGOs that had previously been given an "Excellent" rating. The experts concluded that none of the 30 SGOs merited an Excellent rating and that at least 20 should have been given a "needs improvement" or "does not meet" rating.

At present, there is no incentive for teachers to write especially challenging SGOs. On the contrary, because there are small monetary bonuses attached to attaining one or both SGOs, there is a strong incentive to write SGOs that a teacher's students are almost certain to meet. Additionally, although the SGOs were originally conceived to provide insights to teachers on student learning (Slotnik, et al., 2004), feedback received from teachers participating in focus groups about the SGO process indicated that the SGOs were largely reviewed at the beginning of the year, and revisited at the end of the year (Diaz-Bilello, 2012). That is, little to no efforts were made to progress monitor students during the course of the year and therefore little insights were gained on how students were moving toward meeting critical learning goals.

Results from focus group interviews suggest that SGOs are considered the easiest to manipulate of ProComp incentives (Briggs et al, 2014, pp. 89-90). Empirical evidence in support of this perspective is the finding that the percentage of teachers participating in ProComp who attained both SGOs tends to be about 11-14% higher than teachers not participating in ProComp (Briggs et al, 2014, p. 6). Furthermore, among teachers who use their own tests as outcome measures, the proportion attaining both SGOs was 10% higher relative to those teachers who used national or district-developed tests (Briggs et al, 2014, p. 6). If SGOs were to be associated not only with ProComp bonuses but with teacher evaluation ratings, the incentive to submit gameable SGOs would only increase.

Purpose of the 2013-14 SLO Pilot Project

From both a DPS leadership perspective and from a DPS educator perspective, if SGOs are to be used as a component in the rating of teachers for evaluative purposes, then it is critical that the underlying SGOs can be shown to have at least a minimal degree of validity, reliability and fairness. As part of a longer-term effort to improve the SGO process, DPS's Assessment, Research and Evaluation (ARE) Department contracted with the National Center for the Improvement of Educational Assessment (NCIEA) and the University of Colorado's Center for

Assessment, Design, Research and Evaluation (CADRE) to engage teachers in three DPS schools with a program of professional development in *SLO* development. The more general term "SLO" was embraced to signal a break from the status quo of SGO development under ProComp. A fundamental purpose of this exploratory pilot project was to find out what might be possible if DPS teachers were given the sort of intensive support in SLO development that a subset of teachers had once received prior to the district-wide implementation of ProComp. More specifically:

- Would these supports promote the perspective of the SLO as a critical element of good pedagogy, more than a compliance based activity?
- What benefits can be achieved when teachers are given support to work on the SLO process collaboratively across grades?
- Would these supports help teachers shift their perspective about assessment as something that is embedded in everyday activities, more than a standardized test?
- Would these supports help teachers become informed consumers of externally developed student assessments?

Another purpose of this project was to generate new ideas that could be used to improve the SLO development process. In particular DPS wanted to explore (1) an SLO scoring approach that could incentivize teachers to set challenging learning goals by essentially giving teachers credit for "degree of difficulty" and (2) a teacher ratings approach for student SLO attainment that was more nuanced than the present met/not met dichotomy.

A central innovation in this pilot was to re-conceptualize the SLO process relative to the framework of the Assessment Triangle as described in the National Research Council report *Knowing What Students Know* (Pellegrino et al, 2001), and recent research on learning progressions/learning trajectories that has been conducted in the science and math education research communities. This framework is necessary in our view because SLOs are intended to provide evidence about student growth, and such evidence can only be gathered and evaluated if clear hypotheses are available that delineate the domain of growth and what, exactly, is expected to change in student thinking over some defined instructional period. In a nutshell, we were eager to see if we could move teachers from a perspective that kids either "get it or they don't" to a perspective in which there is greater interest in understanding the "messy middle" that exists

between a novice understanding of some big picture idea (in math, the visual arts, or any subject domain) and the target understanding of that big picture idea after some defined instructional period (e.g., proficiency). Without this more nuanced understanding of student learning, there is very little that can be done when a student demonstrates a lack of proficiency other than reteach the same material and hope for a better outcome.

Overview of this Report

A large portion of this report is dedicated to describing the work undertaken at each pilot site and in each content area since this provides baseline information and data that:

- Explains how the foundation for this work was laid at various district sites;
- Illuminates both site and content specific challenges in year one that can be re-assessed to determine whether those challenges have been adequately addressed by the end of year two;
- Serves as a critical marker for tracking the evolution of the Tier 3 pilot over time.

We anticipate referencing this report at the end of each year as we continue to evaluate and monitor both progress and changes made to the design of the Tier 3 pilots in year 2 and beyond.

This report is divided into two major sections addressing the mathematics pilot first and then moving into the visual arts pilot. In Chapter 2, we present our basic model for SLO development and scoring, a model that is premised upon an operationalization of the assessment triangle idea that can be found in the National Research Council report, *Knowing What Students Know*. A focal point of our model is to get teachers to define a learning progression that characterizes what it means for students to "grow" over the course on an instructional period. The model is presented with the possibility of "scale-up" in mind. While it is intended for use in a context when teachers are given the time and opportunity to collaborate within and across grades, it may be helpful to guide individual teachers faced with the task of writing an SLO appropriate for her students with little to no collaboration with colleagues. The information in Chapter 2 presents the theoretical foundation for the work underlying both the math and the visual arts pilots.

Following Chapter 2, Chapter 3 in Section 1 of the report begins with the logistics of our pilot project to implement this learning progression model with math teachers at three DPS

schools. In this section we present the approach that was intended, and not necessarily the approach that was implemented across sites. Indeed, there were important differences in how professional development activities took place at each site for math. Chapters 4, 5, and 6 in Section 1 describe these differences with respect to activities at our elementary, middle, and high school sites respectively. Chapters 4, 5, and 6 culminate with lessons learned from conducting the pilot at each site.

Following Chapter 6 in Section 1, we move into a description of the work accomplished in the visual arts pilot in Section 2 of this report. The original Tier 3 visual arts pilot took place at two of the DPS schools that were also part of the Tier 3 math pilot. However, in the spring semester, the decision was made jointly by NCIEA and the program manager of the visual arts to include the teachers from one of the school sites into a new Tier 3 pilot to be offered to teacher leaders in the Arts. The original Tier 3 pilot remained intact at Green Beechway high school. Chapter 8 in Section 2 presents an overview of the original pilot and describes the successes and challenges from the work. Chapter 9 uses the same structure as Chapter 8, but focuses on activities accomplished and lessons learned from the arts teacher leader pilot.

We conclude this report with a final chapter that reflects upon accomplishments achieved on initial goals set for the pilot during the planning phase in early June of 2013. These initial goals are revisited since these goals represented some of the early research questions that were developed to evaluate the extent to which the Tier 3 pilots in both math and visual arts could address some of the shortcomings described in this chapter of the SGO system, and support the larger purpose of using learning progressions as a framework for SLOs. Following these reflections, a set of short- and long-term recommendations are provided for the district's consideration as the SLOs are implemented district-wide. We conclude the final chapter by outlining next steps for proceeding with the Tier 3 pilots during the 2014-2015 school year.

References

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

Buckley, K. & Marion, S.F. (2011). A Survey of Approaches Used to Evaluate Educators in Non-Tested Subjects Grades and Subjects. Retrieved from: www.nciea.org.

Goe, L. & Holdheide, L. (2011). Measuring Teachers' Contributions to Student Learning Growth for Nontested Grades and Subjects. Princeton, NJ: National Comprehensive Center for Teacher Quality.

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

Slotnik, W.J., Smit, M., Glass, R.J., Helms, B.J. (2004). Catalyst for Change: Pay for Performance in Denver Final Report. Community Training and Assistance Center. Retrieved from: http://www.ctacusa.com/denver-vol3-final.pdf

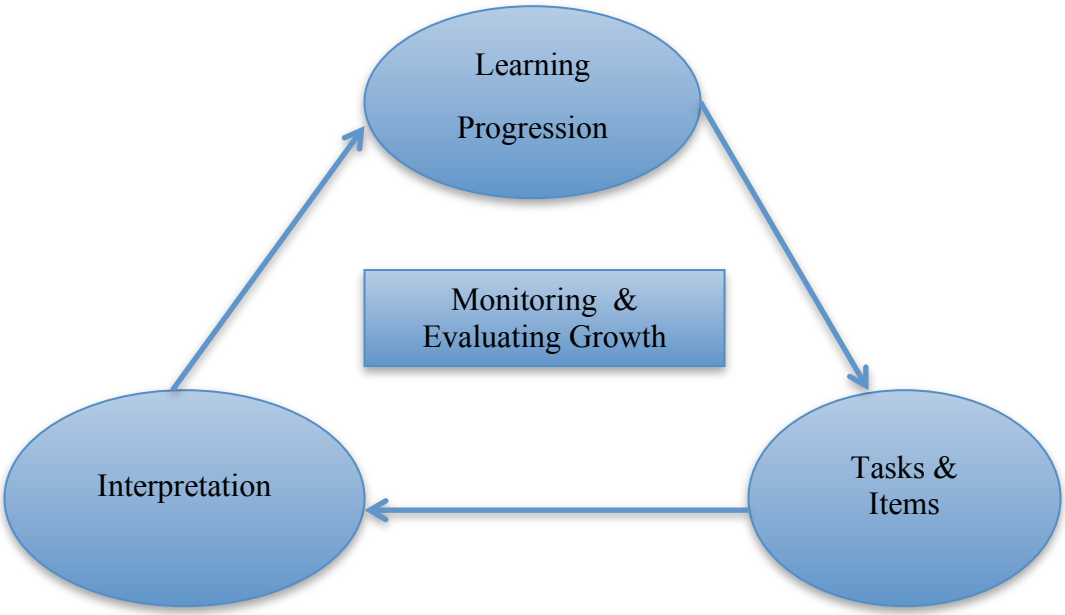
Chapter II: A Learning Progression Model for SLOs

The Assessment Triangle as a Conceptual Framework for SLO Development

Our framework for working with teachers to develop and evaluate SLOs has its roots in the National Research Council report "Knowing What Students Know" (Pellegrino, Chudowsky & Glaser, 2001). In applying this framework, our aim is to conceive of SLO development as part of an iterative and collaborative process Pellegrino et al (2001) describe as the "assessment triangle." The basic idea of the assessment triangle is that any high-quality assessment will always have, at least implicitly, three components: a theory of student cognition (e.g., a qualitatively rich conception of what it is that distinguishes an "expert" from a "novice), a method for collecting observations that elicit evidence about what students know and can do, and methods for making inferences from the fallible information. Although SLO development involves more than creating high-quality assessments, the same three components of the assessment triangle can be tailored to this context by specifying a *learning progression* as a stand-in for a theory of student cognition.

In the simplest sense, a *learning progression* (LP) 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. Within the context of SLOs, an LP can be defined and operationalized in terms of a "big-picture" learning target (or targets) identified by teachers at the beginning of the academic year/semester. These targets would be identified with respect to 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 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 measure that can be used to make a judgment about a student's growth and final position on the learning progression. When these three elements are taken together, they form a principled and systematic basis for monitoring and evaluating growth.

Figure 2.1 A framework for developing SLOs with growth in mind



Unique Features of the Learning Progression Approach to SLO Development and Scoring

To summarize, the SLO framework described above involves the following steps

1. Decide whether an LP will be applied across multiple grades or only within a single grade.
2. Choose big-picture topic for LP.
3. Use standards documents (i.e., CCSS, ELG) to articulate grade-specific anchor(s) for LP.
4. Specify standard 5 level course-specific trajectory that will be basis for assessing student growth toward target.
5. Use assessment items and body of evidence to establish location of students within LP levels at baseline.
6. Conduct score moderation activities as a means of monitoring growth and pilot testing items for end-of-year assessment.
7. Use assessment items and body of evidence to establish location of students within LP levels at end of instructional period.
8. Score student growth.

9. Translate aggregate evidence of student growth into teacher rating.
10. Reflect on the process and make improvements.

Using a Learning Progression to Define an SLO (Cognition Vertex of Assessment Triangle)

The identification of the LP and the learning objective it supports starts with a simple question: "What are the most important things you expect your students to know and be able to do in _____ by the end of the school year?" This usually results in a wide variety of answers that vary in grain size. The goal here is to get teachers thinking concretely about what it is they want to accomplish with their instruction. What aspirations do they have for their kids? Additionally, teachers will need to prioritize the selection of a learning objective for an SLO by considering questions such as the following: What objectives would directly address the specific learning needs of my students? Can these objectives be linked to broader concepts that transcend grade levels and specific instructional units? 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? The next step is to connect these aspirations to the preexisting content standards for the state. It is important for teachers to realize and appreciate that their aspirations can be shown to be consistent with externally imposed expectations. The most important step is to express the learning objective in terms of a big picture target for instruction that represents some combination of knowledge (i.e., content) and the way that students should be able to apply this knowledge (i.e., skills).

In many cases, an LP will include levels that span more than two grades (as we will illustrate in our examples throughout this report). The purpose of such an LP is to focus attention on the changes in student understandings that would be expected to take place across two or more grade levels as students are exposed to instruction targeted to certain core concepts. However, for any given teacher, while it is important to have a big picture perspective on where kids are headed in their understanding, the first priority is to decide how much student growth should be expected within each course. So we draw an important distinction between a multi-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 thing. But usually, teachers will choose some subset of the multi-grade LP as the course-specific trajectory,

and the latter will be the basis for each teacher's SLO, as we will illustrate in this chapter and the ones that follow. Every course-specific trajectory makes it possible to characterize, in a criterion-referenced manner, where it is that students start at the beginning of an instructional period, where they end at the culmination of the instructional period, and how much they have grown in between.

When taking a learning progression approach to develop SLOs, the ideal —something that really distinguishes the approach being described here from the more conventional approaches to SLO development—is to have teachers working collaboratively in vertical teams to establish an LP that tracks the same big picture concept over multiple grades. A basis for such LPs is most readily available in the subjects of mathematics, English Language Arts and science. For example, the common core state standards for mathematics (www.corestandards.org/Math/Content/CC/) provide an organization of standards across grades by the following domains:

- Operations & Algebraic Thinking (grades K-5)
- Number & Operations in Base Ten (grades K-5)
- Number & Operations—Fractions (grades 3-5)
- Measurement & Data (grades K-5)
- Geometry (grades K-8)
- Ratios & Proportional Relationships (grades 6-7)
- The Number System (grades 6-8)
- Expressions & Equations (grades 6-8)
- Statistics & Probability (grades 6-8)

More than one of these domains 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. To continue with this example in the context of the Common Core State Standards for Mathematics, consider the mathematical concept of place value that can be found within the domain of number & operations in base ten. An initial learning progression from Kindergarten to grade 2 could be sketched out as in Figure 2.2 by pulling out standards statements that pertain to target student understandings about place value by grade level.

Figure 2.2 A Hypothetical Learning Progression for Place Value

Grade	Level	Level Descriptor
2	5	A student shall understand that the three digits of a three-digit number represent amounts of hundreds, tens and ones; shall be able to count within 1000, read and write numbers to 1000 using base-ten numerals, number names and expanded form; and compare three digit numbers based on meanings of the tens and ones digits, recording the results of comparisons with the symbols >, = and <.
	4	Partial Understanding ("Messy Middle")
1	3	A student shall understand that the two digits of a two-digit number represent amounts of tens and ones. Student shall be able to compare two digit numbers based on meanings of the tens and ones digits, recording the results of comparisons with the symbols >, = and <.
	2	Partial Understanding ("Messy Middle")
K	1	A student shall be able to compose and decompose numbers from 11 to 19 into ten ones and some further ones and record each composition or decomposition by a drawing or equation. The student will understand that these numbers are composed of ten ones and one, two, three, four, five, six, seven, eight, or nine ones.

In Figure 2.2, each grade level descriptor could be the basis for a different teacher's SLO. Notice that in Figure 2.2, the levels 2 and 4, 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 concepts in the upper grade target, but their thinking about place value is clearly more sophisticated than the lower grade target. In the learning progressions in science literature, this in-between state has been referred to as "the messy middle" because it represents a stage in which students may vary considerably with regard to the ideas central to the learning

progression that they have mastered. 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 having a complete understanding of how these numbers can be decomposed into combinations of tens and ones. This attention to subtle differences in how students are thinking and reasoning is a hallmark of a learning progression 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 within-grade trajectories will usually be undefined. However, over time, teachers will begin to fill this in as they become attuned to patterns in how students develop their understandings.

We now attempt to generalize the learning progression framework such that it could be employed by any teacher for any subject area, with or without a pre-existing 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 being used to define the SLO. Hence, so long as there are standards or goals that a district 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 knowledge and skills that would be expected of all incoming students if they have had the anticipated academic preparation in prior grades (or pre-school in the case of Kindergarten). The *realistic lower anchor* represents the knowledge and skills related to the topic of the SLO that the bottom quartile of students in a teacher's class (or classes) would typically come in with at the beginning of the instructional period. In the place value example illustrated in Figure 2.2, for a 2nd grade teacher, level 5 would represent the aspirational learning target for each 2nd-grade student by the end of the instructional period (e.g., the last month of the school year). Level 3—the understanding of place value expected of students at the end of the 1st-grade—would constitute the aspirational lower anchor; level 1—the understanding of place value expected of students at the end of the Kindergarten—would constitute the realistic lower anchor if it evidence could be provided that significant proportion of a teacher's students enter the 2nd-grade with this level of understanding.

Growth is defined in an absolute sense relative to the movement of most students from levels 1-3 to levels 2-5, where levels 3 and 5 have a well-defined criterion-referenced meaning with respect to what knowledge and skills students have mastered with regard to the big picture concept emphasized by the LP. Growth can be characterized as negative or no growth (a student who remains at the same level or moves backward, minimal growth (movement from one level to the next), aspirational growth (movement of two levels—the distance between level 3 and 5), and exceeded aspirational growth (movement of more than two levels). As long as it would be possible for a teacher (or teachers, or curriculum experts at the district level) in the same subject and grade level to at least operationalize an aspirational learning target (something they are expected to do for any SLO), an aspirational lower anchor and a realistic lower anchor, it would be possible to apply the LP-based approach to SLO development and evaluation. However, the approach is ideally implemented with teachers working in vertical teams to develop a common LP across grades, as presented in Chapters 4-6.

Choosing or Writing Items for Student Assessment (Observation Vertex of Assessment Triangle)

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. The key empirical aspect of this stage is gathering, scoring and comparing student responses to assessment items. This can be done internally when a teacher examines the distribution of student responses within a class, or externally when a set of teachers compares their students' responses to those of other colleagues teaching in the same subject and grade level. This latter process is known as *score moderation*, and is a particularly compelling basis for collaborative professional development.

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 2.2, if a student 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 4 relative to level 5, but not

whether the student is at level 3 or even 2. 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 in the process being 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.

Score moderation activities are central to the LP-based SLO approach, and if done well, can play a dual role of fleshing out the LP in more detail and giving teachers insights into possible formative uses of assessment results. What is score moderation? Score moderation takes place when multiple teachers teaching the same subject at either the same or adjacent grades have administered the same open-ended assessment tasks to their students. Such tasks will typically have scoring rules (e.g., scoring rubrics). In a score moderation session, teachers choose a sample of student responses that they all read and score. Invariably, there will be disagreements about the distinctions being made between student answers—though some answers may be completely right and some completely wrong, many will fall somewhere in between, and these are the ones that will tend to provoke the most discussion and also lead to changes/improvements in how assessment items are chosen, written and scored. In general, the more that the SLO development process can involve teachers sitting together around a table discussing student work and what it implies about learning, the greater the chance that the process will be perceived as more than a compliance-based activity.

Scoring and Interpreting Student Work

It is relatively 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 not necessarily meaningful in the context of a LP because they will depend upon the difficulty of items and the ability of the students. Therefore, 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. In some cases such as the administration of a performance task, the mapping 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 discrete tasks and items, scores from 1 to 20, or 10 to 100, etc. will need to be converted into LP levels. This is analogous to the need to convert scale score on a state's criterion-referenced assessment into achievement levels.

A practical approach that can be taken to convert scores to LP levels is to start by examining the distribution of student scores and dividing the distribution into the same number of equally divided bins (i.e. quartiles, quintiles) represented by a teacher's within-grade LP. This serves as a basis for grouping students who are likely to be relatively similar in their overall performance. The next step is to examine the specific item responses typical of students in each group. For the students in the top performing quantile, the question to be asked is whether the lowest score in this bin is indicative of someone at the top level of the within-grade LP. And so on for each quantile/bin. If, for example, all students in a teacher's class have done well, it would be possible that the demarcation for the top level of the LP is at the 20^{th} percentile of the score distribution rather than the 80^{th} percentile. In contrast, if all students have done poorly, even a score at the 80^{th} percentile might not be sufficient to place a student in the top level of the LP.

A key consideration in the mapping of assessment scores to LP levels is the quality of the underlying assessment. Are the items appropriately targeted to different levels of the LP? Do they cover all the different attributes that characterize knowledge and skills expected of a student at a given level? 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)? Generalizability will be low when there are significant sources of measurement error, and the greater the measurement error, the harder it will be to reliably

distinguish multiple groups of students, especially for assessments that may consist of a relatively small number of items. When this is the case, it may be necessary and advisable to assemble a body of evidence to place students into LP levels. A tool that could come in handy when evaluating the quality of assessments being administered to students would be the "SLO Quality Assessment Check Tool" developed by the National Center for the Improvement of Educational Assessment (see Appendix H)

In this report, we use quantitative and qualitative methods to evaluate the quality of the assessments created at each site. Our qualitative analysis focuses on the extent to which the assessments meet the following criteria from the NCIEA tool:

A high quality assessment:

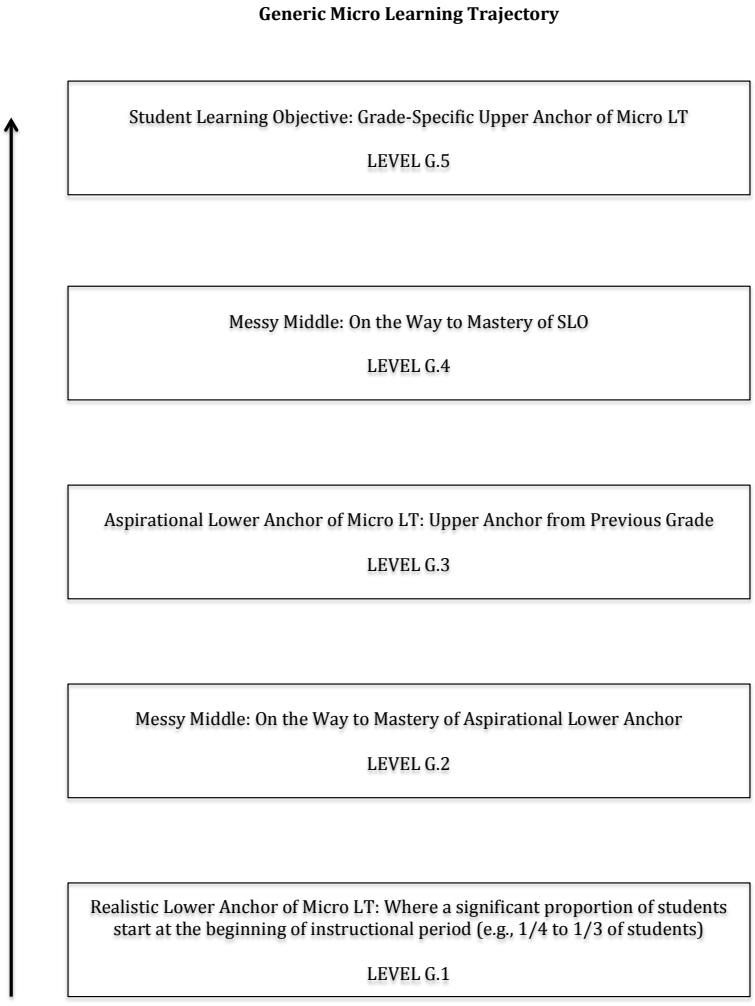
(a) has been determined to be aligned to identified standards and depth of knowledge,
(b) is fair and unbiased,
(c) has a rubric or scoring guide that
    i. allows for reliable scoring,
    ii. appropriately differentiate student performance
    iii. includes evidence to support these rubrics have been validated.

Our quantitative analysis focuses on four quantitative measures: We use Alpha and standard error of measurement (SEM) to assess a test's generalizability. We use item-total correlations to determine the extent to which individual items measure the same construct as the overall test. Finally, we examine the p-values of individual items to determine whether the items on the test spanned a reasonable range of difficulty. For a more detailed description of these statistics, see Math Appendix A.

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. The simplest approach for assessing growth at the student level is to use a body of evidence and a pre-assessment at the beginning of the year to place students somewhere on the bottom three (or possibly) four levels of the within-grade LP at the baseline, and then to

use a post-assessment (or body of evidence) to place students along the within grade LP at the end of the instructional period. Consider the generic within-grade LP shown in Figure 2.3 below. The highest possible growth would be four points from level 1 to level 5. A more complicated approach would involve the use of a "value-table" to assign different point values to different transitions from one level to the next.

Figure 2.3 A Generic Course-Specific Trajectory



Rating Teachers on the Basis of Student Growth

Minimal Growth

If a student is at the same LP level at the end of the school year as he/she was at the beginning, we have a problem. Minimal growth can be defined as the smallest growth score that would move a student up one level of the micro LT. In the example shown in Table 2.1, the most conservative choice would be a growth score of 1, which would be the minimum required to move a student from G.3 to G.4.

Aspirational Growth ("One Year's Worth of Growth")

Aspiration growth should be synonymous with what is meant when educators speak of a child making a "years' worth of growth." One way this could be defined would be to examine the growth score associated with a student moving from the aspirational lower anchor to the target upper anchor.

Exceeding Aspirations ("More than a Year's Worth of Growth")

Any student with a growth score higher than a score defined as aspirational growth would be one who had exceeded aspirational expectations for growth. This could happen either because a student makes dramatic growth starting from G.1 or G.2, or because a student shows growth beyond the target level of the micro LP.

Table 2.1 Example Categorizations of Teachers for Student SLO Growth

Category	%Students Not Making Minimal Growth	% Students Making Minimal Growth	% Students Making Aspirational Growth	% Students Exceeding Aspirational Growth
4 (Higher)	10%	15%	75% Combined	
3	15%	25%	60% Combined	
2	25%	50%	25% Combined	
1 (Lower)	50% or more	Less than 50% Combined		

Note: Percentages in Cells Are All Hypothetical

Many other permutations are of course possible, and constraints on these permutations is something that probably represents a policy decision by district leadership with input from DCTA and other relevant stakeholders. But this approach is flexible enough to accommodate many different perspectives on what should "count" for growth.

The LP-based SLO approach has many similarities with the more conventional SLO development approaches being taken by other states and districts (and previously by DPS as part of the SGO process). Teachers are still expected to set meaningful, standards-based learning targets for students, evaluate where students are relative to those targets at the beginning of an instructional period, monitor their progress over the course of the year, and then assess where they are at the end. Furthermore, the quality of an SLO will be directly related to the quality of the assessments being used to make inferences about what students know and can do at any point in time. There are, however, some important unique features of the model we present in this report.

- Learning Objectives are set at the student level, not at the class level.
- A priority is placed on thinking about *growth* across multiple grades, rather than student status or level of mastery.

- Different targets are not set for different students. Instead differentiation happens in terms of figuring out where kids are at baseline and characterizing them accordingly on the LP.
- Teachers are rewarded for degree of difficulty: more points are awarded for students that start off lower on the LP.
- Clear, criterion-referenced definitions are provided for what it means to have "one year of growth" and what it means to demonstrate minimal growth.
- Score moderation is central to the approach. This fosters teacher collaboration around the careful study of student reasoning, and could serve as the basis for monitoring growth and changing instruction.

Based on the features noted above, growth evaluated under the LP approach no longer requires teachers to predict and develop pre-set expectations on achievement outcomes for students. In other words, growth is no longer evaluated based on the extent to which those predicted outcomes are realized. In most places using SLOs, including the SGO system at DPS, the percentage of students meeting targets set are based upon the percentage of students who meet the pre-set expectations for achievement predicted by teachers at the beginning of the year. For example, a common predicted outcome specified under the SGO system as well as in SLO systems found in other states is: 85 percent (or another percentage) of students will move one proficiency level up from their baseline performance group. The extent to which the pool of students (85 percent of the total class) meets the predicted expectations set by a teacher is used to derive a teacher's rating on the SLO. Under the SGO system, if a teacher did not meet the 85 percent target, he receives a "not met" rating for the SLO. This binary approach (i.e., met or not met) provides little incentive for teachers to set high expectations since a higher bar would lower the probability for teachers to meet the SGO and receive a bonus or a salary bump under ProComp. In other sites using SLOs, ratings are set on different levels of predicted achievement outcomes reached. Figure 2.3 below presents an example of ratings assigned to teacher on SLOs in a different state based on varying percentages attained on predicted achievement:

Figure 2.3: Rating rubric for teachers on SLOs

Rating rubric for teachers with a class of 5 or more students.			
☐ Highly Effective	☐ Effective	☐ Developing	☐ Ineffective
At least 90-100% of students met or exceeded expected target.	At least 75-89% of students met or exceeded expected target.	At least 60-74% of students met or exceeded expected target.	Fewer than 60% of students met or exceeded expected target.

Based on the rules specified in Figure 2.3, to meet an "effective" rating using the previous predicted target example (i.e., 85 percent of students are expected to move one proficiency level up), means that 75 to 89 percent of the 85 percent of the total number of students considered for the SLO would need to move one level up.

In contrast to the typical approach used to evaluate SLO growth in most settings, the LP approach awards growth achieved by any student based on the distance between a student's starting point and where the student lands by the end of the instructional period. Subsequently, teachers avoid setting predicted targets for students at the beginning of the instructional period which has been deemed "unfair" by many DPS teachers (M. Cohen, personal communication, June 25, 2014). As indicated earlier in this section, the goal for all students is to meet the learning objective. However, points are still awarded for students who demonstrate growth along the learning progression continuum.

The next chapter of this report marks the beginning of Section 1. Chapters 3 through 6 in Section 1 describe the structure of the mathematics pilot implemented at 3 school sites and resulting outcomes.

Section 1: Tier 3 Math Pilot

Chapter III: Pilot Implementation of the Learning Progression SLO Approach

Overview

During the 2013-2014 school year, professional development oriented to the SLO model described in the previous section was implemented with teachers in three DPS schools in the subject areas of mathematics and visual arts. The primary purpose of the pilots was to implement the student learning objective (SLO) process using learning progressions as an organizing framework for understanding student work, defining instructional targets, and quantifying the amount of student growth or learning occurring with students during the school year.

This pilot project was the third and most intensive of three tiers of professional development being piloted by the district. The Tier 3 pilots were led by staff from the National Center for the Improvement of Educational Assessment (also known as the Center for Assessment) and the Center for Assessment Development, Research and Evaluation (CADRE) at the University of Colorado Boulder. CADRE led the pilot work in Mathematics and the Center for Assessment led the pilot work in visual arts. Both pilots tested out the use of learning progressions as a framework for implementing the SLO process in the district. In addition to Tier 3, two other SLO pilots took place in the district. The Tier 2 model was led by district assessment coaches and comprised of focusing the SLO implementation work using key standards to inform the work of setting the learning objective. This model was implemented in twelve schools. The Tier 1 model was also led by district assessment staff, but this model only consisted of introducing the SLO process and concept to all district teachers through on-line modules. Both Tier 2 and 3 models were implemented with the intent to understand how the SLO implementation could work using two different levels of support, with Tier 3 providing the most intensive guidance and support to participating teachers. The work undertaken to implement Tier 3 in mathematics, and recommendations made based on lessons learned from the pilot experience are captured in this report. The Tier 3 visual arts work is described in Section 2 which begins on Chapter 7 of this report. The decision to carry out the Tier 3 work in both math and visual arts was based on the district's interest to find out how professional development needs would differ for teachers instructing in a core state-tested subject area relative to a non-tested elective area.

The district recruited three schools to join the Tier 3 pilot: Blue Elementary School, Spruce Middle School, and Green Beechway High School (GB). The three schools were not intended to be representative of DPS schools. The idea was to pick three sites in which teachers and/or the administration had a predisposition to work collaboratively in vertical teams to create challenging and innovative SLOs. In all three sites, administrators made the decision to participate in the pilot. It is fair to say that participating teachers, for the most part, began the process with a healthy and understandable degree of skepticism.

It is important to point out that the idealized SLO model described in the previous chapter was not something that was fully understood or established at the outset of our work with teachers at each school site. What we knew was that we wanted to use learning progressions as a framework for specifying learning targets, choosing assessment tasks, and evaluating growth. What was refined along the way was how these pieces would all fit together and how much time it would take to assemble the pieces into a coherent process. In other words, we were not piloting a known form of professional development around SLOs; rather we were creating a new and unique SLO process. Many of the elements of the process presented in Chapter 2—the distinction between an across-grade LP a course-specific LP, the approach for mapping assessment scores back to LP levels, the scoring of growth, definitions of minimal and aspirational growth, etc.—were things that we developed and discovered over the course of the pilot. We had a big picture sense for novel ways that we thought the SLO process could be strengthened, and an outline for how this process would be implemented, but we anticipated that the details might be subject to change once we had a better understanding of the needs and culture of the teachers in each site.

Each participating site had committed at the outset to roughly 30-40 hours of contact time over the course of five sessions in one academic school year. As an additional incentive, teachers were told they would be eligible to receive a PDU credit (worth $759) through ProComp for participating in the sessions. Our initial plan for the five sessions, expressed in outline form, was as seen in Figure 3.1.

Figure 3.1 Proposed Section Objectives

Session	Description of Session Objectives
1	Introduce concept of the assessment triangle; begin development of a Learning Progression that serves as foundation for the SLO.
2	Refinement of Learning Progression; identifying sources for baseline data, writing/picking assessment items for use to evaluate location of students on Learning Progression
3	Understand required components for developing a high-quality rubric, creating open-ended tasks that can be embedded in classroom activities and then scored with a rubric linked to the Learning Progression.
4	Teachers engaging in progress monitoring via score moderation; Consider revisions to Learning Progression, assessment tasks, and/or scoring rubric
5	Bringing things together, score student growth and examine how this relates to teacher ratings, lessons learned, debrief, making improvements

A central component of the professional development that we had initially envisioned involved the co-creation of an LP with the teachers at each site during sessions 1 and 2. In the first session, after introducing the concept of an LP and providing teachers with an exemplar on the topic of proportional reasoning that we had pulled from the research literature ahead of time, we wanted to let teachers at each site decide on the "big picture concept" in mathematics that would be most useful to them as a basis for crafting student learning objectives for teachers in each grade. In all three sites we did this by prompting them to think about a core concept that was at the root of difficulties students typically had with mathematics from grade to grade. In one of the sites, Spruce, teachers chose to use the same exemplar LP that we had developed in advance. In the other two sites, teachers decided upon different LP topics. Blue teachers chose the topic of place value; GB teachers chose the topic of algebraic manipulations.

One of the things we learned in hindsight was that it was a mistake to commit a full two sessions (and as it turned out at most sites, quite a bit more) to the process of creating a fully elaborated learning progression. That is, fully elaborated in the sense that there would be descriptors for not only grade-level anchors, but for what students were expected to know and be able to do in between the grade level anchors. It was possible to accomplish this with considerable effort at Spruce and GB because of differences in expertise between the teachers at

these sites were small enough to make the process of defining LP levels a meaningful collaborative experience. At Blue, this process was more difficult because the grade span of the LP was twice as big and the number of teachers involved was much. In part, it was this experience that led us to the simplified approach described in Chapter 2 of this report.

Sessions with Blue were led by Derek Briggs and Abby McClelland, sessions at Spruce were led by Jessica Alzen, and sessions at GB were led by Fred Peck. In every session, at least two members of the research team were present with one member taking the lead to facilitate, and one member taking notes and providing assistance to teachers during small-group activities. The dates and length of each session often varied across sites due to idiosyncrasies in teachers' schedules and availability at each school. These differences are detailed in the subsections below that describe characteristics of participants along with session dates and durations by school.

Blue Elementary School

A total of 18 Blue teachers who taught mathematics in grades K-5 participated in SLO professional development activities along with the school's principal. We had originally planned to work with a subset of the teachers in grades 3-5. However, the school's principal requested to include all six grades. Although we did not directly gather data on teachers' teaching experience, anecdotally we learned that some were in their first or second year of teaching, while others had taught for 10 or more years.

Blue is in many ways an exemplar of a school in which it is the norm for teachers to work in "data teams" and collaborate both within and across grades. In our opening session Blue teachers reported that data teams were an important part of their practice and helped facilitate their goal of vertical integration. In previous months, the math teachers had watched videos of classroom practice on the topic of number base 10, viewed in ascending grade level order. This had helped teachers understand both how lower grades were introducing this subject and providing foundational concepts and vocabulary and what understandings and applications teachers at higher grade levels were expecting. This had helped the teachers identify gaps – issues that needed to be introduced or prioritized in previous grades – and begin aligning vocabulary. The teachers also had definite ideas on how the data teams could be improved. They mentioned that vertical alignment meetings tended to have a macro focus, preventing more fine-

grained attention on content between two adjoining grade levels, for example. The teachers also hoped to use the teams to focus more on incorporation of CCSS information and reconciling those standards with the existing "Everyday Math" curriculum.

Table 3.1 below summarizes dates of each session and total hours Blue teachers spent participating in these sessions.

Table 3.1. Summary of SLO PD Sessions with Blue Teachers

Session Number	Session Date	Hours
1	10/1/13	7
2	11/12/13	4
3	1/6/14	4
3	1/23/14	7
4	2/18/14	7
5	4/7/24	4
TOTAL		33

Spruce Middle School

There were four teachers and a math/science instructional coach involved in the SLO Professional Development activities at Spruce: Dave (6th-grade), Cora (7th-grade), Mike (8th-grade), Lisa (Newcomers), and Jan (Instructional Coach). Lisa instructs all 6th-grade students who have been in the country anywhere from just a few months to two years and do not have the necessary skills to be enrolled in the grade-level classes. Spruce has been designated as the school in DPS for all refugee students, hence this position existing at the school. Years of experience for the teachers at this site range from one year (Dave) to over twenty-five years (Cora & Jan).

Since Spruce only has one math teacher per grade-level, the team at this site is very small compared to the others involved in the pilot program. Although this made it impossible to have within-grade collaboration, it enabled every teacher to be deeply engaged in all of the professional development activities. A common theme in every session was active involvement

and vocal participation by each member of the team. Nearly all activities were completed collaboratively with all of the teachers rather than breaking into small groups. This became a strength of the site as every teacher was involved in creating or modifying nearly every artifact for every grade-level.

At our first session, the teachers shared that it is a regular practice for all teachers at Spruce to analyze assessment data and use the results of those analyses to inform instruction. However, since there is only one teacher per subject per grade, there are no actual data "teams". The teachers had previously discussed working more in vertical teams, but had not yet had the opportunity to do so. They expressed this as a way that data teams could be strengthened at their site. Table 3.2 below summarizes dates of each session and total hours Spruce teachers spent participating in these sessions.

Table 3.2. Summary of SLO PD Sessions with Spruce Teachers

Session Number	Session Date	Hours
1	10/10/2013	7
2	11/12/13	4
3	1/23/14	7
4	2/5/14	7
5	3/14/2014	1.5
6	4/11/14	1.5
7	4/28/14	1.5
TOTAL		29.5

Green Beechway High School

Nine teachers (six female and three male) participated in the project at Green Beechway High School. These teachers teach the following classes: Math 090, Algebra I, Geometry, Algebra II, Precalculus, Statistics, and AP calculus AB. Math 090 is a support class for students that are concurrently enrolled in other math classes. The Math 090 teacher participated in all of our sessions, including discussion and artifact creation. However, she did not write an SLO in the

way that all other teachers did, nor did she administer or score any assessments. This is because all of her students were already enrolled in another math course for which there was an SLO written and the attendant assessment. We didn't explicitly collect data on the teachers' experience, but anecdotally it seemed that all teachers had at least a few years of experience.

The teachers reported that, within the math department, there were a few existing structural practices regarding assessment, including data teams, common assessments, and SGOs. Data teams are a district-wide initiative but they are implemented differently within each school. The GB math teachers reported that for them, data teams were used mostly to monitor student grades and attendance, rather than student learning. However, the teachers reported that within the department, they had created other initiatives to monitor student learning, including common assessments and "retention quizzes," which are skill-based quizzes to assess the extent to which students "retain" skills that were learned earlier.

When teachers discussed common assessments, they did not discuss assessment design at all. For them, the bigger issue was what to do with the data one they had been collected. This discussion had two main themes: (a) how to talk about the data as a group, and (b) how to use the data to inform instruction in the classroom. The teachers reported that they needed more training in having conversations about the data that they have collected on common assessments, and they lamented that they had little time for these conversations.

The teachers reported that they wanted to use the data to inform their instruction, but they faced a conflict between what the data were telling them about students, and what their curricula were telling them about what they should teach. In particular, the teachers reported that, in general, their assessments suggested that students were "coming in without the skills at the bottom" – that is, the prerequisite skills that the curriculum assumes that students have. Given this, teachers reported a tension about whether they should teach students these prerequisite skills or teach to the curriculum. Different teachers came down on different sides of this issue, but it's a discussion that often came up during our sessions.

Table 3.3 below summarizes dates of each session and total hours GB teachers spent participating in these sessions.

Table 3.3. Summary of SLO PD Sessions with GB Teachers

Session Number	Session Date	Hours
1	10/15/13	7.5
2	11/12/13	7
3	1/6/14	7
4	2/18/14	7
5	5/30/14	4
TOTAL		32.5

Chapter IV: Results by Math Pilot Site: Blue Elementary School

Using a Learning Progression to Define an SLO at Blue

Blue teachers chose the topic of place value within number within the larger domain of numbers & operations base 10 as the basis for defining an LP that spanned grades K through 5. Along with misconceptions about fractions, this was viewed as a core concept that made it difficult for students to be successful in many other areas of mathematics as they advanced into the upper grades of elementary school. In defining a place value LP, Blue teachers essentially followed the process illustrated in chapter 2; they referenced the CCSS-M and pulled standards statements from the domain of numbers & operations base 10 that had to do with place value. For each grade level, it includes the relevant CCSS standards and content, information about what students should know and be able to do, activities and assessment items relevant to that content, language and vocabulary that students should know and, for some grades, information about common student misconceptions. Figure 4.1 displays a truncated version of this LP, formatted so that it fits within two pages (in the initial draft version of the LP, each grade was a different page) to make it easy to see in glance how students are expected to grow in their understanding of place value from pre-school through the fifth grade.

Working in grade-specific teams, Blue teachers were able to sketch out the place value LP in a few hours in the afternoon of our first PD session in September. However, debate and discussion about the details of this LP, and its relationship to each teacher's course-specific trajectory, continued through the fourth session in March. At multiple junctures, we reviewed the LP to determine whether it fully and adequately captured teacher expectations for what a student would know and be able to do with respect to place value at the end of a given grade. During session 3, teachers raised several overarching concerns. First, teachers were concerned that the LP levels were simply a collection of the relevant Common Core State Standards (CCSS) for the subject matter and grade level. (Although the standards for Kindergarten/ECE are not based on CCSS, the teachers report they are based on another set of standards that covers that grade level.) The levels of the LP did not attempt to integrate the standards, nor did they include any expectations for skills or knowledge that are not included in the standards.

Figure 4.1. First Draft of Blue's Spruce Value Learning Progression (p. 1 of 2)

Level	Place Value Learning Objective
Pre	1. Verbally counts to 10 2. One to one correspondence 3. Can count up to 10 objects accurately 4. Combines and separates up to 5 objects and describes the parts 5. Identifies numerals to 10 by name and connects each to counted objects
0	1. Compose and decompose numbers from 11 to 19 into ten ones and some further ones, e.g., by using objects or drawings, and record each composition or decomposition by a drawing or equation (e.g., 18 = 10 + 8); understand that these numbers are composed of ten ones and one, two, three, four, five, six, seven, eight, or nine ones [K.NBT.1] 2. 10 can be thought of as a bundle of ten ones — called a "ten" [1.NBT.2a]
	Misconceptions that may help define the messy middle: Thinking that the 1 in the tens place equals 1 (not ten); Not understanding that one group can consist of more than one object
1	1. Understand that the two digits of a two-digit number represent amounts of tens and ones. [1.NBT.2] 2. The numbers 10, 20, 30, 40, 50, 60, 70, 80, 90 refer to one, two, three, four, five, six, seven, eight, or nine tens (and 0 ones) [1.NBT.2c] 3. Compare two two-digit numbers based on meanings of the tens and ones digits, recording the results of comparisons with the symbols >, =, and <. [1.NBT.3]
	Misconceptions that may help define the messy middle: Not understanding what the comparison symbols mean; identifying the larger number, not using the complete comparison sentence (verbally or written).
2	1. Understand that the three digits of a three-digit number represent amounts of hundreds, tens, and ones; e.g., 706 equals 7 hundreds, 0 tens, and 6 ones [2.NBT.1] 2. 100 can be thought of as a bundle of ten tens — called a "hundred." [2.NBT.1a] 3. The numbers 100, 200, 300, 400, 500, 600, 700, 800, 900 refer to one, two, three, four, five, six, seven, eight, or nine hundreds (and 0 tens and 0 ones) [2.NBT.1b] 4. Read and write numbers to 1000 [2.NBT.3] 5. Compare two three-digit numbers based on meanings of the hundreds, tens, and ones digits, using >, =, and < symbols to record the results of comparisons. [2.NBT.4] 6. Mentally add 10 or 100 to a given number 100–900, and mentally subtract 10 or 100 from a given number 100–900 [2.NBT.8]

Misconceptions that may help define the messy middle: Equivalence of amounts; A digit in the hundreds place can represent something other than 100; knowing where the comma belongs; inability to use a number grid

Figure 4.1. First Draft of Blue's Place Value Learning Progression (p. 2 of 2)

Level	Place Value Learning Objective
3	1. Use place value to understand to round whole numbers to the nearest 10 or 100. [3.NBT.1] 2. Recognize that in a multi-digit whole number, a digit in one place represents ten times what it represents in the place to its right [4.NBT.1] 3. X 1 digit whole numbers by multiples of 10 in the range of 10 – 90 using strategies based on place value. (Fact extensions/Extended facts) [source?] 4. Read and write multi-digit whole numbers to ten-thousands using base-ten numerals, number names, and expanded form. [source?] Misconceptions that may help define the messy middle: ??
4	1. Read and write multi-digit whole numbers to millions using base-ten numerals, number names, and expanded form. Compare two multi-digit numbers based on meanings of the digits in each place, using >, =, and < symbols to record the results of comparisons. [4.NBT.2] 2. Use place value understanding to round multi-digit whole numbers to any place. [4.NBT.3] 3. Use decimal notation for fractions with denominators 10 or 100. [4.NF.6] 4. Compare two decimals to hundredths by reasoning about their size. Recognize that comparisons are valid only when the two decimals refer to the same whole. Record the results of comparisons with the symbols >, =, or <, and justify the conclusions, e.g., by using a visual model. [4.NF.7] 5. Performs computations by applying conceptual understandings of place value, rather than by applying multi-digit algorithms [source ??] Misconceptions that may help define the messy middle: When & where to estimate; Rounding to largest place for estimates; Not lining up place values; .30 > .3; 50 x 60 – 300
5	1. Recognize that in a multi-digit number, a digit in one place represents 10 times as much as it represents in the place to its right and 1/10 of what it represents in the place to its left. [5.NBT.1] 2. Explain patterns in the number of zeros of the product when multiplying a number by powers of 10, and explain patterns in the placement of the decimal point when a decimal is multiplied or divided by a power of 10. Use whole-number exponents to

denote powers of 10. [5.NBT.2]
3. Read, write, and compare decimals to thousandths [5.NBT.3]
4. Read and write decimals to thousandths using base-ten numerals, number names, and expanded form. [5.NBT.3a]
5. Use place value understanding to round decimals to any place. [5.NBT.4]
6. Performs exact and approximate multiplications and divisions by mentally applying place value strategies when appropriate. [Source ??]

Misconceptions that may help define the messy middle: # of digits over place value; .127 > .9; .30 > .3; "Round up or down"; 4357 rounds to 4457; 4327 to nearest hundred = 300; 50 x 60 = 300 instead of 3,000; Lining up place values

Second, it became apparent that the number of relevant standards varied significantly by grade level, with some grade levels having two or three standards and other grade levels having six or more. This variety in the number of relevant skills a student is expected to master at different grade level concerned teachers.

Course-Specific Trajectories

Our experience with Blue brought to light a tension between the distinction between the full across-grade LP, and course-specific trajectories. Much of our third session was spent with Blue teachers working in grade-specific teams to establish five levels that would be used to assess student growth over the span of their instructional period with the students. To do this, teachers used the LP to define two anchor points on their course-specific trajectories: the aspirational upper anchor (defined by the grade-specific standards found in the CCSS), and the aspirational lower anchor (defined by the standards for the previous grade level found in the CCSS). This left three levels of the trajectory undefined—the realistic lower anchor (defined as the level of understanding expected of the lowest achieving quarter of student at the beginning of the school year), and the messy middles in between the realistic lower and aspirational lower and in between the aspirational lower and aspirational upper (See discussion on in Chapter 2). Our task for teachers was only to define the realistic lower anchor on their basis of their prior experience teaching the concept of place value at their current grade level. (See Figure 2.2 in Chapter 2)

Although many teachers appreciated this approach, since it would help them to distinguish differences in the baseline skills of incoming students and thereby make it easier to

characterize growth, a small but vocal contingent of grade 4 and 5 teachers expressed strong reservations with the distinction being made between the full multi-grade LP and course-specific trajectories. To these teachers, the whole point of the place value LP was to create a common, standardized vocabulary for SLOs. Indeed, the metaphor of choice for these teachers was that the place value LP should be regarded as a "ruler." From this perspective, having grade-specific teams establish a realistic lower anchor (or elements of partial understanding in the messy middle) would run the risk of that teachers in adjacent grades might establish conflicting markers on what should be a standard ruler. To these teachers, it made little sense to force teachers to choose a subset of the full LP to characterize the levels of their course-specific trajectory since the full LP could be used to characterize the levels of each student at the beginning and end of the course. This tension between an approach that is tailored/individualized to the needs of teachers within the same grade relative to an approach that is standardized across grades became a recurring theme in our sessions. We return to this issue again later in this chapter in the subsection on scoring student growth.

Figure 4.3 shows the final version of Blue's LP for place value, which includes common definitions for partial understandings between grade-specific anchors. Notice that in this final version of the LP, with the exception of Kindergarten the decision was made to designate course-specific trajectories only in terms of three levels: the aspirational lower anchor, which always corresponds to the upper anchor from the prior grade, the partial understandings of the messy middle, the upper anchor. However, the teachers left open the possibility that each student's baseline level could be selected from a level in a lower (or higher) portion of the full LP.

Choosing or Writing Items for Student Assessment at Blue

Initial Task Development and Score Moderation

Prior to our second PD session, we had reviewed and assembled a set of exemplar tasks related to the concept of place value. The set of tasks included a task focused on place value for each grade level represented at Blue. In reviewing tasks, we selected ones that we felt would elicit information about the student's thinking, cognitive models, and strategies used to solve a problem. We also evaluated the student information elicited by the task for relevance to the Blue LP to ensure that student responses would provide information that could be related back to the

LP. The exemplar tasks came from sources including PARCC (http://www.parcconline.org/samples/item-task-prototypes) and the New Zealand model trajectory on place value (http://www.nzmaths.co.nz/sites/default/files/images/NumBook1.pdf) .

Figure 4.3: Final Version of Place Value Learning Progression (p. 1 of 4)

****

Kindergarten Trajectory

Level pre K: (Realistic Lower Anchor)

Students can recognize numbers 1-3. Students can verbally count to 3. Students can accurately count up to 3 objects.

Level pre-K: (Messy Middle)

Students can count accurately up to 5 objects with supports. Students can count verbally up to 5. Students can recognize numbers 1 to 5.

Level K.0: (Kindergarten's Aspirational Lower Anchor)

Students can verbally count to 10. Students can accurately count up to 10 objects, with one to one correspondence. Students can separate up to 5 objects and can identify which group has more or less. Students can recognize numbers up to 10 and connect each with counted objects.

Level K.5: (Messy Middle)

Students can verbally count to 20. Students can accurately count up to 20 objects, with one to one correspondence. Students can recognize that a teen number has a "1" in front of it. Students can separate up to 10 objects and identify which group has more or less.

Level 1.0: (Kindergarten Upper Anchor)

Students can compose and decompose numbers from 11-19 into ten ones and some further ones by using objects or drawings and record each composition and decomposition by drawing or equation (e.g. 18=10+8), understand that these numbers are composed of ten ones and one, two, three, four, five, six, seven, eight, or nines. Students can understand that 10 can be thought of as a group of ten ones, called a "ten".

****

First Grade Trajectory

Level 1.0: (First Grade Aspirational Lower Anchor)

Students are able to compose and decompose numbers from 11 to 19 into ten ones and some further ones but are unable to compose and decompose 20-99 into amounts of tens and ones. Additionally, students are able to compare numbers verbally (bigger/smaller).

Standards: K.NBT.1; I.NBT.2a

Level 1.5: (Messy Middle)

Students are able to compose and decompose numbers from 0-99 into tens and further ones may be unable to compare those numbers verbally (using academic language describing place value) or using symbols. Or, students are able to compare numbers using symbols from 0-99 but are unable to compose and decompose the numbers into groups of 10s and further ones*.

*Use of manipulatives allows students to compose and decompose numbers from 0-99.

Level 2.0: (First Grade Upper Anchor)

1. Students will be able to compose and decompose any two-digit number (11-99) into groups of tens and further ones verbally and in writing. Explain that the numbers 10, 20, 30, 40, 50, 60, 70, 80, and 90 are groups of 1, 2, 3, 4, 5, 6, 7, 8, and 9 tens with zero ones.

2. Students will be able to compare two-digit numbers using the terms and symbols for greater than, less than, and equal to and explain the importance of place value (tens and ones) within the comparison.

Standards: 1.NBT.2; 1NBT.2c; 1.NBT.3

****

Second Grade Trajectory

Level 2.0 (Second Grade Aspirational Lower Anchor)

1. Students will be able to compose and decompose any two-digit number (11-99) into groups of tens and further ones verbally and in writing. Explain that the numbers 10, 20, 30, 40, 50, 60, 70, 80, and 90 are groups of 1, 2, 3, 4, 5, 6, 7, 8, and 9 tens with zero ones.

2. Students will be able to compare two-digit numbers using the terms and symbols for greater than, less than, and equal to and explain the importance of place value (tens and ones) within the comparison.

Level 2.5 (Messy Middle)

Students can understand that the two digits of a two-digit number represent the amounts of tens and ones (e.g., 67 equals 6 tens and 7 ones). Students understand that the three digits of a three-digit number represent the amounts of hundreds, tens, and ones (e.g., 706 equals 7 hundreds, 0 tens, and 6 ones).

Students may compare two two-digit numbers based on their understanding of the tens and ones digits, using >, =, < symbols to record their result comparison. *Students may be able to apply skills to two three-digit numbers with the assistance of manipulatives.*

Students can explain their understanding of place value, within context, *and with language supports* (eg., tens and ones, digits, and greater than or less than).

Students read and write numbers up to 100, and count on by 5s and 10s. Students can count on by 10s from any given number to 100.

Level 3.0 (Second Grade Upper Anchor)

Students will be able to read, write, and compare (using <, >, and =) multi-digit whole numbers to the hundreds place using standard form, word form, and expanded form. Students will demonstrate understanding of what each digit represents. Students can read and write numbers up to 999, and count on by 5s, 10s, and 100s. Students can mentally add 10 or 100 to any given number, and mentally subtract 10 or 100 from any given number to 999. Round to nearest ten.

****

Figure 4.3: Final Version of Place Value Learning Progression (p. 3 of 4)

Third Grade Trajectory

Level 3.0 (3rd Aspirational Lower Anchor)

Students will be able to read, write, and compare (using <, >, and =) multi-digit whole numbers to the hundreds place using standard form, word form, and expanded form. Students will demonstrate understanding of what each digit represents. Students can read and write numbers up to 999, and count on by 5s, 10s, and 100s. Students can mentally add 10 or 100 to any given number, and mentally subtract 10 or 100 from any given number to 999. Round to nearest ten.

Level 3.5: (Messy Middle)

Students are able to demonstrate understanding of place value through hundred-thousands place in one or two of the forms (but not all three). Students are able to perform place value tasks, but are not able to explain thinking.

Level 4.0: Upper Anchor From Macro LT

Students will be able to read, write, and compare (using <, >, and =) multi-digit whole numbers to the hundred-thousands place using standard form, word form, and expanded form. Students will demonstrate understanding of what each digit represents. Round numbers to the hundreds. Apply extended facts to the tens place value.

****

Fourth Grade Micro Learning Trajectory

Level 4.0: (4^{th} Grade Aspirational Lower Anchor)

Students will be able to read, write, and compare (using <, >, and =) multi-digit whole numbers to the hundred-thousands place using standard form, word form, and expanded form. Students will demonstrate understanding of what each digit represents. Round numbers to the hundreds. Apply extended facts to the tens place value.

Level 4.5: (Messy Middle)

* Understand the part-whole relationship

* Read, write, identify, and compare place value to the hundred thousands and millions

* Translate and compare numbers to the hundred thousands and millions

* Round numbers to the hundred thousands and millions

Level 5.0: (4^{th} grade upper anchor)

Students can read and write numbers to millions using words, numbers, and expanded form. Round whole numbers up to millions place and compare multi-digit whole numbers using comparison symbols. Students can read, write and convert decimals to fractions with denominators of 10 or 100. Apply extended facts up to hundreds.

****

Figure 4.3: Final Version of Place Value Learning Progression (p. 4 of 4)

Fifth Grade Learning Trajectory

Level 5.0 (5th Grade Aspirational Lower Anchor)

Students can read and write numbers to millions using words, numbers and expanded form. Round whole numbers up to millions place and compare multi-digit whole numbers using comparison symbols. Students can read, write and convert decimals to fractions with denominators of 10 or 100. Apply extended facts to hundreds.

Level 5.5 (Messy Middle)

Students can write read and write and compare decimals to the thousandths. Can write decimal numbers in expanded form to the thousandths place (60.37 = 60 + 0 + .3 + .07). Students will round numbers to any place. Explain patterns in the number of zeros in products or quotients when multiplying or dividing by multiples of 10. (3 x 40 = 120, 20 x 60 = 1200). Students have a conceptual understanding to multiply or divide whole numbers by powers of 10.

Level 6.0 (5th grade Upper Anchor)

Students can read and write numbers from whole numbers that include decimals to thousandths with numerals and expanded form (i.e. 60.37 = 6 x 10 + 3 x 1/10 + 7 x 1/100). Students recognize in a multi-digit number, a digit in one place represents 10 times more than the place to its right and 1/10 the place to its left. Students can explain patterns in the placement of the decimal point when any given number is multiplied or divided by a power of 10. Students have a conceptual understanding to multiply or divide whole numbers by powers of 10.

Level 6.5 (Aspirational Upper Anchor)

Students can use scientific notation for writing and comparing orders of magnitude. Students have a conceptual understanding to multiply or divide whole numbers by powers of 10. Students will apply place value strategies to model math. Students can write and evaluate numerical expressions with whole number exponents.

In session 2, we grouped the teachers with a mix of grade levels represented in each group. We distributed the packet of exemplar tasks and asked each group to review the sample tasks. For each task, the group was asked to develop a rubric with 4 score levels, consider what information about student understanding would be elicited by the task, and consider how information from this task might relate back to the place value LP. After reviewing the exemplar tasks, we grouped teachers by grade level and asked them to develop one assessment task and rubric for their students. We instructed them to use the sample tasks they had reviewed as reference or a starting point and asked them to focus on creating a task that elicited information about student understanding that was relevant to the place value LP.

Some of the grade level groups decided to adopt the exemplar tasks with small modifications. Other groups, such as the grade 5 group, decided to create tasks from scratch or adopt tasks they had already developed. In this activity, especially the portion where teachers developed rubrics, teachers had difficulty shifting from a conception of using assessment tasks to evaluate whether students had learned material to using assessment tasks to elicit richer information that could be related back to a learning trajectory. Several of the draft rubrics paired with tasks with a number of subsections and awarded one point for each subsection that a student got correct. In response to this tendency, we encouraged teachers to think about different ways a student could approach their sample assessment task and how student work might demonstrate which approach an individual student had employed. While some groups shifted focus in response to this feedback, others remained focused on the correctness of student responses, rather than on eliciting information with explicit connections to the LP. As an example of the latter, Figures 4.4 and 4.5 show the task and associated scoring rubric developed by one of the grade 5 teachers.

Figure 4.4. Example of a First Draft of a Grade 5 Assessment Task1a. A baby Koala weighs 0.51 gram at birth. A Joey (baby kangaroo) weighs 0.8 gram at birth. Compare their birth weights using one of these comparison symbols: >, <, or =

0.51 grams _____ 0.8 grams

1b. How much do ten baby Koalas weigh? _____ grams

1c. Explain how you found your answer.

1d. Write 0.51 expanded notation.

1e. One gram (g) equals 1,000 milligrams (mg). How many milligrams does a Koala weigh at birth?

1 g = 1 ,000 mg

0.51 g = _____ mg

Figure 4.5. Draft Scoring Rubric Associated with Grade 5 Assessment task

1a. .51g < .8g

1b. 5.1 grams

1c. Sample: "I moved the decimal point one place to the right to multiply by ten."

1d. 5 x 0.1 + 1 x 0 .01 or 5 x 1/10 + 1 x 1/100

1e. 0.51 g = 510 mg

5 points: All five parts correct, including explanation mentioning powers of ten for C

4 points: Any four parts correct

3 points: Any three parts correct

2 points: Any two parts correct

1 point: Any one part correct

The scoring for this task is based on counting up the number of correct responses to the five questions contained in this task rather than identifying which students have more sophisticated understanding of place value. Implicitly this assumes that any two students who answer any combination of three questions correctly have the same level of understanding and misconceptions about place value. Instead, it is likely that the parts of the task answered correctly or the methods employed by individual students may indicate differing levels of understanding of place value. Scoring with this rubric as written doesn't clearly distinguish between students who may be at different levels of the place value LP. For an example where teachers were making clearer connections between their tasks and the LP, consider Figures 4.6 and 4.7 which provide a task and scoring rubric (with comments) targeted to the upper anchor of grade 2 on the LP. The task for grade 2 places emphasis on student understanding and sophistication of work rather than total number of points earned or parts correctly completed. This reflects greater understanding of the connection between tasks, student work, and the LP.

Figure 4.6. Example of a First Draft of a Grade 2 Assessment Task

Leslie has three number cards.

1. What is the largest three-digit number Leslie can make with her cards?

Explain to Leslie how she can make the largest possible three-digit number with her cards.

2. What is the smallest three-digit number Leslie can make with her cards?

Explain to Leslie how she can make the smallest possible three-digit number with her cards.

Figure 4.6. Draft Scoring Rubric Associated with Grade 5 Assessment task (with Comments)

Target Level on Gust Learning Trajectory: Level 2 (Grade 2 Upper Anchor)

Rubric:

3 - Both number answers are correct. Both explanations include specific place value language (hundreds, tens, ones, value, worth, digit), and why a digit would go in a certain place. For example, "I put the 7 in the hundreds place because it is worth the most" or "The smallest digit would go in the ones place to make the smallest number"

2 – Both number answers are correct. Explanations lacking detail, vocabulary or use circular logic to explain. For example, "762 is the largest number because it is the biggest" or "7 goes in the hundreds, the 6 in the tens, the 2 in the ones" (but does not explain why…)

1 – Both number answers are correct, but there are no explanations (or explanations that do not make sense.) OR One of the number answers is correct, with a detailed explanation.

0 – Neither of the number answers are correct, no explanations.

Comments:

- We thought that this task adequately represented the goal in Level 2. In particular, "Understand that the three digits of a three-digit number represent amounts of hundreds, tens, and ones."
- We thought it was important to have the students explain their thinking in order to show true understanding of place value concepts.
- To make this task more or less difficult, you could increase or decrease the number of digits being used.
- With the rubric, we struggled somewhat with the 2-point answer (and a little with the 1-point) and feel that once we see student work we will be able to modify and improve the rubric.
- We anticipate that many students will be able to create the largest and smallest number, but will not be able to adequately explain their thinking.
- Currently, students are doing this in an independent station (without the explanation component), but they have not yet done number ordering as a whole group.

At the end of the session, each grade level had at least one sample task and a draft scoring rubric. In between Sessions 2 and 3, teachers administered their grade-level task to their students, as well as the task for the grade above and/or below theirs. They then used their scoring rubrics to score the work and provided us with a copy both of the unscored work and a copy of scored work so we could review how they had applied their rubrics in preparation for score moderation activities in session 3. Prior to session 3 we had picked up and reviewed student responses from all the assessment tasks that had been administered. We developed packets of student work for each teacher from assessment responses from their students. In creating the packets, we divided student work into groups of students who displayed the same strategy or level of understanding in responding to an assessment task. For example, the 5th grade task asked students to multiply a decimal by a 10. Students were divided into groups based on their strategies: students who wrote

the decimal 10 times and added it; students who used multiplication and treated the decimal as a whole number; students who used multiplication with the decimal; and students who simply moved the decimal point. These categories were determined based on examining student responses, rather than theorized or hypothesized ahead of time.

During the session, we distributed the pre-prepared packets of student work to each teacher, then asked teachers to work in grade-level teams to review the work, and use this to scrutinize distinctions among the different levels of student sophistication as operationalized by their LP. We asked teachers to review their packets of student work with their initially drafted scoring rubrics. They were to apply the original scoring rubric to each individual example of student work as a grade level team, determining if each teacher would apply the rubric in the same way and finding the range of scores generated for a single example of student work. We then asked the grade-level teams to compare the student work for groups of students with the same overall score to consider whether they thought all the students with the same score were demonstrating the same level of mastery and sophistication. We reminded them that the goal of scoring the assessments was to facilitate mapping a student back onto the LP, so scores should indicate their LP positioning. For example, if a student using sophisticated strategies who had gotten one or two final answers incorrect had the same overall score as a student who had used rudimentary strategies, that might suggest a problem with the scoring rubric.

In reviewing student work, we noticed a strong tension throughout many of the grade-level teams between scoring answers for correctness and scoring answers for the sophistication of the strategies used. For example, a 5th grader asked to multiply a decimal by 10 might get the right answer by writing a column of the decimal 10 times and using addition. Even if that student got the right answer, they would be using a less sophisticated strategy for multiplying decimals than a student who attempted to move the decimal place and moved it in the wrong direction, thus getting the wrong answer. Teachers had difficulty separating their evaluation of the sophistication of the strategy the student employed in answering the problem from the correctness of their final answer. While an advanced student would both use a sophisticated technique and arrive at the correct answer, we felt that placing students on the learning trajectory, especially in gradations of the messy middle, prioritized focus on the sophistication of their strategy rather than solely whether their final answer was correct. Emphasis on the sophistication of strategies employed by students makes stronger connections between student work and the LP.

As students gain more advanced understanding of a concept, they are able to apply more sophisticated reasoning in order to complete tasks. In adopting an LP approach, it should sometimes be the case that just because two students are able to come to the same solution for a task it does not necessarily mean that they have equal levels of conceptual understanding. Emphasis on sophistication of strategies used helps to identify these differences among students.

This tension between evaluating for correctness and evaluating for sophistication created problems for teachers attempting to map scores on student work back to the LP. As teachers tried to align scores from their initial drafts of the scoring rubrics back to LP, they realized that significant changes in the scoring rubric were necessary. Working with student work and trying to do this mapping seemed to be a productive task for illustrating the difference in scoring an assessment for traditional evaluations of mastery from scoring an assessment for evaluating placement on the LP. Engaging in the activity made the difference clear to teachers in a way that abstract discussions had not.

Development of Tasks for End of Course Assessment

Unfortunately, because so much time was devoted to establishing the refining the multi grade LP and its relationship to course-specific trajectories, almost all development of tasks for each grade's end of course assessment took place outside the context of our PD sessions. Here, Blue teachers opted for a standard process for all assessments administered to students in 1st grade and higher. Their agreed-upon goal for assessment development was to have a 10-question assessment task to determine if students had met the end of grade level anchor on the LT. The goal of this assessment was to determine if a student had sufficient proficiency "pass" the end of grade level target. We pointed out that to accurately measure growth, teachers would also need to measure student placement between grade level targets, at the "0.5" levels of the messy middle. The teachers had not discussed assessment tasks for placing students at the 0.5 levels of the messy middle.

The final format was determined for the end-of-year assessment in our final PD session, which would evaluate both whether a student had reached the end-of-year grade level anchor and also whether they had met the partial understanding midpoint of the grade level:

- One set of 5 question pairs focused on determining if the student had proficiency at the 0.5 level. The first portion of each question would be a multiple choice question

and the second component would be constructed response. This might either ask a student to explain their multiple choice response or be associated with the multiple choice question. Each question component would be worth 5 points, for a total of 10 points per question or 50 points for the entire section.

- A second set of 5 question pairs focused on determining if the student had proficiency at the end-of-year grade level. These 5 questions would also be pairs of multiple choice and constructed response questions. Scoring would be the same as the first half of the test.

A consistent theme and priority for Blue teachers was ensuring consistency of assessment tasks, rubrics, scoring, and percentages across grade levels. Teachers expressed concern that without this focus on consistency, teachers in adjoining grades might differ in their assessment of a student's location on the learning trajectory. The principal also raised concerns that differing formats or scoring might mean that student variation in scores across grades would be because of different test forms, rather than true differences in student ability. Although the value of consistency made them want a common rubric, they decided that the differences in content meant that each grade level would have to develop their own rubric. However, they decided to score each assessment consistently and use common cut points for determining whether a student had displayed proficiency at that level. There was also significant discussion of whether the test should consist of multiple choice questions, constructed response questions, or a mixture of the two. Some teachers favored constructed response questions that asked students to explain their previous multiple choice answers, while others preferred constructed response questions associated with the previous multiple choice question but asking them to do more than simply explain their multiple choice response.

Scoring and Interpreting Student Work

In the scoring and interpretation of student assessment responses, Blue teachers favored an approach reminiscent of the way that they had scored SGOs and interim assessments in the past. In a nutshell:

1. The tasks on each grade-specific assessment (usually 10 two-part items) were given weights such that the total score possible was always 100 points.

2. Uniform threshold were then established to translate the percentage of total points earned into ratings of "Advanced", "Proficient," "Partially Proficient" and "Unsatisfactory."
    - 90% and above = Advanced
    - 70% to 89% = Proficient
    - 50% to 70% = Partially Proficient
    - 0% to 49% = Unsatisfactory
3. These proficiency ratings were then translated into LP levels according to the following conversions.
    - Proficient or Advanced = Upper Anchor of Course-Specific Trajectory
    - Partially Proficient = Messy Middle of Course-Specific Trajectory
    - Unsatisfactory = Upper Anchor of Course-Specific Trajectory

While this approach has the benefit of simplicity, it also comes with a number of problems. First, there is no substantive justification for the demarcations being made in terms of proficiency levels relative to the number of points earned. The connections between proficiency levels and criterion-referenced levels of the place value LP are unclear. For example, if teachers in one grade happened to write 10 tasks that were significantly harder or easier than the tasks written by teachers in the adjacent grades, then a 70% could mean different things about student understanding of place value. Second, by combining the proficient and advanced categories to designate a student at the upper anchor of the LP, an overly optimistic picture of student mastery of place value may be portrayed. Indeed, when we examined results from student assessments at each grade, it was typical to find that more that 90% of each teacher's students were designated as having mastered the upper anchor of the grade-specific trajectory. Although this homogeneity of performance is possible, it is an open question whether it would be validated by an externally administered assessment.

Rating Teachers on the Basis of Student Growth

Those teachers who had advocated for the ruler metaphor with regard to the levels of the place value LP also felt that in scoring student growth, equal value should be placed on transitions from one level to the next. Their primary concerns were simplicity, which would both help existing teachers administer and implement the scoring system and would make it easier to teach the system to new teachers arriving at the school. Blue's principal also supported assigning

equal weight to each step, arguing not only for the simplicity benefits, but also expressing her concern that unequal weight might provide unwanted incentives to teachers to focus more instructional time and resources on some students whose growth would be worth more. On the other hand, almost all of the lower grade level teams – K, 1st grade, 2nd grade, and 3rd grade – felt that moving from level to level (at least at the grain size of the current LP) was not of uniform difficulty. For example, 2nd grade teachers discussed how moving a student to upper anchor of their course trajectory, which involves not only getting questions correct but using the most sophisticated techniques and strategies to do so, was often more difficult than teaching a student to get the questions correct. Because of that difference in teaching difficulty, they felt that moving a student to the upper anchor should be accorded more points than moving from the aspirational lower anchor to the messy middle. Ultimately , teachers opted for simplicity and avoiding unintended consequences.

The upshot of this decision was that for every grade, movement from one level of the LP to the next constituted 0.5 points. Hence "a year's growth" or aspirational growth was represented by a student that has moved two levels, earning 1 point. In debating the final category table that would be used to convert student growth scores into a teacher's SLO rating, there were two alternatives. The first was based on a category tables approach along the lines of that described in chapter 2 and taken by Spruce and GB:

Teacher Rating	% Students not making minimum growth (<0.5)	% Student making minimum growth (>0.5 and <1)	% Students meeting or exceeding target growth (>=1)
1 (Lowest)	>= 50%		<50%
2 (SLO Credit for Pilot)	25%	25%	50%
3 (SLO Credit Long-term)	10%	15%	75%
4 (Highest)	5%	10%	85%

The second proposal, developed by the upper grade Blue teachers, involved taking an average of the growth for all students taught by a teacher and applying the following table:

Teacher Rating	Average Growth of Students
1 (Lowest)	Below 0.5
2 (SLO Credit for Pilot)	0.5 to 0.9
3 (SLO Credit Long-term)	0.9 to 1.2
4 (Highest)	1.2 or above

The teachers argued that the averaging system had several advantages in regards to high performing students. They pointed out that in the percentage table, teachers received no additional credit for student growth beyond one grade level whereas in the averaging system, a student who achieved 1.5 grade levels of growth would offset or compensate for students with low grade levels. This was characterized by another teacher as an equity issue, because the percentage table provided no incentive for teachers to focus on students who had already made one grade level of growth. A consistent theme among teacher comments was concern that this accountability system not drive instruction. Teachers were very aware of the incentives inherent in the two different category table proposals and how those might impact their day to day instructional decisions.

There were also voices in support of the percentage table. One teacher argued that the categorizations in the percentage table were more individualized and allowed for more focus on individual students than the group averaging in the averaging table. The principal pointed out that the category table more closely approximated existing district approaches to accountability, where schools are being assessed by the proportion of students in certain proficiency bands, rather than by an overall average student achievement. Through a silent vote, the teachers ultimately selected the percentage table.

Once teachers selected the percentage category table we turned to adjusting or revising the numbers in the table. First, teachers felt that there was no reason to separate students making less than one year of growth into two categories. Teachers felt strongly that every student should be expected to make at least one year of growth and that differentiating between making over or under a half year of growth was unimportant and distracting. One teacher suggested that students who were further behind may be behind because they had problems making one year of growth, but other teachers disagreed, arguing that every student could be expected to make one year of

growth. Also, teachers felt that the percentage distributions between the 0 to 0.5 and 0.5 to 1 units of growth categories did not make sense for the number of students they had. With a class of only 30 students, the difference between 5% and 10% of students making below minimum growth might be only 1 student, or even a fraction of a student. They did not feel like these allowances were meaningful to them.

The principal embraced this suggestion of collapsing those two categories for administrative reasons, suggesting it would be easier for both teachers and for her to track information in only one column (students making at least one year of growth) than a full category table. After significant discussion and suggestions from the principal, the final table adopted by teachers was as follows:

Teacher Rating	% Students making at least 1 year of growth
1 (Lowest)	Under 50%
2 (SLO Credit for Pilot)	50%
3 (SLO Credit Long-term)	85%
4 (Highest)	90%

We pointed out that these requirements were quite stringent relative to the rigorous criteria of the LP. Teachers dismissed this concern, reiterating their expectation that every student make at least one year of growth and their confidence that they would be able to ensure that every student in their classroom would make that growth.

Concerns about Placing Students into Retroactive Baseline Levels

Blue teachers were asked to place their students into baseline levels of their course-specific trajectories in early January on the basis of a draft assessment task that had been administered in December, and their own recollections of their students' performance at the start of the academic year. We noted that this was an artificial exercise because we had not yet established an LP or assessment tasks at the start of the school year. We stressed that safeguards would be built in to ensure that they would not be penalized if they were underestimating student growth. Ultimately however, even though Blue teachers went through the process of

establishing baseline levels, with the exception of Kindergarten teachers they were unwilling to use these baselines to compute student growth points that could then be used for teacher ratings according to the table above. Nonetheless, given that almost all students were designated as achieving the upper anchor of each teacher's course-specific trajectory, unless a significant proportion of Blue students were already at the messy middle at the start of the school year (i.e., level 1.5 instead of 1.0), it is highly likely that most Blue teachers would have received a rating of a 3 or 4. But as mentioned earlier, it could be argued that these ratings could be inflated because of the way student assessment scores were mapped onto LP levels.

Teacher Reactions and Feedback

Concerns about Learning Progression

Teachers seemed primarily to be worried that there would be student growth that would not be captured in their SLO measurement, underestimating their actual growth achievements and potentially penalizing them in their evaluations. In the beginning of the year, we had not developed the point systems by which student growth would be converted to teacher evaluation ratings, so it was difficult to convince teachers that failing to wholly capture the growth of one or two outlier students would not have a significant effect – or even a measurable effect – on their teacher evaluation ratings. We hoped that when we showed them the conversion calculations in the future, we could demonstrate to them that these one or two students whose growth would not be wholly captured by a teacher's course-specific trajectory would not significantly affect their evaluation scores.

Additionally, teachers expressed concern about the accountability impact versus the instructional impact. Several teachers raised the concern that the growth of some students would count less, indicating that fewer instructional resources should be devoted to that individual student. We encouraged teachers to see that it was still instructionally meaningful to create academic growth in each individual student, even if the growth of one student would not be perfectly captured by the SLO system. Teachers seemed receptive to that differentiation, but several raised the concern that if they knew the growth of an individual student would not be captured in this SLO growth framework, it would, consciously or unconsciously, motivate them to spend less time and energy on that student.

Teachers were concerned about the practical issues of writing an SLO keyed to an LP level when the level had a number of bullet points or components. Teachers were also concerned that writing a single growth objective that captured all 6 components of an upper anchor would be unwieldy or even impossible. Another concern about the place value LP was how to incorporate the idea of math practices. Some levels of the macro LT included statements that students should "understand" how to do specific math strategies, which teachers felt was an explicit and inherent incorporation of the idea of math practices. Other levels of the LP, however, did not include "understand" statements and did not have any explicit or implicit suggestion that math practices were an essential part of mastering that level. We suggested two possible refinements to teachers. First, they could revise the LP to more explicitly incorporate the need for students to understand and use relevant math practices in executing the relevant subject matter. Second, they use their second SLO to focus exclusively on math practices and articulate their expectations for student math practices.

A last concern about the LP concerned the scope of the LP and associated SLO. Teachers expressed concern that the SLO and focus of the LP might limit the focus and scope of standards. Because the LP focuses only on base 10 concepts and does not cover the full scope of the relevant math standards for each grade level, the teachers were concerned that growth in other areas would not be captured by the SLO process. They were also concerned that the SLO process might encourage more focus on base 10 concepts than on other aspects of math standards. Similarly, they were concerned that tasks evaluating growth on the course-specific trajectory did not cover the full range and scope of math standards for each grade.

Scoring Student Growth

Teachers also had many questions when allocating point values for student shifts between levels of the LP. They again focused on the outlier students, wondering how the point system would measure growth for a student who entered the grade already very high on the course-specific trajectory and who had little additional room for growth before reaching the target upper anchor. They also asked about students who entered very low on the course-specific trajectory and, even with significant growth, did not progress beyond the lowest level. To address this perception that teachers needed to get the highest amount of points for each individual student in order to obtain a satisfactory teacher evaluation categorization, we discussed how approach taken

to rate teachers for growth would group students and take into account that some students might have very low levels of growth, even if growth for most students was exceptional.

Opportunities for Collaboration and Vertical Alignment

Blue teachers were the most uniformly enthusiastic about the opportunity that the PD sessions presented for them to collaborate with teachers in the same grade level on writing student assessment and interpreting the results, and to align their assessment practices vertically with teachers in adjacent grades. There was, however a flip side to this enthusiasm in that they were skeptical that this new model of SLO development could be scaled out without a comparable investment in PD resources for teachers at all schools.

Chapter V: Results by Math Pilot Site - Spruce Middle School

Using a Learning Progression to Define an SLO at Spruce

The Spruce teachers took up our sample LP on Proportional Reasoning for several reasons. First, Ratios & Proportional Relationships is one of the major domains of math content emphasized in both the 6th and 7th-grade Common Core State Standards in Mathematics (CCSS-M 6th & 7th-grade RP.1, 2, & 3) and in the DPS scope and sequence documents. This topic encompasses many important life-applications such as mark-ups and discounts, sales tax, and measurement conversions. Further, there was significant discussion at the site about how 8th-grade builds on the concept of proportional relationships in slope when dealing with linear relationships and graphing linear equations (CCSS-M 8.EE.B.5). The team agreed that this LP covered a significant portion of the curriculum across all grades at the site and was important for conceptual understanding of several key areas in mathematical understanding (e.g. rates, unit rates, ratios, scale drawings, making predictions).

Figure 5.1 Anchors from CCSS-M for Proportional Reasoning

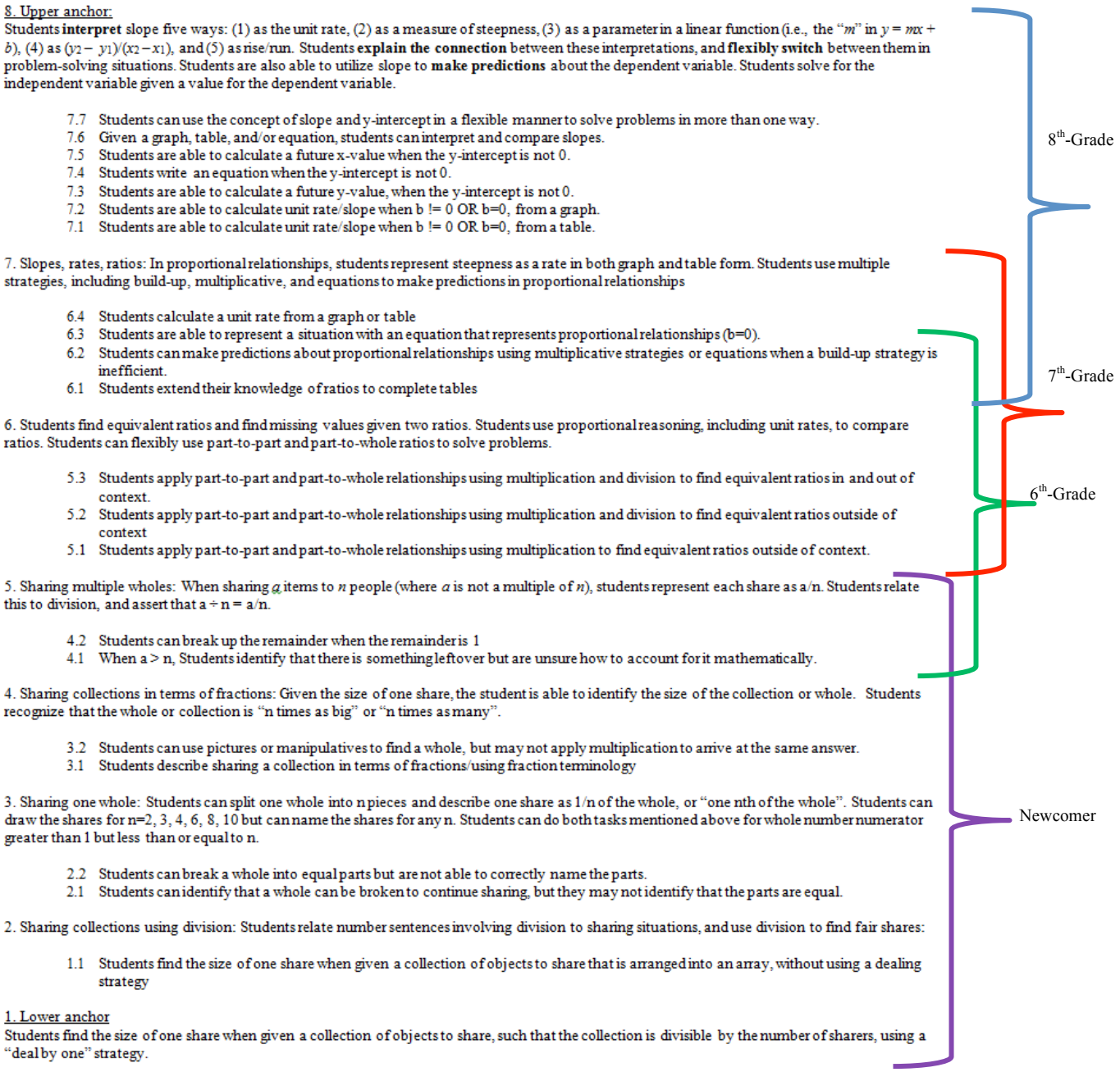
Anchor	Level	Level Descriptor
Upper anchor: End of 8th-Grade	4	Graph proportional relationships, interpreting the unit rate as the slope of the graph. Compare two different proportional relationships represented in different ways
		Partial mastery of 8th-grade Content ("Messy Middle")
End of 7th-Grade	3	Compute unit rates associated with ratios of fractions, including ratios of lengths, areas and other quantities measured in like or different units. Recognize and represent proportional relationships between quantities. Use proportional relationships to solve multistep ratio and percent problems.
		Partial mastery of 7th-grade Content ("Messy Middle")
End of 6th-Grade	2	Understand the concept of a ratio and use ratio language to describe a ratio relationship between two quantities. Understand the concept of a unit rate a/b associated with a ratio a:b with $b \neq 0$, and use rate language in the context of a ratio relationship. Use ratio and rate reasoning to solve real-world and mathematical problems, e.g., by reasoning about tables of equivalent ratios, tape diagrams, double number line diagrams, or equations.
		Partial mastery of 6th-grade Content ("Messy Middle")
End of 5th-Grade		Students can relate sharing and division to fractional parts of a whole or multiple wholes.
		Partial mastery of 5th-grade Content ("Messy Middle")
Realistic lower anchor	1	Students have some understanding of sharing one whole or sharing a collection among a group. These skills are used to develop understanding of making equal groups.

With an LP already in place, the teachers at Spruce worked together to revise the details of the LP. This was first done by discussing the pre-existing levels as we proposed them. Teachers made sure they understood all of the language and made modifications to match teaching on their campus. They also added and deleted levels to represent their own emphasis regarding the topic. Revision of the LP continued through four of the sessions. Teachers made modifications not only based on their own teaching experience and expectations of student

learning, but after student work was collected, teachers continued to adjust the LP to better reflect evidence of student learning as seen in student responses to items developed through the SLO process. A strength of this LP is that it covers the lowest level of math ability at Spruce all the way through completion of 8th-grade, and all teachers involved in this curriculum worked together to develop it. The final LP used at Spruce Middle School as well as the course-specific trajectories identified for each teacher is given in Figure 5.2.

A few important aspects of the LP are evident. First, note that each course-level trajectory spans multiple grades and reflects the general process discussed in Section 2. The bottom of each course-specific trajectory reaches down to what teachers identified as their "realistic lower-anchor" for the level of ability they tend to see in their classrooms each year. For each teacher at Spruce, this is about one grade below the "aspirational lower-anchor", the level students are expected to enter a class if they show grade-level mastery of the content. Each teacher except for Dave defined the "upper anchor" at the appropriate grade-level. Dave decided to set his goal slightly higher and into the 7th-grade curriculum because he had students at the beginning of the year who had already mastered the 6th-grade standards on proportional reasoning.

Figure 5.2 Spruce Final LP and Course-Specific Learning Trajectories



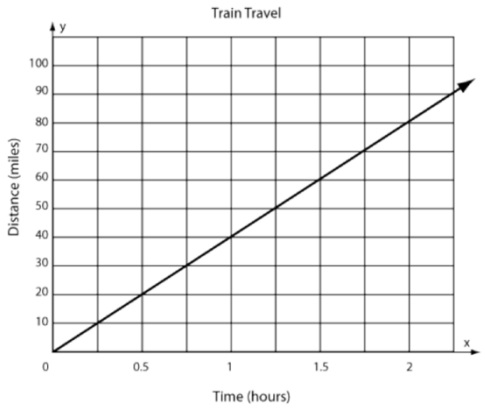
Choosing or Writing Items for Student Assessment at Spruce

In order to elicit information about where student understanding was reflected along this LP, teachers needed to create tasks specific to the progression. In order to facilitate this process, teachers were first given a collection of sample tasks collected from a variety of sources (e.g. Partnership for Assessment of Readiness for College and Careers, Smarter Balanced Assessment Consortium). The teachers generally used these tasks as inspiration for developing their own tasks tailored to the LP. However, Cora did take two items presented to her and made small changes to adapt to the LP. The group worked together to make the changes to these tasks and develop the rubric. The team decided that both tasks were needed to address all elements of the 7th-grade upper anchor and that it was perhaps possible to use one rubric to grade both tasks. Efforts were made to create a rubric that defined differing levels in the sophistication of student learning and understanding rather than simply a checklist of discrete skills. These items as first presented to students and the associated rubric are given in Figure 5.3.

Figure 5.3 7th-grade Tasks and Rubric After Session 2

7th-Grade Sample tasks & rubric

1) Use the graph below to answer the following questions:



a. How many miles does the train travel each hour?

b. Write an equation you could use to find the distance a train travels in any amount of time.

c. Use your equation to predict how many miles the train travels in 12 hours.

2) Reynoldo is planning to drive from New York to San Francisco in his car. Reynoldo started to fill out the table below showing how far in miles he can travel for each gallon of gas he uses.

Gallons	2	4		8	10	12
Miles	56		168	224		

Use the information in Reynoldo's table to answer the questions below.

a. Complete the table for Reynoldo. Assume the relationship in the table is proportional.
b. Based on the table, how many miles per gallon did Reynoldo's car get? Explain your reasoning in words.
c. Write an equation that Reynoldo can use to find the distance (d) he can drive on any number of gallons of gas (g).
d. When Reynoldo's tank is full, it holds 23 gallons. How far can Reynoldo drive on a full tank of gas?

Level	Description
3	Student can interpret rate of change from both a graph and a table. Student uses build-up or multiplicative strategies and write and use equations to make predictions in proportional relationships.
2	Student can interpret rate of change from both graphs and a tables. Student use either the build-up or multiplicative strategy. Also, students can use equations to make predictions in proportional relationships.
1	Students understand ratios only as a build up strategy. Students can find a unit rate from either a graph or a table.
0	Blank or no related work

In order to develop the levels of the rubric, the team started with the extreme values. A very broad descriptor was given for level 0, which was largely tied to thinking about scoring for the purpose of giving grades rather than eliciting information about where a student might fall on the LP. The description for level 3 was essentially a modified description of the upper anchor. In order to create descriptions for levels 1 & 2, teachers discussed common errors students make on tasks like these as well as what differences could exist in student understanding that would be shown in these tasks (i.e. a student could use either a graph or a table but not both, OR students could use build-up strategies but not generalize to a formula). Because of limited time, there was not much opportunity to really discuss the quality of this rubric and if it would give good information regarding where students fell on the LP. However, it was expressed to teachers that using these tasks and rubrics would provide a starting point for developing more refined tasks and rubrics and that looking at student work would help with continuing to refine the LT. Additionally, these tasks would elicit some information about what students could do with relation to the LP.

While most of the work at Spruce was done with all teachers in collaboration with one another, the team ran out of time to develop tasks and rubrics for all grades together as a group. As a result, the other teachers had to develop these tasks and rubrics primarily on their own afterward. The tasks were then administered to students and score moderation occurred at the next training session. Ideally, each task was to be administered by two different teachers so that teachers could discuss grading with one another. However, the fact that Spruce has only one teacher per grade level limits this to some extent. The tasks designed for 8^{th}-grade were only given by Mike as none of the other teachers had students approaching this level on the LP. Additionally, only Lisa gave her tasks because of some miscommunication with Dave.

When they first created their grade-specific tasks, the teachers felt they could use one rubric to score two tasks that each had multiple parts. After reviewing student work samples on these tasks, the teachers decided that one rubric for two tasks did not provide enough information about differences in student work, so they created additional rubrics. These new rubrics had revised and more specific level descriptors and were directly aligned to levels on the LP. This was to help in classifying students and understanding student growth. The goal was to make explicit connections between the tasks and the LP. The updated rubrics for Cora's tasks after this session are provided in Figure 5.4.

Figure 5.4 Revised 7th-grade rubrics after Session 3

Task 1

Score	Description	Level(s) in LT
4	Student can calculate the unit rate from the graph. Can write an equation and find the distance at 12 hours by either using a multiplicative strategy or the equation.	6.4
3	Student can interpret rate of change from the graph. Student use the multiplicative strategy or equation to find the distance at 12 hours. Also, students could not write an equation to describe the relationship.	6.3
2	Students can find the distance at 12 hours using a multiplicative strategy of use the equation.	6.2
1	Can only find the distance at 12 hours using a build up strategy. This is akin to creating a table to solve the problem.	6.1
0	Blank or no related work.	6.0 or below

Task 2

Score	Description	Level(s) in LT
4	Student can calculate the unit rate from the table. Can write an equation and find the distance that could be driven on 23 gallons by either using a multiplicative strategy or the equation. Also, they complete the table correctly.	6.4
3	Student can calculate the unit rate from the table. Student use the multiplicative strategy or equation to find the distance for 23 gallons. Also, students could not write an equation to describe the relationship. Also, they complete the table.	6.3
2	Students can find the distance for 23 gallons and complete the table.	6.2
1	Students complete the table.	6.1
0	Blank or no relevant work.	6.0 or below

As a first step, the teachers were given 3 examples of student work for each of the tasks the teachers had previously developed. Teachers were asked to individually score this work with the rubrics created in the previous session. Afterward, teachers compared and discussed differences in assigned scores. Teachers quickly realized that the rubrics were not specific

enough to score a variety of student responses. In order to help with this issue, we asked the group to add a column to each rubric that identified where student work at that level would map to the LP. This process helped in three ways: (1) Teachers added and deleted content in the LP so that it better reflected elements they emphasized at their site, (2) student misconceptions in understanding were used to refine the level descriptions in the rubrics themselves, and (3) teachers became aware of elements in their tasks that did not actually align to the LP.

Teachers commented that further refining the LP was very much helped by having spent time looking at student work. This enabled them to have a better understanding of how students progressed in their learning with respect to proportional reasoning. Teachers also commented on how this process helped them to develop skills for writing new tasks in the future. They were more attuned to writing items that could elicit information about multiple levels on the LP rather than items that essentially served as a binary checklist for lists of discrete skills. Notice that the new rubrics are more specific to the individual tasks. Additionally, they were aware of how elements of a task that addressed other LPs would distract from assessing where students were relative to Proportional Reasoning skills.

The new level descriptors provide detailed information about what would need to be seen in student work to represent more sophisticated levels of understanding on the LP. For example, in Task 2, a student who receives a score of "1" can complete the table but they are unable to extend their knowledge to a value not in the table that is also not a multiple of another number in the table. A student who scores a "2" can extend knowledge to find the correct value outside of the table, but is not able to generate a formula or use unit rate to make predictions for any value. In order to score a "3", the student needed to show facility in finding the unit rate of the problem, and a score of "4" was reserved for students who could generate a formula and use it to make predictions about any value. This parallels the LP in discussing a student's growing sophistication in learning to develop and use a build-up strategy, a multiplicative strategy, the unit rate, and an equation in order to make predictions about proportional relationships.

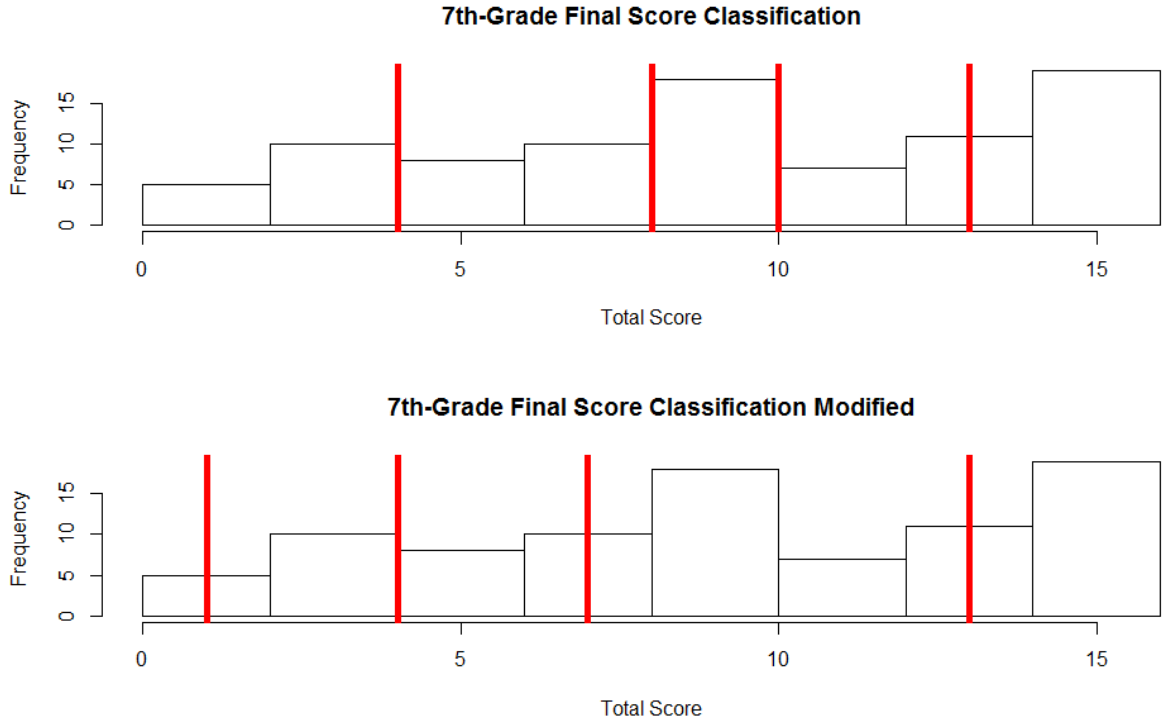
At the end of the SLO process, the Spruce teachers developed 32 constructed-response tasks in total with individual rubrics. In order to create final assessments, teachers chose the items from this bank that covered the skill levels they expected to see among their students. For Mike and Cora, this included tasks for their respective grades and the grade below. Dave used only 6th-grade tasks since he did not originally give any tasks aimed below 6th-grade. This

limited him in his ability to show growth for students that may have started below grade-level. Lisa only gave her tasks as any students who showed enough ability to answer grade-level tasks are exited from her classroom and into Dave's classroom throughout the year.

Scoring and Interpreting Student Work at Spruce

Linking tasks to the LP is an incomplete exercise without a way to translate scores on these tasks such that they provide information about student growth on the LP. As noted in Section 2 pg. 23, it is common for teachers to sum up total points on an assessment and express points earned as a percent of the total correct. This sort of scoring is not meaningful with respect to our LP. Instead, most of the teachers at Spruce took up the process developed in Section 2 of this report. They first examined the distribution of student scores and divided the distribution into equally divided bins. The number of bins varied by teacher and depended on the number of levels specified in each course-specific trajectory (4 for Dave; 5 each for Mike and Cora). This was the starting point for grouping students who were likely to be relatively similar in their overall performance. Next, Dave and Cora examined the specific item responses typical of students in each group and considered making adjustments to cut scores based on the level of understanding represented in each collection of student work. Mike chose to randomly select a handful of students from each bin and identify if he felt each was appropriately categorized or not. For each teacher, the goal was to identify if students in the top performing bin had demonstrated the sort of sophistication in their responses akin to what would be expected someone at the top level of the course-specific trajectory. This process was repeated for each bin. Both Dave and Cora made adjustments to the cut scores for their bins, but Mike felt that his were sufficient as originally defined. Figure 5.7 shows Cora's student score distribution along with the original and revised cut scores.

Figure 5.5 Initial and Modified Score Bands for Final 7th-Grade Student Classification



Cora's original bin definitions had all students scoring below 4 points as showing understanding at the level of the realistic lower-anchor. When she looked at student work, she found that students scoring from 2-4 points on her collection of tasks generally showed evidence of understanding Proportional Relationships at a level higher than the realistic lower-anchor, but still not to the level of the aspirational lower anchor. She also noted that students scoring from 5-7 points were demonstrating knowledge at the aspirational-lower anchor, but students scoring from 8-13 points were showing understanding just below the aspirational upper anchor.

Although Cora's process reflects what was described in Section 2 and what was generally followed by Dave and Mike, Lisa was in a different situation. Lisa had only 13 students in her class, but also had a course-specific trajectory that covered 10 levels. Her approach was to instead order her student work from lowest to highest score and then group and classify students holistically based on the sophistication of understanding evident in their work. This process was reasonable for her because of the small class size and range of ability level to which she must address in her class.

Rating Teachers on the Basis of Student Growth at Spruce

The process of developing bin ranges and identifying student levels on the LP is how we suggest students should be assigned at both the beginning and end of the SLO process. In order to calculate growth along the LP and rate teachers, there must be some quantification of movement along the LP. The simplest application of this process would be to assign 1-point to each level of growth on the LP. Section 2 of this report defines Minimal Growth as students showing growth of just one level on the LP. This is the minimum amount of growth in student understanding we can hope to capture based on our model of cognition. The hope is that the vast majority of (if not every) student show at least this level of growth within a year. Aspirational Growth was defined as that growth expected for a student who enters at the appropriate grade-level understanding and grows to the aspirational upper anchor. If using the very basic LP as defined in Section 2, each grade would then have an expected growth of 2-points per year (one point to move from the aspirational lower-anchor to the messy middle and then another point to move from the messy middle to the upper anchor). Exceeding Aspirational Growth would then be any student showing more than one year's worth of growth on the LP, or movement of at least 3-points in our simplest approach. Teacher classification then occurs based on the percentage of students who make below minimal, minimal, aspirational, or above aspirational growth in the year. When 1-point is assigned to each level of growth on the LP, this is a fairly straightforward process and relatively easy to then translate student growth scores into teacher classifications. A sample of how this could be done was presented in Section 2 and repeated in Table 5.2.

Table 5.1 Categorizations of Teachers for Student SLO Growth

Category	%Students Not Making Minimal Growth	% Students Making Minimal Growth	% Students Making Aspirational Growth	% Students Exceeding Aspirational Growth
4 (Higher)	10%	15%	75% Combined	
3	15%	25%	60% Combined	
2	25%	50%	25% Combined	
1 (Lower)	50% or more	Less than 50% Combined		

Note: Percentages in Cells Are All Hypothetical

However, the teachers at Spruce actually decided to assign from 1-2 points for movement between adjacent LP levels based on the perceived difficulty of getting students to make that growth in understanding. This complicated the final teacher classification process in a few ways. First, calculating student growth at the end of the year was not as simple and subtracting the final level from the beginning level. Instead, care had to be taken to address how many points were assigned to the levels reflected in the student's final and initial placements. Additionally, this made the distribution of growth points for students more difficult to classify as those making minimal growth versus those making expected growth. That is, if minimal growth was 1 point and aspirational growth was 3 points, it was unclear how a student who grew 2 points should be classified. As a result, when analyzing data from Spruce, the decision was made to assign only 1-point to each level of growth on the LP. Figure 5.8 shows the overall distribution of student growth at Spruce, and Figure 5.9 shows the distribution of student growth levels by teacher.

Figure 5.6 Distribution of student growth levels for all Spruce students.

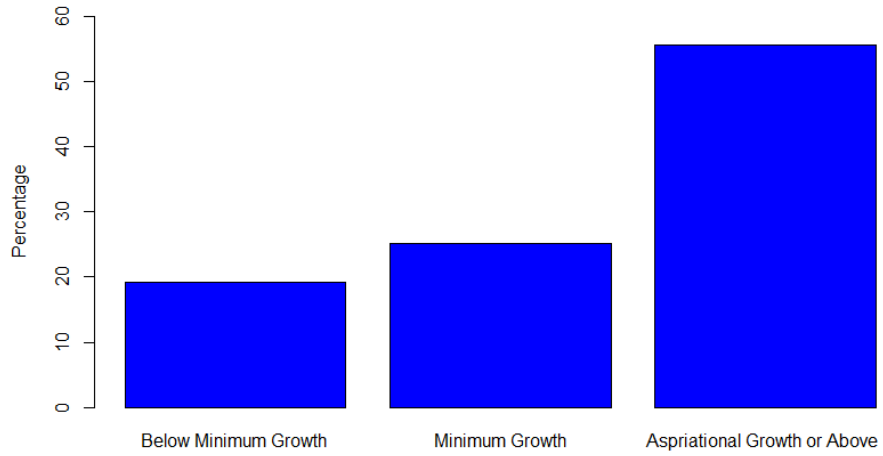


Figure 5.7 Distribution of student growth levels by teacher at Spruce.

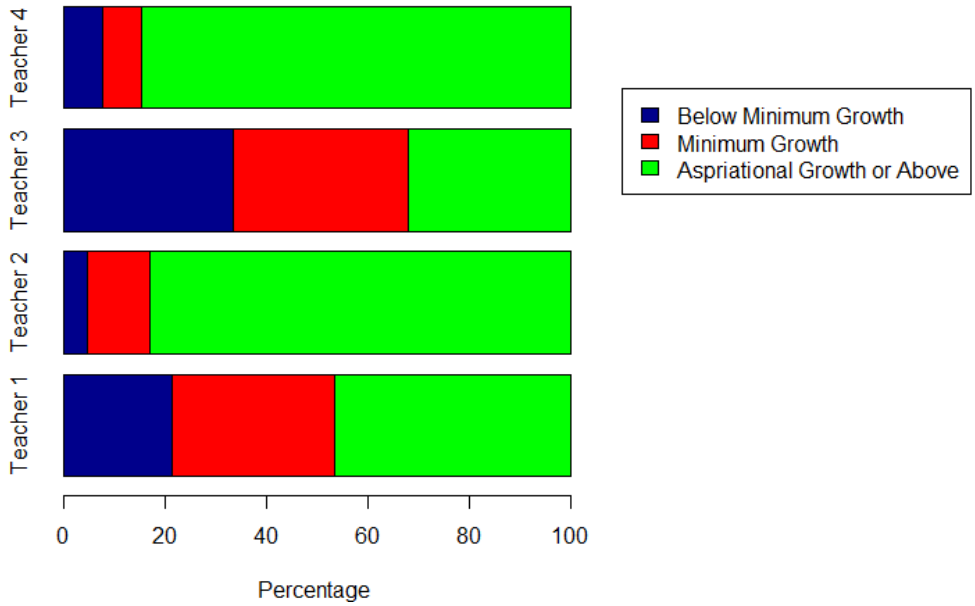


Table 5.2 provides a hypothesized method for classifying teachers based on their student growth. The teachers at Spruce were asked to design such a classification structure as well as an academic exercise (See Table 5.2). They knew this proposed classification structure would not be used for their evaluations during the pilot, but would instead be used to inform DPS when determining a similar classification structure in the years to come.

Table 5.2 Parameters for Teacher Categorizations

Category	%Students Not Making Minimal Growth	% Students Making Minimal Growth	% Students Making Aspirational Growth	% Students Exceeding Aspirational Growth	Narrative Description
4 (Higher)	< 5%		≥ 45% Combined		At least 45% of students showing aspirational growth and less than 5% showing below minimum growth
3	< 5%	< 75%	≥ 25% Combined		At least 25% of students showing aspirational growth and less than 5% showing below minimum growth
2	< 50%		≥ 50% Combined		At least 50% of students showing minimum growth
1 (Lower)	> 50%		< 50% Combined		More than 50% of students showing below minimum growth

Based on this classification structure and the distributions shown in Figures 5.8 & 5.9, all teachers at Spruce would be rated as a 2 because each of them had more than 5% of their students show less than minimal growth. In order to score a 3 or a 4 for teacher final classification, the Spruce teachers determined that no less than 5% of students should show less than minimum growth. Teacher 2 had 72% of her students showed a grade-level's worth of growth or more, but 6% of the students showed below minimum growth. Based on the way Spruce defined teacher classifications, Teacher 2 would be classified with a rating of "2" despite the high number of students with aspirational or higher growth. This is a notable example for district personnel to consider when creating final classification rules for teachers with the SLO process

Teacher Reactions and Feedback

Evaluation forms were distributed at Spruce after all but the first and last session. The first session was omitted as an oversight. The last session was purposely omitted as the session was essentially conducted as a focus group discussion to evaluate the entire SLO process. This is different from the process at the other sites where a computer-based survey was administered to acquire participant responses regarding the professional development experience.

Not a single written evaluation form at Spruce expressed negativity toward the SLO process. Teachers were generally always positive about the experience and appreciative of the time to work together in vertical teams. However, a recurring conversational theme throughout the year was skepticism regarding how development of Learning Progressions would occur in the future years when less time and training were devoted to the process.

Though teachers at Spruce were concerned with how rolling SLOs out to the entire district will go, they were largely supportive of this work and found it beneficial. Teachers commented multiple times that this work should be done in the time allocated for data teams on their campus because the information they gleaned from the SLO process was more informative than from their district benchmark results. Cora noted that at the beginning of the year she didn't understand how the SLO process was more than just a complicated SGO process, but at the end of the year saw the value in identifying student growth and felt the extra work was totally worth the benefits she gained in her classroom. The teacher noted that the process was especially motivating to some of her lower-performing students who previously would have simply given up because they would repeatedly be labeled as unsatisfactory. However, knowing they would receive some recognition for what growth they did make caused both the students and her to be more motivated. Because of this mindset, Cora said she saw much more growth in her lower-achieving students than previously and was surprised by how much growth they showed at the end of the process. During the final session when teachers presented their reflections on the process, all of the teachers noted the amount of work that was required, but also argued that the benefits to classroom instruction were worth the time and effort.

In discussion about how their thinking about teaching and assessment may have changed after being a part of this process, Mike mentioned that this method made him more aware of prerequisite skills that were necessary for students to be successful. He further commented that

this helped instruction as well as writing tasks that truly assessed what was pertinent to the trajectory at hand and not unrelated skills. He specifically stated: “I feel like I’m a pretty good teacher and have a good grasp of the work I do. This project made me more thoughtful about student misconceptions and where I need to start instruction.”

However, Mike also stated that he felt “these tasks were moving away from problem solving in general”. This particular teacher used to give a large, problem-solving task at the end of every unit that connected all relevant skills, but gave no specific direction on how to solve the problem. The teacher stated that “this last task was often too big for the kids.” Students would have to problem-solve to figure out what to do. He felt that the collection of tasks we gave afforded the students more direction on what exact skill was needed for each task and took away the need for greater problem-solving. Mike noted that he would be interested in doing an SLO or building an LP about problem solving, but he wouldn’t want it to count for his evaluation because there’s so much he doesn’t know about this process. All teachers stated they would be interested in research in this area as it was a valuable skill but one none of them felt comfortable assessing or placing on a trajectory.

Cora stated that the process made her more cognizant of what she was really trying to assess when creating tasks. Additionally, this process helped to isolate where students really were struggling and gave her ideas for how to move forward in instruction and/or remediation. Dave discussed having a more cognitive/developmental approach to planning and teaching. He was thinking more about where his students were at the current point and what the next steps should be in their learning. He noted that he’s not necessarily convinced that learning is linear in this way, but that it was a change in what he did in the classroom. This also helped teachers to think about how skills are related among LPs and things that need to be taught together or prior to one another. Lisa reported that figuring out the messy middle was very helpful for her. She’s always had a broad spectrum of skill to work with, and knew what the end goal was (6^{th}-grade standards), but was always a little fuzzy on how to move from where the students started to where they needed to be. The trajectory was very helpful in her understanding this for students at least with regard to this one trajectory.

In comparing SGOs to SLOs, the teachers reported that SGOs had virtually no collaboration or value for teacher-created artifacts. They noted that there was lots of teacher creation involved in the SLO process, and they appreciated this aspect of the project. Though

teachers always had the option to use their own assessments for SGOs, it seemed to them that it was preferable for them to use the assessments on the SGO website. Teachers made comments like the following:

- "I liked that this was more about creating something that you think really attaches to what you're teaching and being trusted that what we're doing is at least as valid as all those other tests out there. It was kind of nice."
- "If this is done well, it can fit more nicely into grade-level standards than some of the tests from the list of choices. There's really no problem with there being a list if the list is connected in the right way with what everyone is on the same page with in terms of standards, but there's been a disconnect."

Also, the teachers stated SGOs often compared growth on two totally unrelated assessments or tasks that didn't seem comparable to teachers. SGOs weren't always seen as a legitimate growth measure, while the SLO process seems to be much more so.

Teachers were overall very positive about the time spent in the pilot this year. No one identified any topics or activities that they felt were useless or should be cut from the process. They noted that it would have been helpful to start off with a completed example so that the end goal was more concrete at the beginning, but also acknowledged that since this was a pilot, such a thing did not exist at the onset of the project. They also stated that they would have liked for the language of the standards to make the grade-level positions on the trajectory (i.e. Aspirational Lower and Upper Anchor). They expressed that it would be good to give teachers a skeleton based on the standards and have them flesh out the messy middle or to give them pre-packaged trajectories if we have them.

Teachers generally stated that they liked being involved this year and that they would do it again next year, but that they don't necessarily think that this process was worth all of the work when related to ProComp bonuses. Essentially, the discussion surrounded how much of the teacher evaluation process this would ultimately be worth. If it's just tied to the bonus, the process is more work than it's worth. If it's only a small part of the teacher evaluation, still likely not worth the work, but if it's a substantial part, then the work may be more relevant. None of them are against the process in general and acknowledge that it can make them better teachers, but the payoff might not be enough for the level of commitment it requires. However, the teachers did see this work as valuable, just not valuable enough to be placed on top of other time

requirements they already have to fill. There was also concern about teachers in non-tested subjects. If the SLO process alone was going to be 50% of the evaluation, the teachers were concerned about the incentives there. This situation, in their opinion, would not be the incentive to do this work well and right but to just make sure the goals were met.

Chapter VI: Results by Math Pilot Site - Green Beechway High School

The teachers at GB chose to focus their work on algebraic manipulation. They chose this topic for three reasons: First, it is a topic that is present in all four years of math courses taught at GB. Second, the teachers considered it to be a foundational skill for much of the rest of the content taught at GB. Finally, the teachers reported that students at GB typically struggle with algebraic manipulation, and therefore the teachers were motivated to devote the year to in-depth consideration of the topic.

Using a Learning Progression to Define an SLO at GB

At GB, we began by creating an across-grade LP. First, teachers worked together to write the lower and upper anchor of the across-grade LP. To create the upper anchor (i.e., the algebraic skills that the teachers expect a graduating senior to have), teachers drew on both their aspirations for students and curriculum and standards documents including the Common Core State Standards for Mathematics (CCSSM) and district curricula. To create the lower anchor, teachers made an appraisal of the algebraic skills that students can generally demonstrate when they walk in to GB on the first day of ninth grade. That is, the lower anchor was written based on the teacher's perception of the reality at GB, rather than those skills that would be expected based on curriculum documents.

Having written the upper and lower anchors, the research team filled-in the end-of-year anchors for each course at GB (Algebra I, Geometry, Algebra II, and Pre-calculus). Our original plan was to develop these anchors collaboratively with teachers during the first session but we ran out of time in the session. To create the anchors, we drew on the CCSSM and the DPS scope-and sequence documents to find the ELGs that were relevant for each level. The teachers then made modifications to the anchors based on their vision for the LP and their understanding of the curriculum documents. For the most part these changes were minor. A major change involved the notion of "understanding." When the research team wrote the anchors, we included knowledge, skills, and understandings related to algebraic manipulation. The teachers decided that they wanted the LP to be focused on skills, and that knowledge and understanding would be shown by correct execution of skills. As a group, we re-worded the anchors to reflect this. An

example is given in Table 6.1, which shows an example of how we modified the anchor for Algebra II.

Table 6.1 How we modified the Algebra II anchor to associate understanding with algebraic skill. Bold face is used to indicate text that changed.

Level 3 (algebra 2) before modification	Level 3 (algebra 2) after modification
Students understand solving equations as a process of reasoning based on the assumption that the original equation has a solution, and they justify their solution steps. Students solve quadratic equations over the set of complex numbers using multiple methods as appropriate (inspection, taking square roots, complete the square, factoring). **They understand positive, negative, and non-integer exponents**, and they solve exponential equations using properties of exponents or by converting to logarithmic form and evaluating the logarithms.	**Students show work when solving**. Students solve quadratic equations over the set of complex numbers using multiple methods as appropriate (inspection, taking square roots, complete the square, factoring). **They demonstrate understanding of positive, negative, and non-integer exponents by using them to solve exponential equations**.

At this point, we had created an LP similar to that described in Section 2. As shown in Figure 6.1, this LP has 9 levels: five anchors (indicated by whole-number level numbers) and four "messy middles" (indicated by decimal level numbers). The descriptions of each anchor are cumulative. They represent the changes and additions from the previous anchor without repeating previous statements.

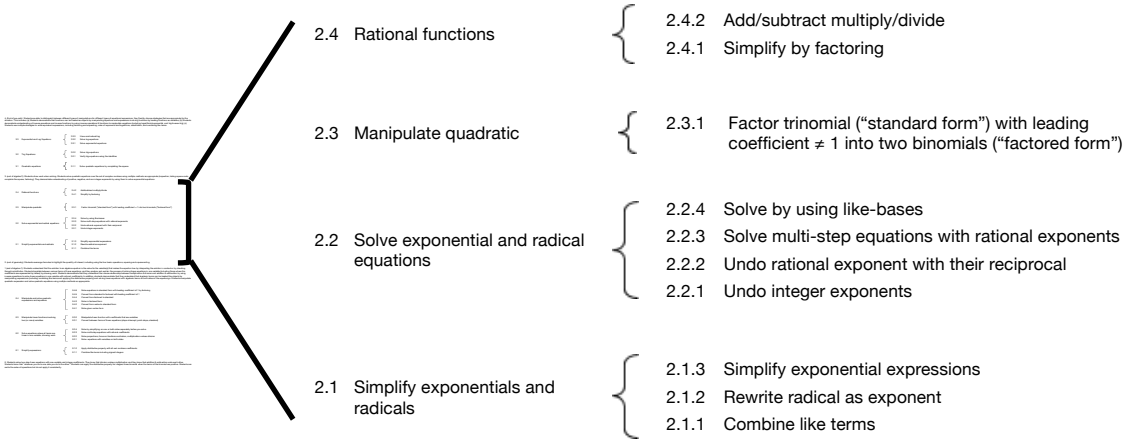
Anchor	Level	Level Descriptor
Upper anchor: End of Pre-calculus	4	Students are able to distinguish between different types of manipulations for different types of equations/expressions, they flexibly choose strategies that are appropriate for the situation. This includes: (a) Students demonstrate that functions can be treated as objects by manipulating equations and expressions involving functions by treating functions as variables; (b) Students demonstrate understanding of inverse operations and inverse functions by using inverse operations & functions to manipulate equations (including logarithmic/exponential, and trig/inverse-trig); (c) Students use multiple strategies to write equivalent expressions, including factoring and expanding, rules of exponents and logarithms, substitution, and combining like terms.
	3.*x*	*Partial mastery of Pre-calculus ("Messy Middle")*
End of Algebra II	3	Students show work when solving. Students solve quadratic equations over the set of complex numbers using multiple methods as appropriate (inspection, taking square roots, complete the square, factoring). They demonstrate understanding of positive, negative, and non-integer exponents by using them to solve exponential equations.
	2.*x*	*Partial mastery of Algebra II ("Messy Middle")*
End of Geometry	2	Students rearrange formulas to highlight the quantity of interest, including using the four basic operations, squaring and squarerooting.
	1.*x*	*Partial mastery of Geometry ("Messy Middle")*
End of Algebra I	1	Students understand that the solution to an algebra equation is the value for the variable(s) that makes the equation true by interpreting the solution in context or by checking through substitution. Students translate between various forms of linear equations, and they analyze and explain the process of solving linear equations in one variable (including those where the coefficients are represented by letters) by showing work. Students demonstrate that they understand the inverse relationship between multiplication & division and addition & subtraction by using inverse operations to solve linear equations in one variable with rational coefficients. In addition, students demonstrate that they understand that algebraic terms can be treated like objects by manipulating expressions (including combining like terms and applying the distributive property) and solving linear equations with algebraic terms on both sides of the equals sign. Students manipulate quadratic expression and solve quadratic equations using multiple methods as appropriate.
	0.*x*	*Partial mastery of Algebra I ("Messy Middle")*
Realistic lower anchor	0	Students solve two-step linear equations with one-variable and integer coefficients. They know that division undoes multiplication and they know that addition & subtraction undo each other. Students know that "whatever you do to one side you do to the other." Students can apply the distributive property for integers times binomial when the terms of the binomial are positive. Students can recite the order of operations but do not apply it

consistently.

Figure 6.1 The anchors used at GB

Following to the method described in Section 2, one would stop here and use the LP shown in Figure 6.1 as the basis for SLOs. At GB we went further and attempted to flesh-out the messy middles. This resulted in a number of "micro-levels" in-between each anchor. The teachers created these micro-levels using DPS scope-and-sequence documents. The micro levels are discrete algebraic skills and they are ordered based on the order in which they are taught at GB. Figure 6.2 shows a miniature version of the LP, along with a "zoomed-in" version of the 2.*x* micro-levels. Recall from Figure 6.1 that these micro-levels correspond to the "messy middle" of Algebra II. Thus, the ordered levels constitute a hypothesis about how students might progress from Level 2 (end of geometry) to Level 3 (end of Algebra II). The complete LP with all micro-levels can be found in Math Appendix D.

Figure 6.2 Schematic of the across-grade LP created at GB, with a zoomed-in version of the 2.*x* micro-levels. For the complete LP, see Math Appendix D



The complete LP defines the progression of a student from the beginning of Algebra I to the end of Pre-calculus. As such, it is too expansive to be useful at a given grade. To create useful LPs for each grade, teachers worked in course-specific teams to determine which portions

of the LP were most relevant to each course level. In some sense this had been done already—as shown in Figure 6.2, the section between Levels 2 and 3 corresponds to the skills that students are expected to learn in Algebra II. However, it would not be appropriate to limit the Algebra II LP to just these levels. Doing so assumes that students come to Algebra II at level 2. The teachers assured us that this was not usually the case, and that, in practice, students often entered a particular course below the top level of the previous course (e.g., students enter Algebra II below level 2). Thus, the Algebra II LP should begin lower then level 2.

Of the three anchors that define each course-specific LP (realistic lower anchor, aspirational lower anchor, and upper anchor), the aspirational lower anchor and upper anchor are defined by the anchors shown in Figure 6.1. The task for teachers was to define the realistic lower anchor. To define the realistic lower anchor, we asked teachers to come to a consensus on the upper bound of the lowest 1/3 of students in a given class and use this level as the realistic lower anchor.

Figure 6.3 shows the span of each course specific LP. The course-specific LPs are represented as rectangles against the background of the complete across-grade LP. Within each rectangle, the thick lines represent, in ascending order, the realistic lower anchor, the aspirational lower anchor, and the upper anchor. Math Appendix E shows each course-specific LP.

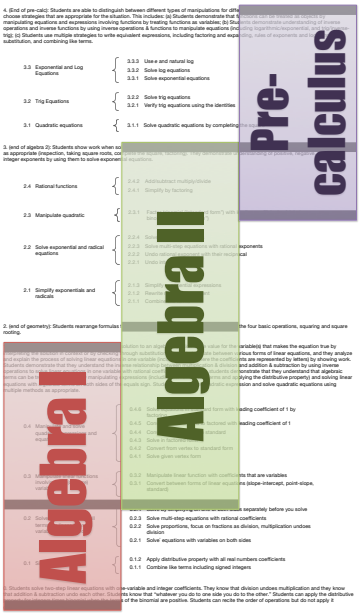


Figure 6.3 Schematic that shows the span of each course-specific LP against the background of the across-grade LP. Thick lines represent, in ascending order, the realistic lower anchor, the aspirational lower anchor, and the upper anchor of each LP

Choosing or Writing Items for Student Assessment at GB

The teachers at GB developed 28 tasks with rubrics. All of the tasks are constructed response. Twelve tasks are scored dichotomously (i.e., 0 or 1 point), and 16 tasks are polytomous. Figure 6.4 shows an example of a dichotomous task and Figure 6.5 shows an example of a polytomous task. All 28 tasks are aligned to the across-grade LP, as shown in Figure 6.6.

Task D

Problem:

Convert the equation below to slope-intercept form. *Show your work.*

$$y = 5 + 3(x - 2)$$

Rubric

Description	Score	Level(s) in LT
Equation is converted correctly	**1**	0.3.1
Equation is not converted correctly	**0**	Below 0.3.1

Figure 6.4 Task D; an example of a dichotomous task

Task A

Problem:

Solve the equation below for x. *Show your work or explain your method.*

$$\frac{2}{3}x + 5 = 4 - 2(x - 6)$$

Rubric

Description	Score	Level(s) in LT
Problem is solved correctly, for an exact answer (i.e., not a decimal approximation). Simple arithmetic errors are okay, but the key is that the student handles the fraction correctly.	**3**	0.2.4 or above
Problem is solved completely (including distributing and using inverse operations to collect terms, possibly with minor arithmetic errors not involving the fraction) except that student mishandles the fraction.	**2**	0.2.2 – 0.2.3
Student uses inverse operations to collect terms with an *x* on one side and constant terms on the other (arithmetic mistakes okay), but mishandles the fraction, or stops there without completely combining like terms. Student may not distribute, or may have arithmetic errors in distributing and combining like terms.	**1**	0.2.1
The student does not make any inverse operations to solve for *x*, or does not collect terms with an *x* on one side and constant terms	**0**	Below 0.2.1

on the other

Figure 6.5 Task A; an example of a polytomous task

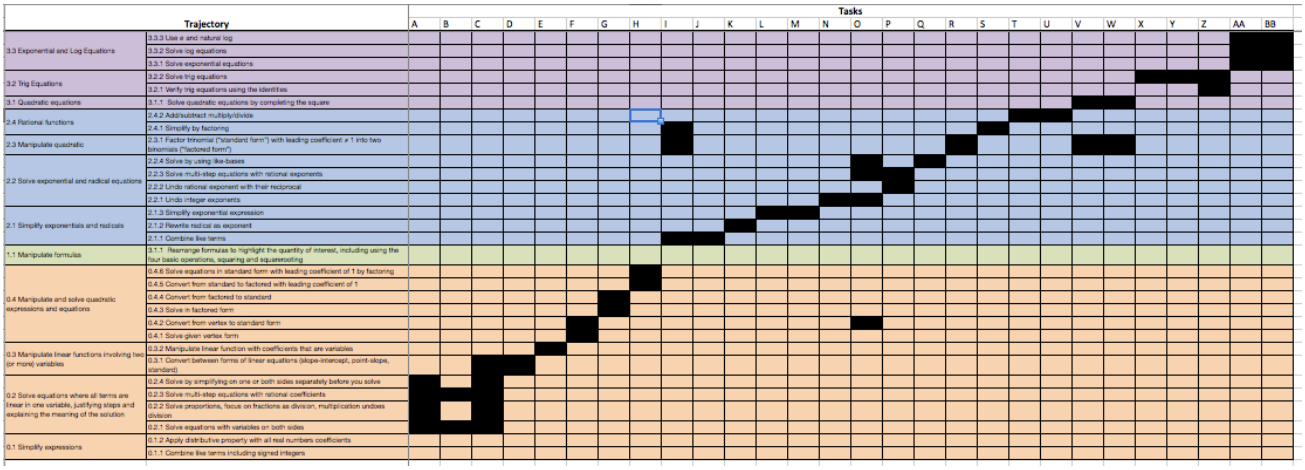


Figure 6.6 The alignment of tasks with the across-grade LP. Each row is a level in the LP and each column is a task. Black squares indicate a correspondence between the task and the level in the LP.

The teachers did not create all 28 tasks at once. First, they developed five tasks (two for Algebra I, two for Algebra II, and one for Pre-calculus). They gave these tasks to their students and scored them according to the rubrics. Then, as a group, we engaged in a score moderation activity. During this activity we focused on three aspects of the tasks and rubrics: (1) alignment of the task to the LP, (2) alignment of the rubric to the LP, and (3) clarity of the rubric, such that tasks can be scored by different teachers with high inter-rater reliability.

With respect to the first focal area, we found that some tasks asked questions that were not related to the LP, and that some rubrics differentiated aspects of student work that were not meaningful in terms of the LP. For example, one task asked students to analyze the algebraic steps of two hypothetical students, Amy and Ben. Amy made a mistake with the distributive property. Ben made a mistake manipulating signed integers. While the LP has levels corresponding to Amy's mistake, it does not have any levels that correspond to Ben's mistake. In fact, the teachers made a conscious decision to exclude manipulation of signed integers from the

LP because, even though students often make mistakes with signed integers while solving algebra equations, the teachers decided that manipulating signed integers is not itself an algebraic skill. The teachers modified this task to remove the analysis of Ben's mistake.

With respect to the second focal area, we found that some rubrics were written such that variance in score corresponded to variance in the amount of correctness in the students' work, rather than variance that was meaningful in terms of the LP. To be sure, these sources of variance are not necessarily mutually exclusive, and will often overlap. However, some rubrics were written such that the rubric levels did not correspond with variance in the LP. To help teachers write rubrics where variance in score was meaningful in terms of the LP, we asked the teachers to connect each rubric score with a range (i.e., one or more) of LP levels, modifying the rubric levels as appropriate. Figure 6.7 shows how one rubric changed before and after the activity. As shown, the "after" rubric has a column to map the rubric score to the LP level. Furthermore, descriptions in the "after" rubric all mention the presence or absence of three key skills: using inverse operations to collect like terms on respective sides of the equation, simplifying one side of the equation (combining like terms and using the distributive property), and working with the fraction. At the time of this activity, these three skills composed levels 0.2.1, 0.2.3, and 0.2.4 respectively in the LP. Thus, although the rubric still describes a schedule of progressive correctness in solving the problem, the schedule is aligned to the levels in the LP.

Before session 3

Description of student work	Score
Simplify the right side of the equation first or subtract 4 to both sides. The student will solve the equation for X and get a correct response.	**3**
Student does incorrect distribution and does 4 – 2 before the distribution or does not apply the negative through the (X – 6). But can algebraically manipulate for X and mishandles the fraction.	**2**
Student attempts algebra steps but does not apply the rules correctly	**1**
Student does not demonstrate any clear idea on how the solve the problem.	**0**

After session 3

Description of student work	Score	LP level
Problem is solved correctly, for an exact answer (i.e., not a decimal approximation). Simple arithmetic errors are okay, but the key is that the student handles the fraction correctly.	**3**	0.2.4 or above
Problem is solved completely (including distributing and using inverse operations to collect terms, possibly with minor arithmetic errors not involving the fraction) except that student mishandles the fraction.	**2**	0.2.2 – 0.2.3
Student uses inverse operations to collect terms with an *x* on one side and constant terms on the other (arithmetic mistakes okay), but mishandles the fraction, or stops there without completely combining like terms. Student may not distribute, or may have arithmetic errors in distributing and combining like terms.	**1**	0.2.1
The student does not make any inverse operations to solve for *x*, or does not collect terms with an *x* on one side and constant terms on the other	**0**	Below 0.2.1

Figure 6.7 How the rubric for task A was changed during the score moderation activity.

With respect to the third focal area, we found that some rubrics used language that was ambiguous. For example, consider level 3 of the "before" rubric in Figure 6.7. The description for this level refers to a "correct response." The teachers discussed whether the correct response had to be in exact form, or whether a relatively accurate decimal approximation would be sufficient (for this task, some students converted the fraction 2/3 into 0.66 before solving, resulting in a final answer that is off by one one-hundredth). The teachers decided that the answer had to be in exact form and, as shown in the "after" rubric in Figure 6.7, they modified the rubric accordingly. Notice that the rubric is still written at a fairly general level. For example, the descriptions refer to "the fraction" without explaining what the fraction is, and the description does not explain what "handling correctly" or "mishandling" the fraction means. The teachers decided that such detail was unnecessary and that any algebra teacher would interpret these terms in a consistent way.

The complete set of 28 tasks and rubrics contains the five tasks that we used in the score moderation activity, as well as 23 additional tasks that teachers wrote after the score moderation

activity. In creating the additional tasks, teachers drew on the lessons from the score moderation activity. However, the second set of tasks were not themselves part of any score moderation. The tasks are available on request.

Teachers created assessments by selecting tasks from the bank of 28 tasks described above. We asked the teachers to select tasks based on the levels that the teachers thought the students would be in the LP. We suggested that there ought to be more than one task per level, and that there should be some tasks above and below the anticipated levels. We recommended that the final assessments should take students about one class period to complete, but that the assessment could be given over multiple class periods. The "one period" heuristic was somewhat arbitrary. We chose it because it seemed to be a good balance between having enough tasks to get a relatively reliable assessment, while not taking up too much class time or grading time. We suggested that all of the teachers for a given course should work together to create the final assessment for that course. We made this suggestion for two reasons. First, because we believed that collaborative work would improve the quality of the assessments. Second, because our interpretation model requires a test-specific mapping of student scores to the learning trajectory. The fewer tests there were, the fewer mapping teachers would have to create. Despite our suggestion, most teachers created assessments independently.

Scoring and Interpreting Student Work at GB

In Chapter 2 we described a method for classifying students into levels in the LP, based on the students' total score on an assessment. At GB, the Algebra I, Algebra II, and Pre-calculus LPs have 15, 20, and nine levels respectively (see Math Appendix E). Thus, the job of the interpretation model is to classify students into these levels based on the students' assessment scores.

In Chapter 2 we also described how measurement error has an adverse effect on the accuracy and generalizability of the categorizations. With respect to the course-specific LPs at GB, accurately classifying students into nine, 15, or 20 categories is an enormous challenge. There are two ways to increase the classification accuracy: give more assessment items or reduce the number of categories. As discussed above, the former is probably not realistic given the

constraints on teachers' time. Thus, we asked the teachers, in course-specific groups, to reduce the number of classification categories to no more than seven. We suggested that teachers should create these "accountability levels" by aggregating micro-levels that were related such that differences between them would be difficult to distinguish. After this activity, Algebra I had six accountability levels and Algebra II and Pre-calculus each had five accountability levels. The accountability levels for Algebra I are shown in Figure 6.8.

The next step was to quantify growth between levels. In Chapter 2, we described a method in which movement between any pair of adjacent levels is worth one point. At GB, the teachers decided that some inter-level movements were more difficult than others. Thus, teachers worked in course-specific teams to define "growth scores" between each pair of adjacent accountability levels. We instructed teachers to first imagine that each "jump" between adjacent accountability levels was worth one point. Then, we asked teachers to modify the growth scores to reflect the difficult of the "jump" between levels, such that: (a) more difficult "jumps" had higher point-values, and equi-difficult jumps had equivalent point values. Figure 6.8 shows the accountability levels and growth scores for Algebra I as an example. As shown, the teachers judged the jump from level C to D to be the easiest. The jumps from A-B, D-E, and E-F were judged to be the most difficult.

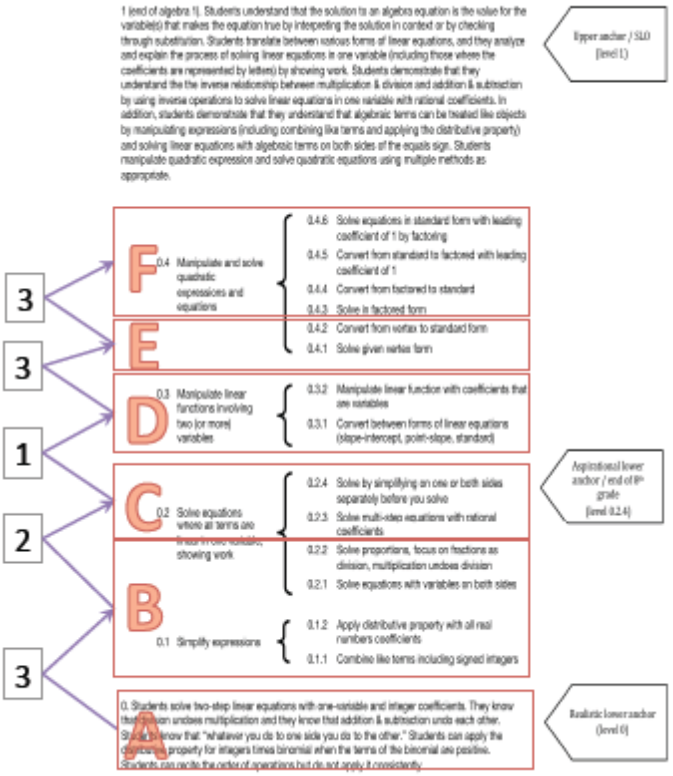


Figure 6.8 Accountability levels and growth scores for the Algebra I LP

Both the number of accountability levels and the growth scores between accountability levels are unique to each course-specific LP. However, as discussed in Section 2, we can compare growth across courses by defining two growth milestones for each course: minimal growth and aspirational growth. Minimal growth is the lowest possible growth score in the course and aspirational growth represents one year's growth (as defined by the sum of the growth scores between the aspirational lower anchor and the upper anchor). For example, for the Algebra I LP shown in Figure 6.8, the minimal growth is 1 and the aspiration growth is 7, because a student who moves from the aspirational lower anchor to the upper anchor would receive 7 points. Table 6.2 shows the growth milestones for each course-specific LP. While growth scores give us a way to compare growth within a course, the growth milestones give us a way to compare growth across courses.

Table 6.2 Summary of growth scores for each course

Course	Number of accountability levels	Minimal growth	Aspirational growth
Algebra I	6	1	7
Algebra II	5	1	4
Pre-calculus	5	1	6

Teachers created five assessments. Following the method described in Section 2, the teachers created a mapping between the students' total scores and the accountability levels in the LP. Figure 6.9 shows an example of this mapping for an Algebra I teacher.

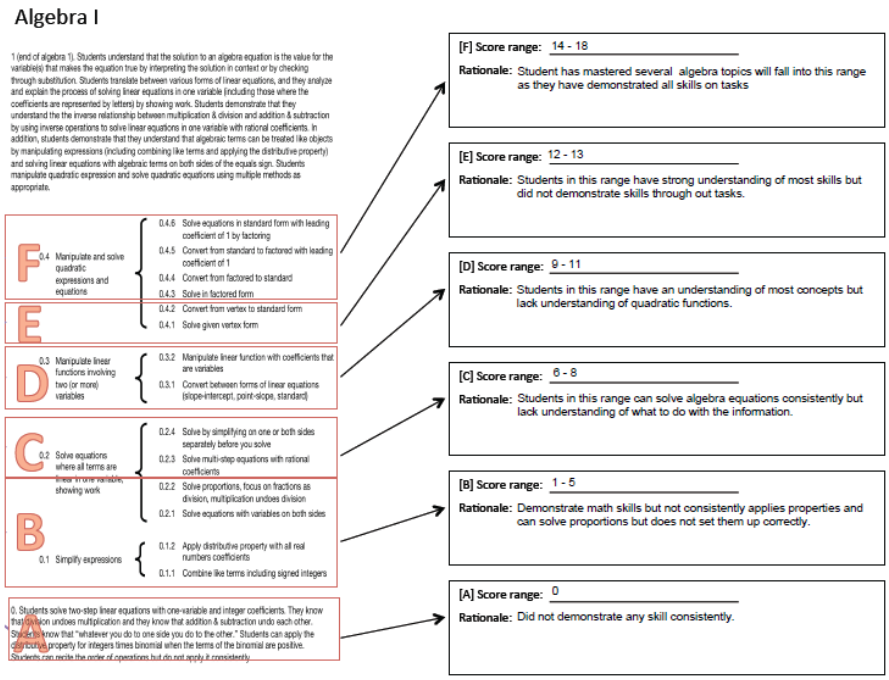


Figure 6.9 The mapping between total score and LP for Test B

As discussed in Chapter 2, the quality of the underlying assessment is an important consideration in the interpretation of the results. In Chapter 2, we introduced three criteria from the NCIEA for high quality assessments. We suggest that these criteria are met by the task-creation process described above. Recall that in this process teachers created tasks and rubrics

that were explicitly tied to the LP (criterion (a)), and engaged in score moderation activities to ensure that the rubrics met the criteria listed in criterion (c). With respect to criterion (b), we asked teachers to consider this criterion while creating tasks, and the teachers felt that the tasks were at a minimum unbiased. Furthermore, because all of the tasks are straightforward algebraic manipulation tasks, we believe this criterion is met, although we did not engage in any formal process to ensure this.

In addition to the qualitative criteria discussed above, we also examined the four quantitative measures of assessment quality discussed in Chapter 2. In general, the tests had low SEMs and relatively high reliability, but possibly not high enough to support the number of categorizations they were used to make. In terms of items, more than 90% of items matched the underlying assessment, and the distribution of p-values suggests that the items spanned an appropriate range of difficulty. See Math Appendix C for more details.

In general, then, the teacher-created tests seem to be of generally high quality. We now turn to discussions of student growth, using the benchmark measures of minimal and aspirational growth. As shown in Figure 6.10, approximately half of students had growth scores between minimum and aspirational growth, while approximately one quarter of students had growth scores below minimum growth and at or above aspirational growth. The distribution of growth levels varied by teacher (See Figure 6.11).

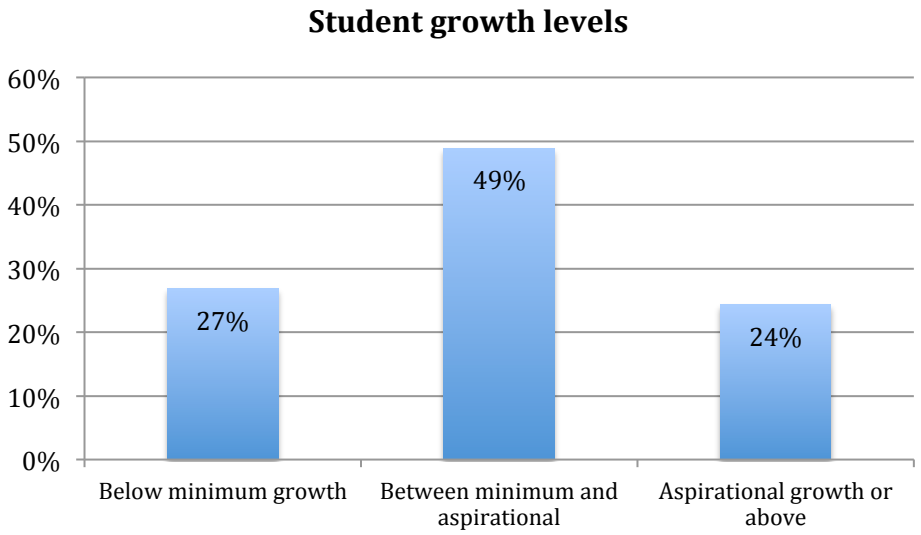


Figure 6.10 Distribution of student growth levels for all students.

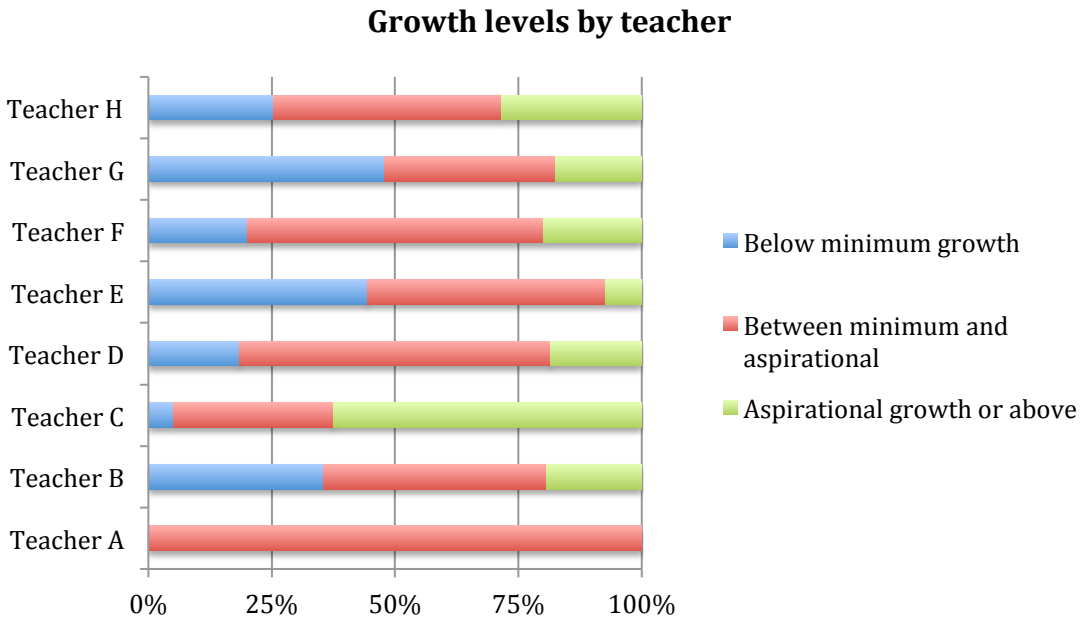


Figure 6.11 Distribution of student growth levels by teacher.

Rating teachers on the basis of student growth at GB

In Section 2, we described a method to classify teachers into four effectiveness levels based on student growth benchmarks. Each school created their own categorization schedule, based on the template given in Section 2. Figure 6.12 shows the categorization schedule that the teachers at GB created. Based on this classification schedule, 4 teachers were classified at Level 2, and 4 teachers were classified at Level 3.

Teacher level	Below minimum growth	Between minimum and aspirational	Aspirational growth or above	Narrative description
4	0%		$> 33\%$	None below minimum growth and at least 33% aspirational growth or above
3	$< 33\%$		$> 15\%$	Less than 33% below minimum growth, and at least 15% aspirational growth or above
2	$< 50\%$	$\geq 50\%$ combined		At least half above minimum growth
1	$> 50\%$	$< 50\%$ combined		More than half below minimum growth

Figure 6.12 The teacher classification schedule at GB

Teacher reactions and feedback

We collected teacher feedback in formal and informal ways. Formally, we distributed a survey at the end of each of the first four sessions. The feedback forms asked three questions: (1) What went well during today's session; (2) What can be improved for next time; and (3) What are some things that you like to learn from this collaboration? We distributed a more comprehensive survey at the end of the final session, in which we asked teachers to reflect on the pilot as a whole. Finally, we observed teachers' PDU presentations, in which they described to their peers their evaluations of their own work in the pilot. Informally, we took notes of teacher feedback as it arose naturally in conversation and group discussions during each session. Our summary of teacher feedback focuses on the key areas that teachers often talked and wrote about: (1) Interest and concerns with the LP, (2) the scope and time commitment of the project, and (3) the ways that the pilot work impacted teacher's classroom practice.

Teachers interest and concerns with LPs

The notion of an LP seemed new to the teachers at GB. However, they were interested in vertical collaboration and in measuring students based on "where they are" as opposed to normative expectation (e.g., district assumptions or those based on CCSSM).

One (crude) way to gauge the initial interest in LPs is to tabulate the number of times that teachers mentioned LPs in the feedback forms form the first two sessions. As shown in Table 6.3, teachers mentioned LPs in about half of their responses to "what went well", and in about 30% of the responses to "what do you want to learn". Sample comments include (note that at GB we used the term 'trajectory' rather than 'progression'):

What went well

- "Idea of trajectory and common threads" (session 1)
- "Understanding the learning trajectory process" (session 2)

What do you want to learn?

- "Building trajectories for more concepts" (session 1)
- "How to use the trajectory" (session 2)

In the 16 feedback forms that we collected over the first two sessions, there was only one negative comment about learning trajectories:

> It feels like some of the trajectory work was repetitive and had already been done through the district scope and sequence (session 2)

Table 6.3 Mentions of LPs on feedback forms.

Section of feedback form	Session 1 (n=8)	Session 2 (n=8)
LP mentioned in "what went well"	3	4
LP mentioned in "what can be improved?"	2 (both wanted more time for LPs)	1
LP mentioned in "what do you want to learn about"	2	3

During sessions 3 and 4, one teacher expressed a concern that student learning might not follow the LP because the LP itself is not hierarchically ordered. This teacher used the example of the micro-levels between levels 2 and 3 (see Figure 6.1 above), as follows. The teacher explained that there is nothing inherent in the ability to manipulate rational functions (the 2.4 levels) that indicates or requires that the student is able to solve exponential equations (the 2.2

levels). The teacher argued that if a student answers a level 2.4 question correctly, this should not be taken as evidence that the student has mastered the skills below level 2.4. It is entirely possible that students may answer level 2.4 questions correctly without having mastered the skills at lower levels in the LP.

This teacher was comfortable with the ordering of the micro-levels within a group (e.g., within 2.1), but suggested that each of the groups ought to be considered independent of the others. Thus, rather than a linear path from level 2 to level 3 there would be a branching, as shown in Figure 6.15. In this scheme, full mastery of each micro-level would still be required for a student to be located at level 3, and micro-levels are still ordered within their group (e.g., levels 2.1.1, 2.1.2, and 2.1.3 are all ordered within the 2.1 group). However, this scheme allows for non-linear development of each group of micro-levels. Thus, a student at level 2.4.2 would be assumed to have mastered the skills inherent in level 2.4.1, but would not be assumed to have mastered the skills in any of the other groups.

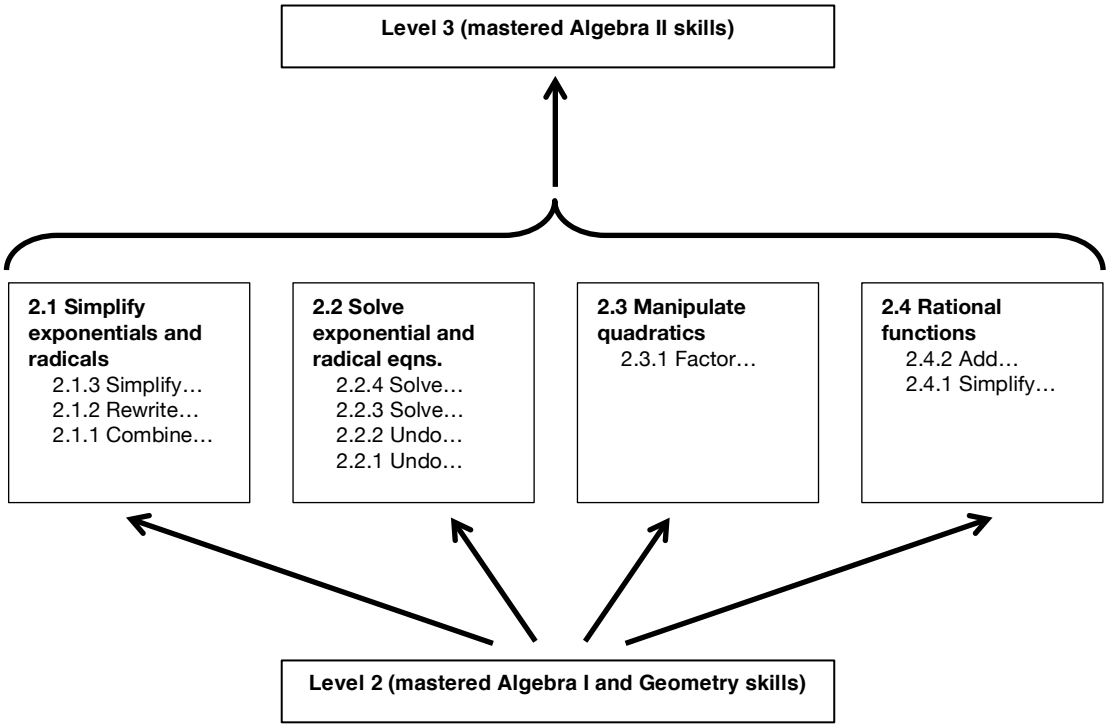


Figure 6.15 The micro-levels between levels 2 and 3 as branches

Each time the teacher brought-up this concern, we discussed it as a group. We all agreed that the linear ordering level groups were based only on the sequence in which the topics were taught at GB. In some cases, this order seems arbitrary from a disciplinary perspective. In other cases however, the order makes sense from a disciplinary perspective. For example, some teachers suggested that the 2.3 group ought to come before the 2.4 group because simplifying rational functions often involves factoring quadratics. Similarly, some teachers argued that the 2.1 group (simplifying exponential expressions) is logically prior to the 2.2 group (solving exponential equations). This led to a discussion that there might be two "branches" within the Algebra II micro-levels.

Even though most teachers agreed with the branching scheme in theory, some argued that the LP ought to maintain a linear progression. They argued that the linear progression represents the sequence that the skills are taught at GB, and thus it is likely that student who have mastered a upper level have mastered a lower level, even if the lower level is not logically prior to the upper level from a disciplinary perspective. The teachers in favor of the linear sequence agreed that there will be students whose learning does not follow the sequence, but the teachers argued that the sequence will probably capture the path of learning for the majority of students at GB. For purposes of the pilot, we decided to maintain the linear nature of the LP. Note that the concern about the ordering of micro-levels can be alleviated by using the generic LP discussed in Section 2, because this LP makes no *a-priori* hypotheses about order within the "messy middle."

We have our own concerns about the LP. As described above, a key feature of LPs is that they are based on an examination of student reasoning:

> Learning trajectories… must be based on careful empirical study of the development of student reasoning and not simply delineated as a logical analysis of mathematical topics (Confrey, 2012, p. 158)

It seems to us that the LP that we created at GB was not solely based on the teachers' expertise on student reasoning. We say this for a couple of reasons. First, because the micro-levels in the LP (see Figure 6.1) are not written in terms of student thinking, but rather in terms of mathematical topics, and the ordering of the topics was based on the order in which topics are taught at GB, rather than through analyses of student reasoning. This suggests that the LP we created at GB is not a "learning progression" in the way that the research community defines the

term. In retrospect, we believe that it would have been better to start with the generic LP described in Section 2, and to use student work to flesh out the "messy middle" over a number of years.

Scope and time commitment of the project.

In every session, teachers brought up the scope and time commitment of the project. Our work together took four full-days and one half-day. The full days were scheduled on district "blue-green days," which the teachers understood to be half-day for PD and half-day for teacher work time. Thus, the teachers perceived our work as cutting into a significant amount of their work time. Although they negotiated a settlement with their administration, the teachers still mentioned their loss of work time in every session. In addition, teachers told us that there was a significant out-of-session time commitment required to administer and score student assessments. They asked us, "Where is this time supposed to come from?" Finally, the time commitment was a prominent theme in the final survey. Nearly every teacher mentioned that the time commitment was great.

Benefits of collaboration

Teachers indexed the benefits of discussion and collaborative work more than any other topic in the "what went well during today's session?" prompt on the feedback forms. Often teachers simply wrote "collaboration" or "collaboration with colleagues" in response to this prompt. Sometimes they were more specific. For example:

- "Good discussion regarding alg 1 micro trajectory"
- "Discussing tasks that are useful and meaningful"
- "Talking with colleagues about SLO's"
- "Discussion of student work"

How the pilot impacted the teachers' classroom practice

On the final survey, teachers indicated two ways that the pilot impacted their classroom practice. First, they reported that they appreciated the focus on growth rather than status (e.g., "I feel that the SLO process provides a more direct path for monitoring student growth and has been helpful in providing a direction for how we monitor our students"). Second, they reported

that the work helped to focus their teaching on specific skills, rather than broadly reteaching a unit (e.g., “I have changed a little in regards to re-teaching very specific concepts to help students make progress.”).

References

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.

Buckley, K. & Marion, S.F. (2011). A Survey of Approaches Used to Evaluate Educators in Non-Tested Subjects Grades and Subjects. Retrieved from: www.nciea.org.

Clements, D. H., & Sarama, J. (2004). Learning trajectories in mathematics education. *Mathematical Thinking and Learning*, *6*(2), 81–89. doi:10.1207/s15327833mtl0602

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.

Confrey, J., Maloney, A., Nguyen, K. H., Mojica, G., & Myers, M. (2009). Equipartitioning/splitting as a foundation of rational number reasoning using learning trajectories. In M. Tzekaki, M. Kaldrimidou, & C. Sakonidis (Eds.), *Proceedings of the 33rd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 1). Thessaloniki, Greece: PME.

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

Duschl, R. A., Maeng, S., & Sezen, A. (2011). Learning progressions and teaching sequences: a review and analysis. *Studies in Science Education*, *47*(2), 123–182. doi:10.1080/03057267.2011.604476

Goe, L. & Holdheide, L. (2011). Measuring Teachers' Contributions to Student Learning Growth for Nontested Grades and Subjects. Princeton, NJ: National Comprehensive Center for Teacher Quality.

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

Slotnik, W.J., Smit, M., Glass, R.J., Helms, B.J. (2004). Catalyst for Change: Pay for Performance in Denver Final Report. Community Training and Assistance Center. Retrieved from: http://www.ctacusa.com/denver-vol3-final.pdf

Wilson, P. H., Edgington, C. P., Nguyen, K. H., Pescosolido, R. C., & Confrey, J. (2011). Fractions: How to share fair. *Mathematics Teaching in the Middle School*, *17*(4), 230–236.

Wilson, P. H., Myers, M., Edgington, C. P., & Confrey, J. (2012). Fair shares, matey, or walk the plank. *Teaching Children Mathematics*, *18*(8), 482–489.

Section 2: Tier 3 Visual Arts Pilots

Chapter VII: The Visual Arts Pilot

Introduction

This chapter focuses on the Tier 3 student learning objective (SLO) pilot work in the visual arts and describes outcomes and lessons learned from the professional development sessions completed with two teams of teachers: 1) two visual arts teachers from Green Beechway High School; and 2) nine teacher leaders in the arts representing various levels (elementary, middle and high school) and schools in the Denver Public Schools (DPS). Content and activities for the Tier 3 work for visual arts were developed by Elena Diaz-Bilello from the Center for Assessment and supported by Rajendra Chattergoon from CADRE. Capucine Chapman, academic program manager for the visual arts in DPS, and Elena Diaz-Bilello co-led the content development work with the second team of arts teacher leaders during the spring.

The plan of the Tier 3 pilot was to provide supports and professional development workshops on the SLO process using *learning progressions* or LPs (see Chapter 2) for teachers at each of the three selected pilot sites (Blue Elementary School, Green Beechway High School, and Spruce Middle School) in visual arts. As outlined earlier in Chapter 2, visual arts was picked for the Tier 3 pilot to determine what type of outcomes could be achieved for a subject area in which a teacher was either likely to be the sole person instructing the subject in the entire school, or the subject was one which most schools do not dedicate resources and time to building a strong data team or collaborative culture among singleton teachers[2].

Of the original three sites, only two started the visual arts pilot in October since the single visual arts teacher at Blue opted to not join. The pilot work initially began with two teams of teachers, one at Spruce and one at Green Beechway (GB) High School. At the end of the fall semester, the Center for Assessment joined forces with the Visual Arts Department at DPS to offer a series of workshops focused on leveraging the ongoing units of study development work as an entry point for exposing teacher leaders in the arts to the new SLO process and to the Tier 3 focus on LPs. The two visual arts teachers at the Tier 3 pilot site at Spruce decided to join the teacher leader workshops in order to collaborate with a larger group of arts teachers and to

[2] A secondary objective for the visual arts pilot was to see if a collaborative and supportive environment could be fostered with "singleton" (a term used by the district to describe a teacher who is the only instructor for a subject in the entire school) teachers to strengthen the buy-in for the SLO process at each school site.

reengage with the new SLO process using their units of study. As documented later in this report, the Spruce teachers faced tremendous difficulty participating in the original fall pilot, preferring to apply their knowledge in developing units of study to the new SLO process. The work with the two visual arts teachers at GB continued as originally envisioned in a separate pilot for the remainder of the school year. In this report, we refer to our work with the GB teachers as the "original pilot".

As elaborated upon in Chapter 2 of this report, the cornerstone of our work in the visual arts was the use of a LP as a model of cognition for evaluating student learning on a critical objective embodying a big topic area that transcended grade levels. The workshop activities of developing a LP, specifying the tasks to evaluate students on each level of the progression, and using the scores from the tasks to inform teachers about where students were located relative to the learning objective served to operationalize the National Research Council's (2001) "assessment triangle" (please refer to pg. 15 for a discussion of the assessment triangle).

The visual arts pilot initially followed the same steps as the mathematics pilot. Teachers at GB and Spruce focused on developing a multi-grade LP that spanned across grade levels, and a *course-specific trajectory* that specified the learning objective for the SLO as the final aspirational goal for all students in each teacher's course. The purpose of developing a multi-grade progression for the pilot was to help teachers identify big picture concepts in the visual arts, pinpoint what the learning objectives supporting those concepts should look like in each grade or grade range, and connect these concepts and learning objectives to the Colorado Academic Standards for the visual arts.

A *course-specific trajectory* represented a segment (i.e. one level) of the multi-grade LP and served as a major focal area for undertaking this work with the arts teachers in both the original Tier 3 and the teacher leader workshops. In contrast to the multi-grade LP, the course-specific trajectory focused on identifying key attributes that students are expected to demonstrate in order to move toward the learning objective (formerly referred to as the "learning content" in the SGO system). The course-specific trajectories were used as a framework to identify instructional steps and supports needed to move students to higher levels along the specified progression. Teachers reviewed student work to identify the characteristics of students located at each level of the progression and to determine how students needed to progress in order to move

to higher levels of the progression. Teachers also reviewed student work and responses to tasks to help locate their students along the course-specific trajectory.

In the teacher leader workshops, teacher leaders developed only course-specific LPs in a group activity using a pre-existing unit of study developed by a teacher. In contrast to the original visual arts pilot, in which teachers moved through the SLO process and evaluated their students on the learning objectives developed, the teacher leader pilot was focused on the future implementation of SLOs. That is, the teacher leaders were given a structure for understanding how the SLO process and the use of a LP could help frame the ongoing unit of study development and the sequence in which these units could be used during the school year to assess students. This group of teachers did not have to track and evaluate their own students' progress on the learning objectives they developed.

Following this introduction, the visual arts section is divided into two chapters. In Chapter 8, we provide an overview of the structure and activities of the professional development sessions for the original pilot with GB and Spruce, describe the context and the process used to engage teachers with the SLO and LPs work, and describe associated outcomes. In Chapter 9, we provide the same overview of the structure and activities for the teacher leader workshops and also discuss associated outcomes related to that work. Each chapter culminates with a summary of successes and challenges learned from each pilot.

Chapter VIII: The Original Visual Arts Pilot

Professional Development Structure

The professional development sessions for the original Tier 3 visual arts SLO pilot were initially structured as full-day sessions that would take place five times throughout the school year for each pilot site. During the course of the pilot, two of the five full-day sessions were split into half-day sessions due to scheduling conflicts. The proposed key objectives of the sessions were to:

- Introduce participants to the new SLO process and its associated expectations.
- Develop both multi-grade and course-specific trajectories connected to a big idea and a cluster of standards while simultaneously embedding the SLOs into the highest level of each course-specific trajectory.
- Map selected performance assessments used in each semester to each level of the course-specific trajectory.
- Determine the starting points of students relative to each course-specific trajectory developed.
- Define the performance expectations to be achieved by students using the 4-point rubrics developed by each teacher on all of the tasks selected for each course-specific trajectory.
- Evaluate student growth in a semester by comparing students' starting and ending points relative to each course-specific trajectory.
- Determine a final rating for teachers using the pilot rules.

The set of activities and content covered in each session during the school year are reflected in Table 8.1 and correspond to the session objectives proposed to the district. The Spruce teachers completed the first two sessions in the fall before opting to join the larger teacher leader arts pilot. The GB teachers participated in all of the activities noted in the table.

With the exception of session 2, all of the sessions took place at the respective pilot site with one or two DPS central administration staff members observing. School-based administrators did not attend any of the sessions. In session 2, prior to the Spruce teachers leaving the original pilot, we brought the Spruce and GB teachers to a DPS administration building so that they could serve as critical peers on the LPs they developed. In the next section, we provide background information about the participants to provide context about the extent to which the teachers entered the pilot with a foundation for engaging in the SLO work.

Table 8.1. Activities and Content Addressed During Original Pilot Sessions

Session	Activities
Session 1 **(split between 2 days for GB)** **October 2013**	- Overview of new SLO process. - Reviewing new SLO template and each component part. - Assessing student work (introduction to thinking about a developmental model of student learning and differentiating student work). - Introduction to LPs and connection to the SLO process. - Surveying LPs in the visual arts. - Reviewing an example of one multi-grade progression. - Drafting a multi-grade progression.
Session 2 **November 2013**	- Completing a multi-grade progression and completing a draft of one course-specific trajectory. - Mapping the location of selected performance/unit tasks to the course-specific trajectory. - Brief overview of assessment design considerations relative to cognitive complexity and to different student backgrounds.
Session 3 **January 2014**	- Establishing baseline information for students using first semester assessments. - Understanding criteria in the assessment review tool (content coverage, cognitive complexity, scoring, and fairness) and applying the tool to one performance task. - Reviewing student work and identifying criteria for differentiating students on the LP. - Assembling and identifying performance assessments as "body of evidence" for evaluating students on the SLO. - Developing the second course-specific trajectory.
Session 4 **February 2014**	- Completing and refining course-specific trajectories. - Determining the expectations for performance on different assessments to locate students on course-specific trajectories. - Populating value-tables to evaluate student growth for the pilot. - Reviewing and coming to a consensus on the pilot rules used to establish standards for minimum, acceptable, and aspirational growth (i.e., met the SLO).
Session 5 **(split between 2 days)** **April and May 2014**	- Revisiting the assessment review tool and criteria. - Working through the quality tool using a performance assessment. - Recording the final location of students on the learning objective. - Determining teacher ratings based on student outcomes achieved on each SLO. - Discussion of student work and instructional strategies. - Interview with teachers.

Participant Description and Context

Teacher Background and Perspectives on SLOs and SGOs

At GB, one visual arts teacher taught ceramics and the second taught drawing and painting. The ceramics teacher at GB was new to the school and the teaching profession. Before accepting a teaching position at GB, she worked as a student intern at DPS' Slavens School as part of her graduate coursework. Since she was a new teacher, she did not have any input or feedback about the ProComp student growth objectives (SGO) process. The other visual arts teacher at GB was a veteran teacher with over 30 years of experience. In the first session, he stated that he wanted to understand how it would maintain the integrity of the content taught to his students under this new SLO process while still meeting the requirements of the accountability system (personal communication, October 29, 2013). He typically developed his SGOs using one set of pre- and post-tests he constructed on arts vocabulary. Every year, at least 80% of his students met both objectives and he received the base-building award associated with SGOs in the district's ProComp system. The drawing teacher expressed his frustration about struggles with administrators over the years to communicate how performance-based content should be evaluated using a portfolio-based model. The drawing teacher said he was informed that he could only use formal pre- and post-tests for evaluating student growth on his SGOs.

The visual arts teachers at Spruce include two experienced teachers, one of whom specializes in digital arts. Both teach grades K-8. However, the visual arts course is not offered to every grade each academic year. The Spruce visual arts teachers expressed that the SLO language made more sense to them than the SGO language (personal communication, October 10, 2013). For them, the proposed SLO structure was a "better fit" than the SGO structure where their principal wanted them to only use pre- and post-tests. Like the GB drawing teacher, they expressed discomfort with the idea of a pre- and post-test model to evaluate student growth on a content area in which most of the activities were focused on the completion of performance tasks. Both teachers were accustomed to using a body of evidence approach with multiple sources of data to monitor the progress of students. Unlike SGOs written in previous years, as of the 2012-2013 academic year, they were required to incorporate the new Colorado Academic Standards for the visual arts into the learning objectives. Both Spruce visual arts teachers stated that their administrators mandated their participation in the SLO pilot.

Data Teams Context and Professional Development

As indicated by many teachers in DPS, the success of the SLO or SGO process, or the degree to which the process is meaningful for instruction, hinges largely on the strength of data teams or professional learning communities established at schools (Briggs, et al., 2014; Slotnik, et al., 2004). Considering that many schools have only one or two visual arts teachers, it seemed unlikely to us that data teams established at each school site focused on content-specific issues and concerns related to the arts. We also wanted to know the extent to which professional development opportunities were available to help deepen the pedagogical content knowledge in the visual arts, and we wanted to learn whether those opportunities might provide another venue for teacher collaboration.

At Spruce, the two visual arts teachers indicated that they did not have opportunities to collaborate and discuss student work and instructional strategies. As noted by both teachers, their collaboration times did not include instructional planning time and they rarely had time to "even speak to one another" (personal communication, October 10, 2013). However, they participated in the optional unit of study professional development sessions through the DPS visual arts department during the 2012-2013 school year. Based on the input received from both teachers, this was the only content-related professional development opportunity they received from the district. Both teachers praised Capucine Chapman from the visual department for organizing those sessions as collaborative work sessions with other arts educators in the district.

At GB, the two visual arts teachers engaged in conversations with the GB performing arts teachers, but the conversations did not center on student work/learning or on instructional issues and challenges. The veteran teacher had participated in the visual arts department's optional units of study professional development sessions but did not indicate that he was still involved with that work. Both teachers confirmed during the first session of the pilot that they have not had opportunities to collaborate about student learning and instruction, and both indicated that they enjoyed the opportunity given during the first session to "talk about student work" specific to their content area (personal communication, October 29, 2013).

Standards Implementation Work

The four visual arts teachers in the original pilot unanimously expressed a high degree of dissatisfaction with the new Colorado Academic Standards (COAS) in visual arts. According to the four teachers, the new state standards:

- Did not highlight any of the requisite skills that students should learn, although visual arts is a skills-based subject.
- Were too conceptual, without consistent interpretation and understanding of the grade level expectations and evidence outcomes across classrooms in the district.
- Were accompanied by a range of DOK levels that were not understood by most teachers since no training was provided at the district or state level to help teachers understand what the DOK levels meant relative to each of the evidence outcomes.

District guidance to teachers developing the units of study in the visual arts is that they need to address all four of the broader standards (observe and learn to comprehend, envision and critique to reflect, invent and discover to create, and relate and connect to transfer), and address specific grade-level expectations and evidence outcomes in their units. The three teachers' work on units of study helped them identify which grade level expectations and evidence outcomes were addressed in their units. However, despite their exposure to the new standards through the units of study work, the three teachers still expressed frustration with establishing clear connections between the lessons they developed and ensuring that they were addressing all of the required standards.

Assessment Practices

Given the performance-rich nature of their subject, all three experienced visual arts teachers had developed many performance tasks to evaluate their students over the course of the semester. Many of the performance tasks were embedded in units that would be completed over the course of several days or weeks and graded separately. For example at Spruce, one teacher developed a semester long unit that involved multiple students collaborating across multiple grades and culminated with the production of a final sculpture for the Denver Museum of the Arts. At GB, the veteran teacher developed units of study that encompassed several days of work and evaluation, and culminated with a summative task.

In addition to the performance tasks developed for their units of study, all of the teachers administered the pre-and post-tests developed by the Assessment, Research, and Evaluation (ARE) department. However, none of them wanted to use these tests to evaluate their students since they felt that the assessments did not accurately capture what their students know relative to the learning objectives set, and did not provide them with useful information about their students.[3] For this reason, we allowed the teachers to select existing performance tasks to both establish baseline performance on the SLO for students and evaluate students on the learning objective, as long as they could demonstrate that these tasks were aligned to the attributes in their progressions.

Considering that these teachers did not have prior experience working collaboratively with each other at each school site and did not seem to value the standards in their subject area, the pilot started out with some significant challenges. However, the fact that these teachers were accustomed to using performance tasks to evaluate their student meant that one of the initial Tier 3 objectives (i.e., to get teachers to think of assessment as something that is embedded in everyday activities, more than a standardized test) was already met.

In the next section, we describe the process we used with the teachers in the original pilot to develop the multi-grade LP and course-specific trajectories for the SLO process. We begin the next section by providing an overview of research-based learning progressions available in the visual arts, since this information provides context for the resources used to help teachers build the multi-grade LP. We then describe the process the teachers used to develop the course-specific trajectories using the multi-grade LP as reference and describe the resulting outcomes from the work.

[3] Although we did encourage teachers to consider some of the items in the interim assessments for their SLOs, all of the teachers opted to focus on using their own performance assessments to evaluate their students. We intend to map the pre-and post-assessments next year since the district has set policy that available interim assessments must be used as part of a teacher's SLO body of evidence.

Developing Visual Arts Learning Progressions for SLOs

Process for Developing the Multi-Grade Progressions

Unlike other core content areas such as mathematics (e.g., see Confrey, 2012; Clements & Sarama, 2009) and the sciences (e.g., see Catley, Lehrer & Reiser, 2004; Smith, Wiser, Anderson, & Krajcik, 2006), hypothetical or draft LPs have not been developed in the visual arts as part of a researched program of study in the United States. That is, drafts have not yet been developed in the visual arts that meet the following definition of an LP put forward by Masters and Forster (1996): "[a LP] describes the knowledge, skills and understandings of a learning area in the sequence in which they typically develop and provides examples of the kinds of performances and student work typically observed at particular levels of attainment" (p. 4).

In the United States, a few state departments of education (e.g., see Colorado Department of Education, 2009; New Jersey Department of Education, 2009; Washington Office of Superintendent of Public Instruction, 2011) use their grade level or grade-range standards as signposts for sequencing arts curriculum and refer to those as "vertical progressions." Although the standards and associated benchmarks may serve as helpful markers of visual arts concepts that need to be addressed in each grade range or grades, these standards lack an important characteristic of LPs noted by Shepard, Daro, and Stancavage (2013) in being able to provide substantive markers to teachers of increasing proficiency acquired by students in an area. As noted by Whitaker (2006), "there is a need [in the visual arts field] to develop a [LP framework] for teachers where student learning [in the visual arts] is broken down into smaller and more easily identifiable stages" (p. 2).

In contrast to the United States, the Quebec Educational Program (2009) and New Zealand Ministry of Education (2007) organized visual arts progressions according to critical arts processes that increase in sophistication as students developmentally progress in their exposure to and manipulation of art media. The Quebec and New Zealand progressions encompass multiple grades and describe the knowledge and skills that students move through as they acquire competency in core areas of the arts. In both Quebec and New Zealand, the core areas also correspond to key arts processes captured in the new national visual arts standards in the United States (National Coalition for Core Arts Standards, 2014): performing and presenting, reflecting or critiquing, and creating. For example, in New Zealand four multi-grade progressions were

developed according to four core process areas identified by a panel of experts as key "progress indicators":

1. Developing practical understanding and skills.
2. Developing ideas to communicate an intention.
3. Viewing, analyzing, and responding.
4. Viewing and researching.

The set of attributes, or characteristics that distinguish the developmental stages that students move through as they acquire more complex skills was developed for each progress indicator and distinguished by different levels that could either span specific grades or grade ranges. For example, Figure 8.1 shows the attributes identified for each of the four levels of the viewing, analyzing, and responding progress indicator for all primary grades in New Zealand:

Figure 8.1 Visual arts LP for one progress indicator, New Zealand

Level	Viewing, Analyzing and Responding
Level 1	Talks about their own and others' art works.
Level 2	Describes how and why meaning is conveyed in their own and others art works.
Level 3	Identifies the means by which selected objects and images communicate different kinds of ideas.
Level 4	Explores and describes how the use of different media influences the communication of ideas in selected art works.

The developmental aspects of this progression surface in the characteristics captured at each level. That is, in order for students to engage in the process of viewing, analyzing and responding to art works at Level 4, arts educators in New Zealand determined that students would need to be able to demonstrate the ability to carry out the expectations specified in all of the prior levels first. For instructors, this multi-grade progression provides an opportunity to design instructional objectives that need to be met for this progress indicator at each level. For example, the work for a Level 2 class would require that arts educators focus their instructional

efforts on ensuring that by the end of the course students acquire facility with describing how and why meaning is conveyed in their own and others' art works.

This same developmental model was used in Quebec for two key process areas or "competencies" for art where two separate multi-grade LPs were created to highlight key attributes that distinguish where students were located across the developmental continuum in the primary and the secondary grades. In Quebec, the two process areas used to develop the LPs are: creating images (personally based on media) and appreciating works of art and cultural objectives from the world's artistic heritage, personal images, and media images.

The process we implemented to develop the multi-grade LPs with the visual arts teachers in the two original pilot sites at GB and Spruce used the LPs from Quebec and New Zealand as a starting point. We provided examples of the progressions developed in both locations and developed a multi-grade progression as an additional example for their review. These tools helped scaffold the work of identifying key attributes that students would need to demonstrate at each grade level or range and the work of identifying the key big idea or process areas that could be used to develop multi-grade progressions.

Appendix F presents the multi-grade progression developed by the teachers at GB. They defined the levels in their progression by course (beginner, intermediate, advanced) and defined the attributes at the end of the progression according to what they expected a student attending an Advanced Placement (AP) studio art course to demonstrate in the area of "engaging in self-expression." The choice of this big idea appears to have been influenced by one of the process areas or competencies used in Quebec (creating personal images). The teachers also inserted intermediary attributes, or objectives, which started the work of moving into and refining the course-specific trajectories developed in the second session. A third column, skills, was also added since the skills (i.e., learning and applying specific techniques to use media) being cultivated in visual arts courses are a necessary component to evaluate relative to each task.

We inserted the grade-level expectations and evidence outcomes from the Colorado Academic Standards to each of the levels in the progression to demonstrate how the attributes can (and should) be connected to a cluster of state expectations and evidence outcomes for high school students in the visual arts. Tying in a cluster of grade level expectations and evidence outcomes for each level also served as a check that the grain-size of the attributes or learning objectives for each grade was large enough to encompass the academic year. Only one multi-

grade progression was developed since the identified attributes would hold for both drawing and ceramics, but the activities and skills would naturally differ for their respective subjects since different media were being explored. According to both teachers, by the time a student reaches AP studio art at the end of the multi-grade progression, a student should be able to independently create a work of art by making deliberate choices about techniques and media employed to engage in self-expression. It is important to note here that although the GB teachers developed a draft of the multi-grade LP and made a few refinements to the draft in the second session, this draft was never finalized. Both teachers (with agreement from us and the visual arts program manager) realized that they needed to focus and spend more time developing the course-specific LPs, which were more critical to the SLO process than the multi-grade LP.

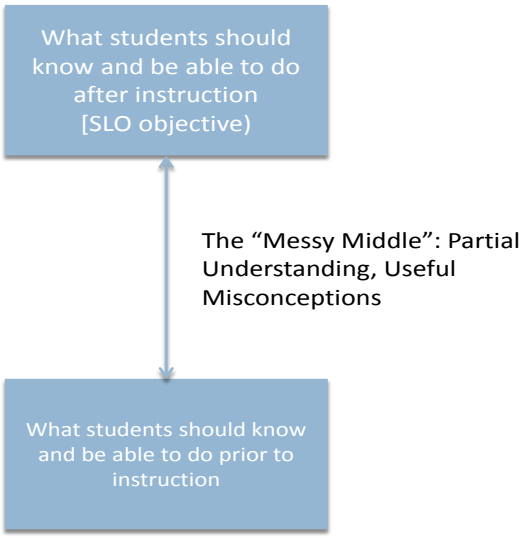
Although the teachers at the Spruce pilot site also developed a multi-grade progression, they faced tremendous difficulty identifying the key attributes or learning objectives that identify different stages of learning along the big idea of interest. They started the process by identifying the activities and skills they taught in grades K-8. Although they stated in their evaluations that they enjoyed the collaborative time (i.e. as indicated by one of the teachers: "this activity gave us a chance to see what the other person does with those same grades!"), their multi-grade progression contained a number of disconnected activities and skills with many attributes missing for each level specified. In their final session (session 2) participating in the original pilot, the Spruce teachers received feedback on their multi-grade progression from the visual arts teachers at GB. The GB visual arts teachers noted that they did not understand the order of activities and skills connected to the attributes presented in the Spruce teachers' LP. Additionally, the GB teachers noted that they could not see the progression in the attributes leading to the final level. Following the November joint session with teachers at GB, the Spruce teachers opted out of the original pilot and decided to join the teacher leader workshop pilot to reengage with the SLO process through the units of study development work. Hence, the multi-grade progression developed by the Spruce teachers is not included in this report.

Process for Developing the Course-Specific Trajectories at GB

The two visual arts teachers at GB were given the option to either develop one SLO and one SGO or to develop two SLOs. Both teachers opted to develop two SLOs, and this decision required that the teachers develop two separate course-specific trajectories each for their SLOs.

The diagram in Figure 8.2 provides a conceptual illustration of the course-specific trajectory used with the teachers. Similar to the process used to construct the multi-grade progression, the teachers were asked to identify the key attributes that separated the *lower anchor*[4] (i.e. the bottom box of Figure 8.2), or starting point of the progression, from the *upper anchor*[5] (i.e., the upper box of Figure 8.2). Although they were directed to specify the lower and upper anchors of the course-specific trajectory by using the attributes located on the multi-grade progression, the drawing teacher decided to modify the attributes for the upper anchor for one SLO. The visual arts teachers were then asked to identify the "messy middle" or the levels in-between that would lead to the upper anchor. As indicated by Figure 8.2, the attributes identified in the upper anchor represent the SLO learning objective.

Figure 8.2. Illustration of the course-specific trajectory



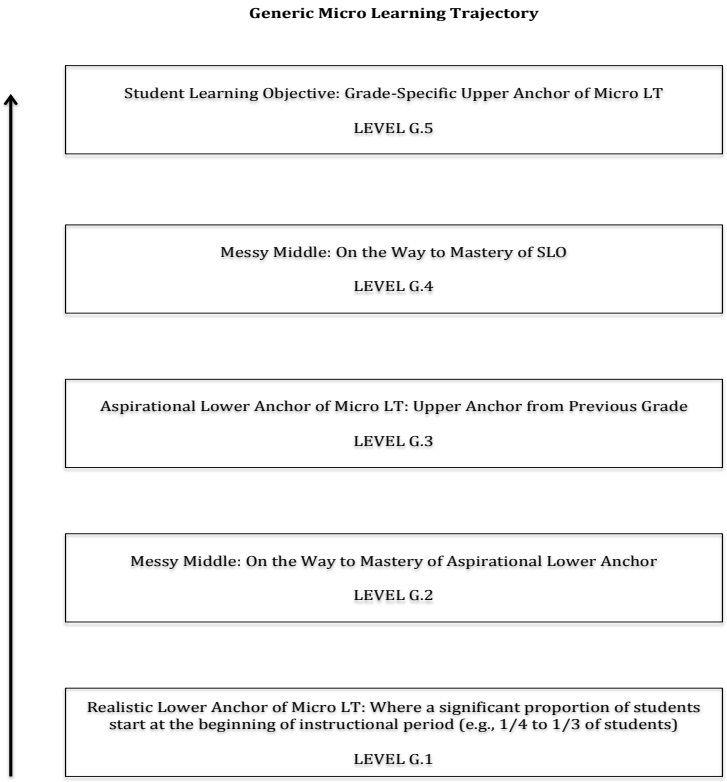
Both teachers opted to develop SLOs for two different courses: a beginning class and an intermediate or advanced class. Different classrooms were selected to test out how student growth could be evaluated for groups of students situated at different locations on the multi-grade progression. Since both teachers taught different sections of the beginning class, they each selected one beginning class to track and monitor over the course of the school year. Similarly,

[4] The lower anchor is referred to as the *lower aspirational anchor* in the math pilot. We maintained the use of the term lower anchor in the visual arts pilot since we did not want to change the terms already introduced to our teachers in session one.
[5] The upper anchor is referred to as the *aspirational learning target* in the math pilot.

both taught different sections of the intermediate or advanced courses and chose one higher-level course to track and monitor students.

However, after reviewing the performance profiles of their students in the beginning class, the teachers realized that they were working with students that had considerably lower levels of knowledge and skills and were well below meeting course expectations at the beginning of the year. Figure 8.3 presents a generic learning progression developed by the math pilot that depicts the two additional levels that are located below the lower anchor. The math team developed this example since their teachers also indicated that a significant number of their students started well below grade level expectations.

Figure 8.3. Generic learning progression



For the beginning visual arts courses, the teachers added two additional levels: the *realistic lower anchor* in the general progression in Figure 8.3 was referred to as *not ready for course* by the visual arts teachers and the *messy middle: on the way to mastery of aspirational lower anchor* was referred to as *almost ready for course*. The terminology differed for these two lower levels from the math pilot because the course-specific trajectories were being defined solely in

relationship to the course itself and were no longer wedded to the information in the multi-grade LP. For example, in the math pilot, the realistic aspirational lower anchor would correspond to a level located somewhere in the multi-grade LP. However, since this "anchor" was not defined by the multi-grade LP, the visual arts teachers felt comfortable defining their own set of descriptors that would differentiate their not ready for the course students from their almost ready students, and from students at the lower anchor.

For the drawing teacher, since all of the students entering the intermediate arts course had experienced the beginning arts course, the lowest level of the course-specific LP started at the almost ready for the course level. However, the ceramics teacher started her intermediate arts course progression at the not ready for the course level since the school had enrolled a group of students that had no prior exposure to ceramics in her intermediate class and she was given "no choice" but to accept their enrollment in her class.[6]

The process of developing the course-specific trajectories was highly iterative, and involved multiple revisions based on reviews of student work and input provided by us, the visual arts program manager and a DPS assessment coach with an arts background. Final copies of their course-specific trajectories are located in Appendix G.

Refining Selected Tasks and Rubrics

After the progressions were largely established, the teachers worked on identifying and selecting existing performance tasks that they thought best evaluated the instructional target, or attributes, and the set of skills they defined at each level on the progression. In contrast to the mathematics pilot, we had the teachers engage in task refinement rather than task development. Considering the limited amount of time available for this pilot, we opted to spend more time engaging in the process of talking about what they were learning about their students based on how they were performing on each task and developing performance criteria to differentiate levels of performance on each task. In session 3, we spent a half day introducing assessment quality criteria to the teachers and practiced using the criteria on a selected task to demonstrate how they could refine their other tasks to meet quality criteria expectations. The activities and tasks used to place students along the course-specific trajectory were mostly developed by the visual arts teachers independently. The assessment coach from ARE with an arts background

[6] Next year, the ceramics teacher will impose the requirement of taking beginning ceramics before enrolling in intermediate ceramics.

worked with us to help the teachers check on the match between selected tasks and the progressions.

Although we revisited the assessment quality criteria pertaining to scoring in session 3 and depth of knowledge (DOK) in session 5, each of these assessment quality topics could only be treated lightly during the workshops. After being introduced to the SLO Assessment Quality Check Tool (see Appendix H) in session 3, the GB visual arts teachers shared prompts of visual arts tasks they administered during the fall semester and used to place students in baseline performance groups. They also shared rubrics for each task and qualitatively discussed corresponding student work samples for each level of the rubric. The GB visual arts teachers then used the tool to identify some areas for improvement. The drawing teacher identified some areas of his scoring rubric he wanted to change. The ceramics teacher wanted to be more explicit about the clarity and focus of her task and to re-word language that may be difficult for English Language Learners.

In the first part of session 5, we devoted additional time to looking at tasks using the SLO Assessment Quality Check Tool. In advance of the workshop, both GB visual arts teachers shared copies of a task and rubric used to evaluate students in their course-specific trajectory. The drawing teacher shared the "Charcoal Drawing of an Object" task and corresponding rubric located at the upper anchor of his course-specific trajectory for the Beginning Drawing course (see Figure G.1 in Appendix G). The ceramics teacher shared the "Final Research Paper and Sculpture" task and corresponding rubric from the upper anchor of her course-specific trajectory for the Intermediate Ceramics course (Figure G.2 in Appendix G). During the workshop, we reviewed both tasks and rubrics using the SLO Assessment Quality Check tool. Both teachers' tasks met criteria in the areas of "Clarity and Focus", "Content", and "Fairness", but did not meet any of the three criteria defined under "Scoring". Both teachers noted modifications they would like to make to the tasks in the future but did not modify the tasks during the pilot period because the tasks were already in the process of being administered. After the workshop, we provided detailed written feedback on the tasks to the teachers and drafted sample rubrics for components of the tasks that were scored ambiguously. Copies of the two tasks, rubrics, and written feedback from the facilitators are included in Appendix I.

In contrast to the mathematics pilot, one major drawback to the visual arts pilot around scoring or making inferences about student performance (i.e., the third vertex on the "assessment

triangle") was the fact that our teachers could not share performance tasks for their SLOs and engage in score moderation work on scores assigned to students since the skills being evaluated in their respective visual arts specialty areas (drawing and painting vs. ceramics) differed significantly. The teachers were able to experience a score moderation process during session 1 when they evaluated samples of student work together, but could not compare the scores they assigned to students on their SLO tasks. Without having the teachers engage in a score moderation process with their SLO tasks, there was no external check on the scores assigned to each individual student. In the next section, we describe how the GB visual arts teachers used the tasks in conjunction with the course-specific trajectory to create performance groups to track student growth within their courses.

Establishing Baseline Performance Data

In sessions 3 and 4, GB visual arts teachers used data sources available from the first semester to place students into baseline performance groups for the beginning of second semester courses. The purpose of establishing the baseline data was to define where students were located on the progression at the beginning of the course and calculate the amount of growth achieved by students at the end of the year. After identifying the data sources, the teachers created four baseline performance categories that were associated with levels of the course-specific trajectories. These four baseline categories were "extremely ready for semester 2 course" (the messy middle), "ready for semester 2 course" (the lower anchor), "almost ready for course" (between the lower anchor and the lowest possible level of entering students), and "not ready for course" (the lowest possible level for students entering the course).

Following the selection of data sources used to establish the baseline categories, the teachers determined the performance expectations required on each different data source they selected to place students into a baseline performance group. The teachers also examined and referred back to student work from the baseline data sources to check the placement of students into baseline categories and made adjustments to performance group assignments if they felt that their initial assessment of students needed to be modified.

Defining Performance Expectations on Tasks to Determine Location on Progressions

After reviewing tasks from the GB visual arts teachers using the SLO Assessment Quality Check Tool in session 3, we asked the teachers to identify the set of tasks they were planning to use to evaluate students on both SLOs. Similar to the approach used to evaluate a body of evidence to place students into baseline performance levels, the visual arts teachers developed performance expectations for students using selected second semester tasks to locate students on the course-specific trajectories by the end of the instructional period (i.e. the end of the spring semester). Figures 8.4 and 8.5 identify the tasks, or SLO data sources, chosen by the GB visual arts teachers and their articulated performance expectations on the tasks for each level of the four course-specific trajectories.

Figure 8.4. SLO performance expectations for drawing courses

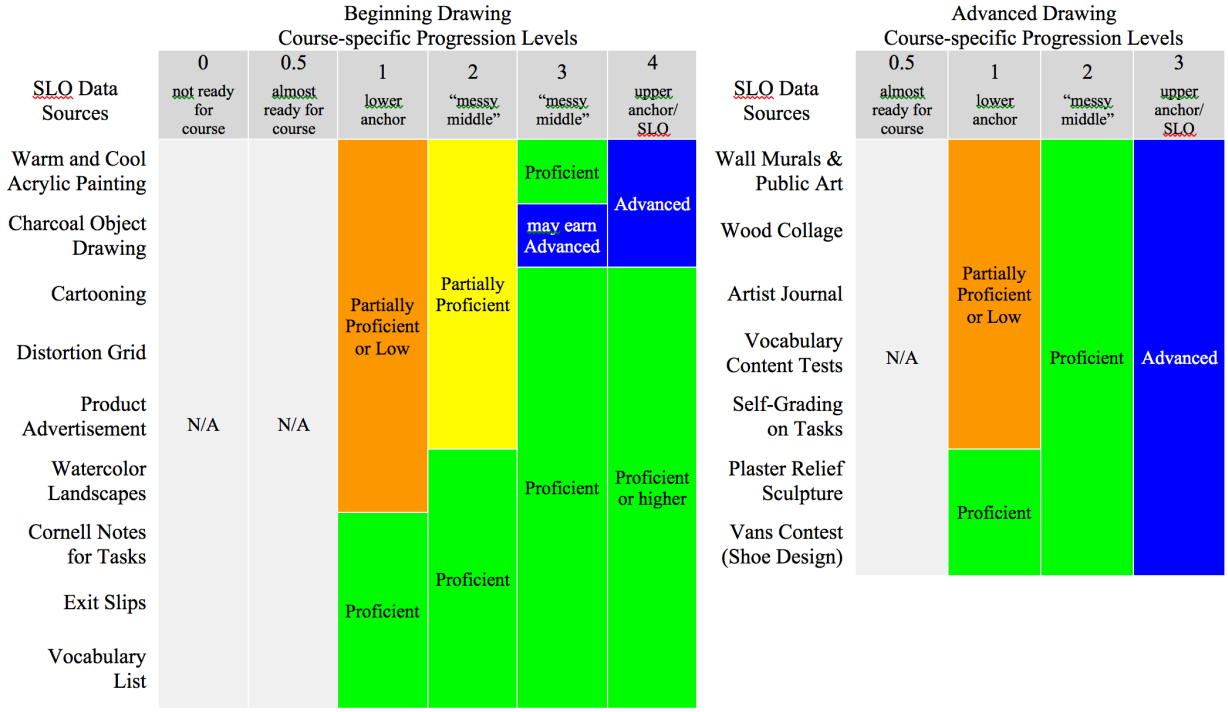
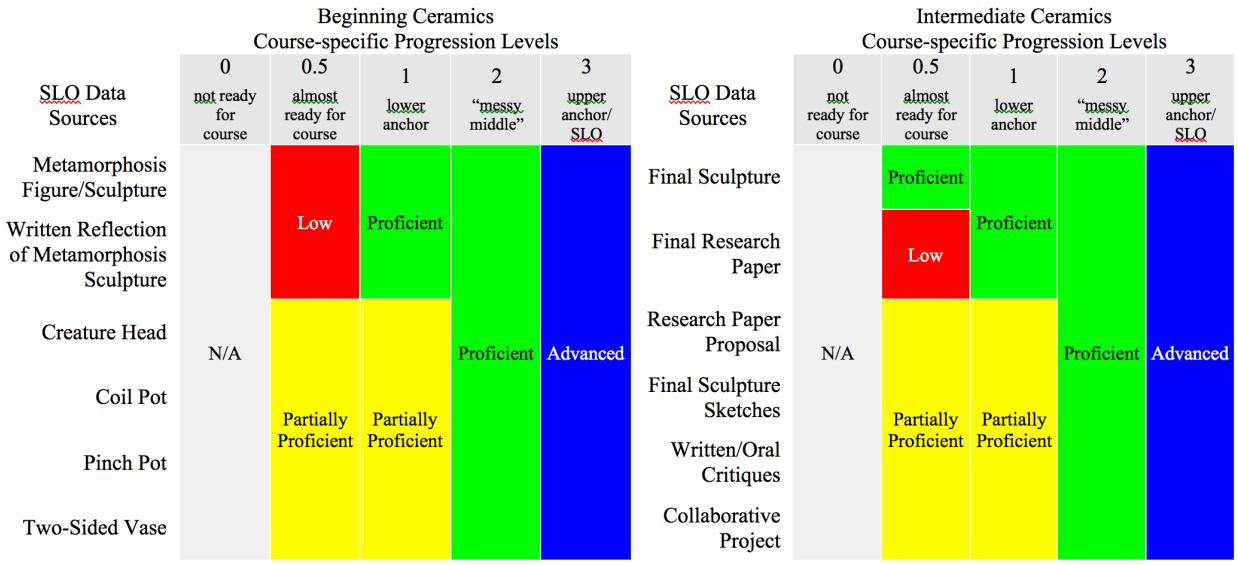


Figure 8.5. SLO performance expectations for ceramics courses



Like the baseline evidence sources, all of the SLO data sources are listed in the activity dimension of the course-specific trajectories. Unlike the baseline evidence sources, several of the SLO data sources assess students at the highest level of the course-specific trajectory (the SLO) to determine whether students met the learning objective.

The visual arts teachers were asked to independently check their tasks to ensure that they met the criteria in the SLO Assessment Quality Check Tool, and had clearly defined performance levels and expectations in the scoring rubrics. Tasks toward the top of the course-specific trajectories were given more weight than those near the bottom of the LP because tasks at the top of the course-specific trajectory required advanced applications of the skills acquired from the lower levels.

In contrast to setting baseline data, which relied on previously administered tasks, creating a body of evidence for the SLO required the existence of well-planned activities that would be administered during the period of instruction. The drawing teacher selected tasks from his preexisting task bank. Although he wanted to modify some of the rubrics after our discussion of quality criteria, he did not implement modifications during this pilot year. The novice ceramics teacher did not create her tasks in advance of selecting them for the SLO body of evidence. She had a general idea of activities that she wanted to do with her students. However,

the curriculum and tasks that she drafted in advance of the pilot did not necessarily align with the course-specific trajectories that she later developed, nor was she able to draw on the resources of other ceramics colleagues for pre-existing well-developed tasks. For these reasons, her SLO data sources were modified after the course-specific trajectories were finalized in session 4. For example, she originally planned to administer a task that required students to make a dinnerware set but instead developed a collaborative ceramics project that was more clearly aligned to the course-specific trajectory.

Additional time on visual arts task development would probably have helped both teachers craft higher quality tasks with enough time to administer them. For both teachers, the performance expectations defined on the various tasks to determine where students were located on the trajectories carried assumptions that they would hold across all of their students. Although we questioned whether these assumptions could hold across tasks, the teachers strongly believed that the skills acquired during the semester are cumulative in nature, and that advanced students in their arts courses consistently perform at advanced levels on all assigned tasks.

Scoring Student Growth

After assigning students to baseline performance groups and determining their location at the end of the course, the teachers were asked to come to a consensus on scoring, or quantifying, the growth made by their students. To quantify student growth on the SLO at the end of the period of instruction, we worked with the GB visual arts teachers to develop a scoring scheme that assigned points to each possible movement on the course-specific trajectory. First, we shared a hypothetical value table with the GB visual arts teachers in session 4. In this value table, we assigned points to each possible upward movement on the course-specific trajectory, giving more weight to movements spanning more levels of the course-specific trajectory. For example, a student who began the period of instruction at level 0 (not ready for the course) but ended at level 3 (upper anchor) was given 6 points, but a student who began at level 2 (the "messy middle") but moved to level 3 (upper anchor) was awarded 2 points. In addition, movements that we hypothesized were less difficult were given less points. For example, the movement of a student from level 0 (not ready for course) to level 1 (ready for course) was assigned 0.5 points since there were very few skills that the students needed to learn.

After introducing the hypothetical value table to the GB visual arts teachers, we invited them to create their own value table. We expected the teachers to disagree with our assessment of the difficulty of moving between levels. Instead, both visual arts teachers felt that the hypothetical value table fairly assigned points and chose to adopt the hypothetical value table without modifications to measure student performance on the SLO. This value table is presented in Table 8.2 and describes the amount of points awarded for each movement on the course-specific trajectories. If a student remained at the same level or moved down a level of the course-specific trajectory, we chose to assign 0 points to measure that student's growth.

Table 8.2. Scoring Scheme for Measuring Student Growth

Baseline Location on Course-specific progression	End of Course Location on Course-specific progression			
	Level 1 lower anchor	Level 2 "messy middle"	Level 3 "messy middle"	Level 3/4 upper anchor/SLO
Level 2 "messy middle"	0	0	1	2
Level 1 lower anchor	0	0.5	2	4
Level 0.5 almost ready for course	0	1	3	5
Level 0 not ready for course	0.5	2	4	6

Categorizing Teachers on the SLO

Before transforming the points from quantifying the student growth achieved into teacher categorizations, we shared the SLO rating expectations for the pilot with the SLO team at Department of Assessment, Research and Evaluation. Table 8.3 describes the teacher-rating scheme used in this pilot. We distinguished between three types of possible growth associated

with each rating: aspirational growth, acceptable growth, and minimal growth. Aspirational growth was defined as student attainment of the SLO regardless of baseline position. In the SGO system, aspirational growth was the only type of growth that could be measured. Acceptable growth occurred when the average (median) student growth score in a class reached 2 or above. The value of 2 was selected since this score is associated with moving at least 2 levels up for the two lowest levels and for meeting the objective for the highest baseline group. Minimal growth was defined as a student moving up one level of the course-specific trajectory regardless of baseline position. Although we shared the definitions of aspirational, acceptable, and minimal growth with the GB visual arts teachers in session 4, we did not share the descriptions for the four teacher ratings described in Table 8.3 until session 5. Our early conversations did not attach percentages to the number of students meeting the SLO as described in Table 8.3.

Table 8.3. SLO Ratings

Ratings	Description
4	80% of students met the SLO and at least 50% made minimal growth
3	the median score equals or surpasses the definition of acceptable growth
2	50% of students made minimal growth
1	less than 50% of students made minimal growth

Both visual arts teachers were rated on two SLOs. Figures 8.6 and 8.7 distinguish between students attaining minimal, acceptable, and aspirational growth and summarize the teacher ratings. To arrive at individual teacher ratings, we first computed a growth score for each individual student using that student's baseline level of the course-specific trajectory, position on the course-specific trajectory at the end of the period of instruction, and the point scheme described in Table 8.2. We then computed the average (median) score for a class using the individual student growth scores. We used the definitions from Table 8.3, the median score, percentage of students making minimal growth, and percentage of students making aspirational growth to obtain final SLO ratings for each GB visual arts teacher. The results from Figures 8.6 and 8.7 were reported to DPS.

Figure 8.6. Distribution of student growth levels for all visual arts students

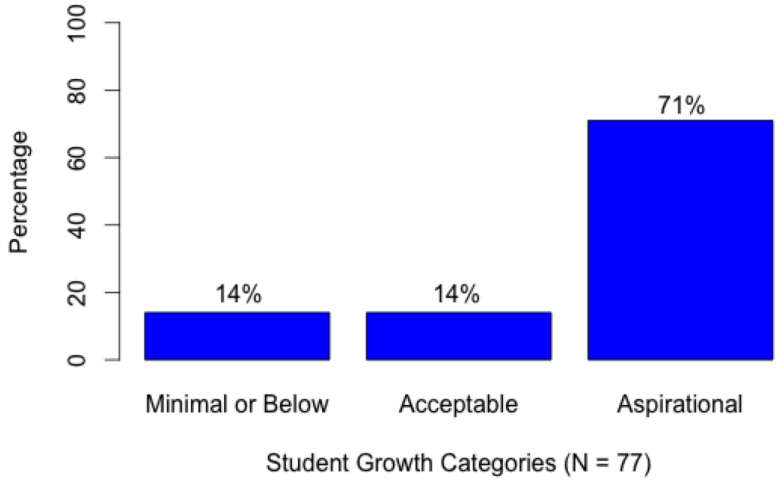
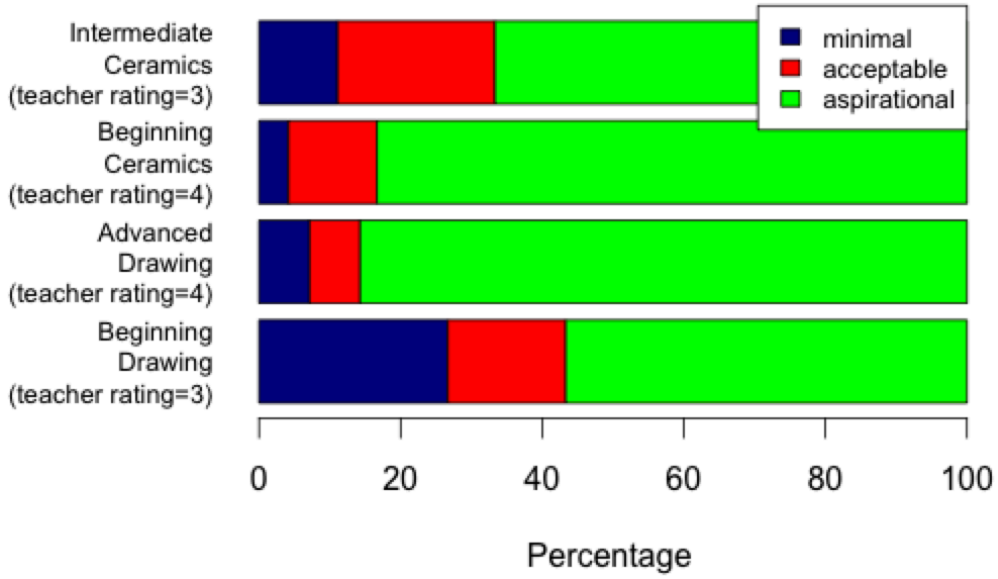


Figure 8.7. Distribution of student growth levels by visual arts course



An interesting point to highlight from this pilot is that under the old SGO system, both teachers would have only achieved one objective since only one class for each teacher met the aspirational growth standard of 80 percent. According to both teachers, they were pleased to see that growth could still be rewarded under this scoring and rating configuration. Their focus

during this pilot became less about whether all students met the objective, but instead whether all of their students "were showing growth and showing evidence of learning" (personal communication, May 30, 2014). As noted by the drawing teacher, "all I used to care about in the past was whether I met that 80 percent objective, and now this approach makes me care more about whether all of my students are growing [and]…what can I do to help my students grow." (personal communication, May 30, 2014).

In the next section we share reflections about the original Tier 3 pilot from our participants from GB and the visual arts program manager. Some of the concerns noted in the reflections are addressed and incorporated into a key set of recommendations that follow to improve upon the content and structure of these workshops in anticipation of the 2014-2015 school year.

Reflections on the Original Pilot

We gathered participant feedback and input based on content covered at the end of each workshop session using an evaluation form, through an exit interview conducted with the participants at the end of the pilot, and via the Professional Development Unit (PDU) peer review session held at the end of the school year with guest participants in attendance. The guest participants in attendance at the PDU review session were the current president of the Denver Classroom Teachers Association and two members of the PDU ProComp team at DPS. This section highlights three key challenges and three key successes from the pilot.

Challenges

- Lacking adequate time and consistent contact hours for undertaking Tier 3 work.

Considering the range of topics and content covered and the fact that both teachers said they had not been exposed to many of these topics (e.g., LPs, DOK, considering alignment of tasks to the LP levels, and to the Colorado Academic Standards for visual arts), both teachers indicated that more contact time was probably needed for this pilot. That is, a total of 40 hours dedicated to the pilot was insufficient, and the sessions should be held closer in time to ensure that less time was spent at the beginning of each session revisiting topics from the prior sessions.

However, both teachers also noted that it would have been difficult to build in more contact time for the pilot since other mandatory professional development and district meetings consumed their schedules.

- Needing additional guidance on refining assessments and developing scoring criteria in the rubrics.

Both teachers indicated that they needed more guidance and assistance in refining their tasks and developing strong rubrics and scoring criteria for their tasks. Although we had spent some time refining one rubric and task with the teachers, they noted that they would have liked to go through the same process with all of the tasks and rubrics they used for this pilot.

- Needing guidance on defining the LPs.

Both teachers found the process of developing and refining the progressions over the year to be difficult and not intuitive. The fact that both teachers were asked to continuously refine and also to switch the order of a few levels specified in their progressions revealed that both teachers underwent a significant learning curve to engage in this process. Furthermore, both teachers indicated that the multi-grade progression created in the first session was perhaps the most challenging progression to develop since this required having to think about what different stages of mastery of the big picture ideas would look across the different courses taught. Both teachers said they had not engaged in this type of thinking in the past and encountered some difficulty envisioning the key milestones that students need to demonstrate as they moved toward the final learning objective.

Successes

- Using the course-specific trajectories to guide the instructional planning and evaluation of their students.

Both teachers indicated that the descriptions of performance, the set of activities planned to evaluate students on each level, and the set of skills being evaluated made the instructional steps they wanted to take with their students this year more explicit. That is, they were able to prioritize how the tasks should be sequenced and determine the set of skills that needed to be taught prior to moving into higher-level tasks. The veteran drawing teacher who was skeptical

about the pilot in the beginning of the year noted that "if the district sticks with this approach, this would be really helpful for a lot of teachers…because it helps teachers organize their instruction and also helps them focus their evaluation of students" (personal communication, May 30, 2014).

- Focusing on student work as a central feature of the pilot and the involvement of students in understanding expectations for performance.

According to both teachers, the focus on student work made the performance criteria for each level on the progression clearer and provided them with the means to better articulate expectations for moving all students toward the learning objective. Both teachers noted during the PDU review session that their students knew which tasks would "count more" than others, and also understood what they were being evaluated on for each task. For example, for the beginning drawing course, the charcoal drawing of an object task carried more weight in the assessment of students than the vocabulary test, and students understood that their ability to express and correctly identify different shading values and depth would be the focal area of their evaluation on the charcoal drawing task. Interestingly, both teachers stated that the criteria being used to evaluate student progress made the evaluation process "less arbitrary" to students and that any arguments about how students were performing could be quickly addressed by going over the performance expectations and criteria set for each task.

- Focusing on the growth of individual students as opposed to focusing on whether the SLO objective percentage target was met by the entire class.

Both teachers noted that one of the highlights of being part of this pilot was that the growth made by students who started well below mastery (e.g., had no arts exposure during their middle school years) could be tracked and acknowledged. That is, their focus on the SLO process was on how and whether each and every individual in their classroom was making progress and what they needed to do to guide students to move to higher levels on the progression. The veteran drawing teacher who had developed SGOs since ProComp was implemented noted that this was the first year that he "didn't focus on whether [his] class met the 80 percent target, but focused on whether each student was demonstrating progress…and what [he] needed to do to help that student grow" (personal communication, May 30, 2014).

In addition to the reflections gathered from the participants, the visual arts program manager was asked to provide feedback on the pilot based on her observations and participation in each workshop. The program manager thought that this original pilot should be folded under the teacher leader workshops to ensure that a critical mass of teachers is reached through the professional development sessions offered. The program manager made the observation that at times, there was a ratio of four staff members to two teachers in each session and that both district and Center for Assessment resources were not being optimized to ensure that more arts teachers would benefit from the content.

Additionally, the visual arts program manager communicated that the two visual arts teachers needed to work with a larger group of teachers to benefit from collaborative work. This input was based on her observation during several PD sessions of a lack of active collaboration between the two GB visual arts teachers, specifically their reluctance to provide input on each other's progressions. In stark contrast, the teacher leaders in those workshops enjoyed working in teams to develop the course-specific trajectories and actively negotiated how the levels, activities, and skills should be defined. Her observation of both settings allowed her to see the benefits of a slightly larger collaborative setting and the downside of having only two teachers serviced by the original Tier 3 pilot work.

In Chapter 9 of this report, we move into describing the process used to carry out the second pilot with teacher leaders in the arts. The second pilot included two music teachers and one drama teacher. The mix of teachers participating in the workshops provided valuable insights into key differences and issues found between these disciplines.

Chapter 9: The Teacher Leader Workshop Pilot

Professional Development Structure

The teacher leader workshops were co-led by Elena Diaz-Bilello from the Center for Assessment and Capucine Chapman from DPS. The sessions touched on much of the same content as the original pilot, but aimed to integrate the LP concept with the unit of study development work and the SLO process. The decision to move forward with a different pilot involving the teacher leaders was inspired by input received from the two visual arts teachers from Spruce. Since the visual arts department had conducted a series of workshops in the 2012-2013 school year focused on developing units of study to implement in the classrooms, those teachers said they wanted to understand how that work could be folded into the new SLO process and how a LP could be used to frame the unit of study work. The two visual arts teachers from Spruce decided to leave the original pilot and join the teacher leader workshops.

All three sessions took place during the spring semester. Two of the sessions were held at Manual High School and the final session was held at DPS' Acoma Campus. All sessions were voluntary and a total of nine teachers registered to participate and earn ProComp professional development unit credits for completing the work. The sessions comprised three full-day sessions. The Arts department funded substitutes for all teachers.

The set of activities and content covered in each work session is outlined in Table 9.1. We intended the sessions to provide an overview and introduction to engaging in the new SLO process using existing units of study and to ensure that the arts teacher leaders would be in a better position to implement this new process in the 2014-2015 school year. As indicated earlier, the two Spruce visual arts teachers joined the workshops and used the opportunity to modify and refine their units of study for next year. All teacher leaders including the Spruce teachers submitted student objectives to DPS using the SGO system rather than the pilot SLO process.

Table 9.1. Activities and Content Addressed During Teacher-Leader Pilot Sessions

Session	Activities
Session 1 **January 2014**	- Overview of new SLO process as part of teacher evaluation and connection to units of study. - Overview of LPs as an organizing framework for the SLO process, and units of study. - Examining a sample unit of study and defining the learning objective connected to the study. - Group work mapping out a LP in the sample unit of study. - Identifying a body of evidence used through units of study developed for evaluating students on SLOs. - Overview of DOK and Bloom's Taxonomy and application of DOK to objective specified and the sample unit of study. - Developing a learning objective for a selected semester-long unit of study. - Peer review of learning objectives.
Session 2 **February 2014**	- Continued work on cognitive complexity and deconstructing the learning objective developed for the selected unit of study. - Understanding assessment quality criteria in the assessment review tool. - Running one assessment embedded in the selected unit of study through the review tool. - Modifying the assessment as needed.
Session 3 **April 2014**	- Example of walking through scoring rules and guidelines for considering student performance across assessments embedded in the unit of study. - Evaluating and modifying criteria used in rubrics to evaluate students on tasks/assessments embedded in the unit of study. - Walk through example of assembling body of evidence for the SLO based on a semester-long unit.

In contrast to the activities in the original Tier 3 pilot, teacher leaders focused on refining their units of study and using those units as a tool for engaging with the SLO process. The LPs were used as an organizing framework to help sequence and structure the lessons and tasks embedded in their units. Although many of the topics addressed in the original pilot were also addressed during these teacher leader workshops, the participants were not required to set baseline data or monitor student progress toward meeting the SLO. These workshops were focused on ensuring that participants understood what would be expected from them in the new SLO process without requiring them to try out the process in the spring.

Participant Description and Context

Teacher Background

The nine teachers in attendance included the two Spruce visual arts teachers, one of whom was a teacher leader. All teachers (with the exception of one of the visual arts teachers from Spruce) were teacher leaders and had several years of experience teaching in the district and serving as mentors for other arts teachers. Six of the teachers were from the visual arts, and the remaining teachers were from the performing arts (one drama and two music teachers). All of these teachers had mentoring responsibilities that required them to meet with a small group of 2-5 teachers on a regular basis both in person and on-line to help them engage in the work of developing units of study in the classroom.

Although all teachers were experienced and used units of study in their classrooms, the music teachers expressed concerns about integrating the SLO process with units of study, since the majority of teachers in the performing arts are not developing units of study. According to the visual arts program manager, the performing arts as a whole is planning to move forward with implementing units of study, but only a few teacher leaders in those subject areas (dance, drama, and music) have been exposed to the units of study development work and almost no teachers in the performance arts use units of study in the classrooms.

In the following sections we present an overview of the process used to develop the LPs into organized units of study, the process used to refine the tasks embedded in the unit of study, and the process used to strengthen the rubrics used in the units. Since the LPs were being developed for a unit of study used in a specific course, the work of constructing progressions focused only on the course-specific or unit-specific progressions. If a unit-specific trajectory spanned the entire length of the course, there was no distinction made between the unit- and the course-specific trajectories.

Developing the Unit-Specific Learning Progressions

The process of developing the unit-specific progressions took place during the first session and was identical to the process used in the original Tier 3 pilot to develop course-specific trajectories. However, one key difference in the approach used to construct the

progressions in the teacher leader workshop is that the participants were divided into teams of three and given the assignment to strengthen a sample unit of study developed by a DPS teacher on the use of illuminated manuscripts in a high school visual arts course. The sample unit failed to specify the type of tasks that students would be asked to complete to demonstrate learning in the unit and also did not provide any information on the instructional sequence that would be used to carry out the objectives. The teacher who developed the unit noted in the unit that all of this information was "located in her head and there [was] no need for her to explicitly write this information down". In their respective teams, the teacher leaders reviewed the unit together and worked to:

- Identify the learning objective that would be used to help organize the smaller unit objectives identified by the teacher who constructed the unit;
- Select and order the smaller unit objectives that were critical milestones that students need to understand prior to moving to the next level; and,
- Develop examples of the types of activities and tasks they would want students to perform to demonstrate their understanding at each level.

The performing arts teacher leaders were sub-divided among the visual arts teachers so that they could benefit from the expertise of those teachers. The "messy middle" defined for the unit-specific progression was easier to negotiate and define since the unit contained many smaller unit objectives. These smaller objectives were used to help define what the messy middle should look like. That is, teachers only had to review and agree on which of the available objectives could be consolidated and serve as a level within the progression and then order their selection. After each group developed their unit-specific progression and identified the activities and skills associated with each level, the three groups posted their results on the wall. The three groups then appraised each progression together and provided input to one another on skills/activities that should be considered. They also discussed the relevance of ordering the levels in the LP in different ways. Figure 9.1 presents an example of the attributes for the five levels specified for one of the three unit-specific progressions developed by one team of teachers.

Figure 9.1. Sample unit-specific learning progression

Level	Attributes
4	Students will be able to explain and demonstrate understanding of the process used to develop illuminated manuscripts and address their role and relevance to society.
3	Students gain familiarity with the media and techniques used to develop illuminated manuscripts by planning for and developing a draft of an illuminated manuscript.
2	Students research characteristics and elements used to develop illuminated manuscripts such as exploring color mixing (shades) and blending used as well as exploring common composition and lines used for manuscripts.
1	Students research the use of illuminated manuscripts in medieval society and focus their research on identifying, analyzing and determining the symbolism of animals, people, colors and shapes used in manuscripts.

Although the time period for this unit was not specified, two high school teachers in the group noted that assuming students meet one to two times a week during the semester, this unit would likely take an entire semester to complete. For example, they felt that the exploration of media and techniques used to develop a draft and produce a final manuscript (level 2 of Figure 4) would take a few weeks. Additionally, they wanted students to reflect upon the process they used to develop their manuscripts, address the relevance of the symbolism, and address the role of manuscripts within the context of medieval society to obtain evidence for students reaching the final level. At the final level, the last task of producing an illuminated manuscript would also provide evidence about the extent to which students understand the process used to develop the manuscripts.

During the start of the second session, the drama teacher leader asked for feedback from the larger group to develop a unit-specific progression for her unit of study. The drama teacher outlined her unit objectives and then the larger group went to work to define the SLO for the unit of study. The group engaged in a process of backwards-mapping to identify the key instructional targets that would lead students toward the learning objective. This process of defining the unit-

specific progression took approximately two hours to complete since the teachers negotiated which steps to define and the sequence of those steps.

Process of Refining Unit Tasks

The topic of refining unit tasks was addressed during the second session. The process of refining unit tasks was identical to the process used to refine tasks in the original pilot. That is, teachers first discussed each of the criteria considered on the SLO Assessment Quality Check Tool (Appendix C) and then ran the tool through one task they used within a unit of study they developed. However, a key difference between the original pilot and this session was that the teachers were willing to exchange tasks, run the tool using the other person's task, and provide input to one another on areas of improvement. In the original pilot, the visual arts teachers asked for more input directly from the facilitators rather than from one another. This could be due to the fact that one teacher was new to the profession and felt uncomfortable providing input to a veteran teacher. On the other hand, all of the teacher leaders (with the exception of one Spruce teacher) present were experienced in their profession and considered subject experts in their field. An interesting area to note is that the music teachers expressed frustration with the assumed expectation that their younger elementary school students should be accessing higher DOK levels as noted in the DOK ranges associated with the evidence outcomes of the Colorado Academic Standards for their subject. This issue is revisited in the reflections section for the teacher-leader workshops.

Process of Refining Rubrics

The process of refining the scoring rubrics took place during the final session and followed the same process used in the original Tier 3 pilot. However, in contrast to the original pilot, the teachers engaged in collaborative activities to improve their own rubrics. Two teachers broadcasted their own rubrics to get collective input from peers on ways to improve their scoring descriptors. An example of one rubric used by a teacher leader and shared with other teachers during the workshop is located in Appendix J. Some teachers noted that evaluative statements in the rubric developed by the teacher should be removed and substituted with clearer performance expectations. Other teachers in the group provided input on clarifying the criteria being defined for each level. At the end of the third session, five of the teachers stayed for an additional hour

to get individual input from both Elena Diaz-Bilello and Capucine Chapman on other rubrics they were using for different units of study.

Reflections from the Teacher Leader Workshops

As indicated earlier, these sessions were motivated by the Spruce visual arts teachers noting that the LP concept was difficult for them to tackle without understanding how it applied to their units of study. Reframing the progressions as a guide for developing their units of study helped clarify the concept for the Spruce teachers and also helped clarify ways to develop SLOs. The three key challenges and successes described below are based on input received from participants at the end of each session.

Challenges

- Need more examples and guidance to specify the learning objectives.

Considering the amount of time it took in each workshop to clarify and specify the SLO and the instructional targets, a few teachers indicated that viewing some examples of good objectives set in their respective content areas would be helpful to review. This would allow them to more easily develop their own objectives and to share these examples with their mentees.

- For music and dance, performance tasks and learning objectives will likely remain at the DOK levels 1 and 2 for the vast majority of courses offered at the elementary level.

This is not so much a challenge, but rather a realistic assessment by teachers about the frameworks used in these fields (and physical education) to instruct students relative to the instructional time available to elective areas. That is, the vast majority of courses in music, dance, and P.E. in the elementary levels are focused on developing psychomotor skills, which may correspond to a DOK of 2 at the most, but will rarely reach a 3 or a 4. According to the music teachers, they do not see their elementary school students frequently enough to move beyond a DOK of 2. Important objectives in their subjects are whether students can learn how to automate skills learned, such as reaching the right pitch or hitting the right notes for a tune before engaging students in the process of evaluating and critiquing performances. Both performing arts teachers noted that higher DOK levels may be reached in music courses offered in middle and high schools, but the expectations for students to reach those levels would vary based upon

whether the course meets regularly and whether those students enter the course above the novice level (personal communication, February 5, 2014).

- Need more time and guidance to refine rubrics and tasks.

Similar to the input received from the original Tier 3 teachers at GB, the teacher leaders noted that they would have liked more time to continue working on their unit tasks and rubrics. Since time was limited, the sessions were confined to working on just one selected task in their unit of study. All teachers noted that they needed more practice and time to get more comfortable with undertaking this task on their own and with others.

Successes

- Collaborative and group work advanced the understanding of teachers on workshop topics discussed.

The most common and positive feedback provided by all teachers was the fact that the collaborative work structured around the activities helped deepen their understanding of the topics covered by allowing them to apply the concepts and interact with each other on all topics. They also noted that they enjoyed the cross-disciplinary teams since the different disciplines within the arts provided alternative perspectives that they would not have considered within their own content domains.

- Using the unit-specific LP to structure and organize their units of study.

The second most common piece of feedback received from teachers was the value of using a unit-specific trajectory to identify the larger SLO, the accompanying instructional targets to organize the set of topics addressed in the unit, and assessments used to evaluate students. Although this backwards-mapping process was familiar to many of the teachers, all teachers noted that they have not been given many opportunities to do this type of work in professional development sessions offered by the district and they valued the time spent on developing the progressions in the Tier 3 workshops. One teacher noted in the "parking lot" used to collect input from participants at the end of the day: "I love using this [progression] approach to SLO writing and using it with the unit of study…it makes me understand better where I am (in terms of what the teacher has accomplished with students) and where my students are (in terms of

location relative to the learning objective) and where I want to lead them by the end of the course."

- Clearer understanding of DOK as it relates to the unit tasks, the standards, and the SLO learning objective.

Although all teachers had seen DOK levels noted with the evidence outcomes in their standards, none of them were provided with any guidance on interpreting or understanding the meaning of those levels. Additionally, only one of the nine teachers was familiar with Bloom's Taxonomy. The workshop session spent considerable time (almost a full day) addressing DOK levels and providing concrete examples of activities representing different DOK levels in the arts to help teachers unlock what was demanded under each level. However, although many teachers noted the value of getting clarification on DOK levels, the performing arts teachers questioned the value of this framework relative to how their discipline is structured across most schools in the district.

References

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

Catley, K., Lehrer, R., & Reiser, B. (2004). Tracing a prospective learning progression for developing understanding of evolution. Commissioned paper for the National Research Council Committee on Test Design for K–12 Science Achievement Workshop, Washington, DC.

Clements, D., & Sarama, J. (2009). Learning and teaching early math: The Learning Progressions approach. New York: Routledge.

Colorado Department of Education. (2009). Visual arts: Prepared graduate competencies at grade levels. Retrieved from http://www.cde.state.co.us/coarts/statestandards

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.

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

National Coalition for Core Arts Standards. (2013). National core arts standards: a conceptual framework for arts learning. Retrieved from: http://nccas.wikispaces.com

New Jersey Department of Education. (2009). Visual and performing arts learning progressions. Retrieved from: http://www.state.nj.us/education/cccs/progressions/1/1.htm

New Zealand Ministry of Education. (2007). Visual arts: Achievement objectives. Retrieved from: http://artsonline2.tki.org.nz/ecurriculum/visarts/ao.php

Quebec Educational Program. (2009). Progression of learning visual arts. Retrieved from: http://www.learnquebec.ca/en/content/curriculum/arts/visualarts/viscurr.html

Shepard, L. A., Daro, P., & Stancavage, F. B. (2013). The Relevance of Learning Progressions for NAEP. Commissioned paper by the NAEP Validity Studies (NVS) Panel. Retrieved from: www.air.org/common_core_NAEP

Slotnik, W.J., Smit, M., Glass, R.J., Helms, B.J. (2004). Catalyst for Change: Pay for Performance in Denver Final Report. Community Training and Assistance Center. Retrieved from: http://www.ctacusa.com/denver-vol3-final.pdf

Smith, C. L., Wiser, M., Anderson, C. W., & Krajcik, J. (2006). Implications of research on children's learning for standards and assessment: A proposed LP for matter and the atomic molecular theory. Focus Article. Measurement: Interdisciplinary Research and Perspectives, 14, 1–98.

Washington Office of Superintendent of Public Instruction. (2011). Washington State K-12 arts learning standards. Retrieved from: http://www.k12.wa.us/Arts/Standards

Whitaker, S. (2006). Developing a Progression of Student Learning for the Visual Arts. Unpublished manuscript. Retrieved from: http://esolonline.tki.org.nz/content/

Chapter 10: Reflections on Goals Achieved in Pilots, Recommendations and Next Steps

Four goals were established during the planning stages of the Tier 3 pilot for both visual arts and mathematics. These goals were as follows:

1. To promote the perspective of the SLO as a critical element of good pedagogy, more than a compliance-based activity.
2. To find out what benefits can be achieved when teachers are given support to work on the SLO process collaboratively across grades.
3. To get teachers to think of assessment as something that is embedded in everyday activities, more than a standardized test.
4. To help teachers become informed consumers of externally developed student assessments.

As the project unfolded, we found that it was primarily the second of these goals that took precedent, in part because we had relatively limited time to work with teachers on evaluating the quality and utility of their assessment results. Furthermore, this "goal" was probably better cast as an open-ended question "What range of outcomes are observed when DPS teachers are given support to work on the SLO process collaboratively across grades using a newly developed learning progression framework for student growth?"

There is some evidence that teachers in our participating sites began to see the SLO process as more than a compliance-based activity (especially when the process we facilitated is compared to the conventional SGO process). It is also possible that many of the teachers began to think of assessment as something that is more than a standardized test, and that they felt better equipped to be informed consumers of externally developed student assessments. But because we did not implement formal pre-post instrumentation or observation protocols to evaluate these goals, most of the evidence along these lines comes from anecdotes or informal observations during our training sessions. And even this evidence is equivocal because to the extent that we observed changes in teacher perspectives, it is unclear whether this can be attributed to the new SLO process we introduced, or whether this is merely a product of giving teachers more time to work together. In the section that follows, we summarize what we took away as the most important findings from the Tier 3 pilot, and we make recommendations for DPS on the basis of

these findings. Findings 1-3 and their associated recommendations are intended to reflect short-term decisions that will need to be made as the SLO system is developed. Findings 4-5 also apply to short-term decisions, but have the greatest relevance with respect to longer term decisions should the district attempt to scale up this approach.

Finding 1: Choosing and Creating a Fully Elaborated Learning Progression from Scratch is asking too much of Teachers

The development of a multigrade Learning Progression along with course-specific trajectories is a time-consuming endeavor. At two of our math sites, Blue and GB, teachers chose the topics for their Learning Progressions first and then used existing standards documents to create grade-specific anchors. Although this first task was accomplished relatively quickly (within a few hours during a single session), it took multiple sessions to create course-specific trajectories for each teacher that were intended to include well-defined criteria for levels in the "messy middle." Even at the one math site (Spruce) that decided to use a previously established LP from research, multiple sessions were spent elaborating the course-specific trajectories. This was equally true among the visual arts teachers from GB who remained in the original pilot. Even without needing to establish a multigrade LP first, each teacher's course-specific trajectory went through many time-consuming iterations.

The reason the process is so time-consuming is that teachers typically have a very hard time delineating the "messy middle" level of sophistication that distinguishes a student who may have mastered some, but not all of the standards that collectively comprise the upper anchor of a course-specific trajectory. For some teachers this might be because they are inexperienced; for others it might be because they have not previously focused their attention on how students are thinking about the one big picture idea that is the basis for the learning progression. But even for experienced teachers who are typically much attuned to how their students come to know and understand key concepts in their discipline, it can be difficult to reconstruct a perceived order in which students learn certain knowledge and skills from memory, let alone to come to a consensus about this with other colleagues.

During the math and visual arts sessions, we spent nearly half of the PD time working on the multi-grade LP. Even with this time investment, the LPs that we created were problematic

because they were not based on a careful analysis of how student reasoning becomes more sophisticated over time. Such analysis is not what teachers are trained to do and it takes more work than teachers have time for in just a few PD sessions.

Recommendation 1: *Use existing LPs from the standards documents or the research literature, and do not attempt to flesh-out the messy middles right away.*

We recommend that schools use existing, research-based LPs. If no LP is available for the topic that the school wishes to focus on, then the school should use the general approach described in Chapter 2, in which the "messy middle" is initially left unspecified. Over time, teachers can fill-in the messy middle through collaborative score moderation activities. As teachers work together to interpret student work, they can make hypotheses about how student reasoning develops over time in the messy middle, and through this process can begin to fill-in the messy middle with concrete statements about student reasoning. This also creates a more authentic process in which teacher-input and interpretation of student work is key in the development of artifacts used for the SLO process.

Finding 2: Learning Progressions that span multiple grades or courses may be inappropriate at the outset in subjects without a well-established curriculum and content standards that cannot be vertically aligned

Of all the subjects taught in grades K-12, mathematics is probably the one with a curricular structure most amenable to the use of learning progressions because there are many core ideas and concepts that are introduced, reinforced and integrated into new and more complex ideas and concepts. The elective courses offered in visual arts during middle school and high school seem to represent the opposite extreme. There is not a standard visual arts curriculum across the district, nor are students expected to take visual arts courses in a clearly defined sequence, making it almost impossible to delineate what, for example, a 10^{th} grade student should know about visual arts relative to a 9^{th} grade student. In addition, to the extent that content standards have been developed for the visual arts, they have been written in such a general manner that they are very difficult to operationalize as performance indicators. This reality is likely to be shared to a great extent by other elective courses offered in middle school

and high school (with foreign language courses as the exception), and to a lesser extent by courses in English Language Arts and social studies. This is not meant to imply that there is no value in developing multigrade or multicourse LPs in subjects outside of math and science, only that the investment to do so may be prohibitive in the short-term.

Recommendation 2.1: Develop course-specific trajectories for the elective areas in year 2 (and beyond) as a framework for constructing the SLOs.

Considering the difficulty for many elective areas to develop a multi-grade LP when common curricula and resources are not readily available to help guide the development work, the multi-grade LP should be tackled as a longer term project. Instead, teachers in elective areas should focus on developing course-specific or unit-specific trajectories as a basis for their SLOs. The visual arts in DPS can provide a model for approaching this LP work with teacher leaders in other elective areas. In the visual arts, a group of teacher leaders were first exposed to course-specific and unit-specific trajectories to help facilitate the process of backwards mapping instructional activities and assessments. Currently, a few select teacher leaders in coordination with the visual arts program manager are now developing a multi-grade LP example to help facilitate the process of developing unit- or course-specific trajectories for the visual arts. The LP being developed in the visual arts can provide a basis for which teacher leaders can develop a stronger awareness of the expectations for students in specific areas of the content as they move from the lower to the upper grades. The LP can also be further refined by having teacher leaders evaluate the LP and provide their input on the attributes and skills that define each level.

Recommendation 2.2: Consider establishing common scheduling expectations for the elective areas across schools.

One area that should be considered by policymakers in the district is that elective areas, such as the arts, are challenged by the considerable variability of schedules used across schools to engage students in these subjects. For example, all four pilot teachers pointed out that each school makes the decision on how much exposure to the arts students should get during the school year. This exposure can range from daily exposure to none for some schools. Within Spruce, elementary students received minimal exposure to the arts, but the teachers interacted more regularly with middle school students in some

grades in specific years. The GB teachers noted that some students enrolled in their beginning courses had taken their last art course in the elementary grades. The imposition of common visual arts standards on teachers assumes or implies that schools are providing similar or the same levels of exposure to the arts across students, but since this is hardly the case, it seems unrealistic to expect that all teachers can work toward meeting the same set of standards due to these scheduling arrangements that are unique to each site. According to the visual arts program manager, this scheduling challenge is not confined to the visual arts, but is a considerable challenge that cuts across many elective areas.

Recommendation 2.3: Work on developing guidelines for translating and interpreting the standards for elective areas.

Standards for elective areas in the performing and visual arts were deemed to be difficult to interpret. Further, the evidence outcomes are not necessarily deemed to be the right set of outcomes that all teachers are expected to address. Developing a common language and interpretation for understanding the grade level expectations will be helpful so that teachers are clearer about the set of evidence outcomes that they need to meet for grades K-8 and for high school. Providing clarity around interpreting the standards will also help teachers better identify the cluster of grade level expectations and evidence outcomes that are being addressed by the learning objectives and the instructional targets they develop for their SLOs.

Finding 3: Teachers Recognize the Potential of the new SLO Approach to be an Improvement on the SGO Process

Although they worried about the amount of new work that it could entail, most teachers at our pilot sites appreciated that the new SLO approach had great potential to benefit them in ways that the previous SGO approach did not. For many participating math teachers, making vertical connections across grades was a clear advantage of the approach. Teachers in earlier grades (lower levels of the LP) would now have a better sense for what teachers in later grades (higher levels of the LP) were expecting of students; teachers in later grades would be able to track student performance on assessments given in earlier grades on the same big picture topic.

For some math teachers, a key advantage of the new approach was the fact that each course-specific trajectory made it possible to show growth even for students who do not reach the upper anchor. For visual arts teachers, specifying the course-specific trajectory helped them to better organize their units of study. Finally, and notably, teachers in both math and visual arts saw the potential for the new SLO approach to provide information to them during the course of the year that could be used formatively to improve instruction.

Recommendation 3.1: Consider developing only one SLO with pilot teachers.

One strategy for easing the SLO implementation burden for new teachers participating in the pilot would be to allow teachers to develop only one SLO to test out in the fall and to use in the spring, or to develop just one SLO for the school year. Considering that teacher engagement with the SLO process, and not results from student growth on the SLO, is being factored into a teacher's evaluation next year, the district may want to consider this recommendation for the rest of the teacher population trying out SLOs for the first time next year.

Recommendation 3.2: Place a greater emphasis on how the SLO approach can be used to improve instruction and provide students with targeted feedback

In order for teachers to truly buy-in to the new SLO process as something that is more than just a compliance-based activity, they will need to see clear linkages made between this activity and what they do on a daily basis in the classroom. For example, a central feature of the SLO process should be score moderation activities, in which teachers meet together discuss their interpretation of student work relative to the course-specific learning trajectory. This is something that was not typically done for SGOs. In year 2 of the pilot project, efforts should be made to incorporate examples and discussions of teachers using the LP and the results of assessments targeted to the LP to adjust their instruction.

Finding 4: Teachers want to follow a standard process for developing and scoring SLOs that is rigorous and fair, yet easy to implement.

A selling point of SLOs as a growth indicator relative to student growth percentiles (for teachers who teach in tested subject areas) is that teachers and/or school leaders can have greater

control over the choice of the SLO and the associated student assessments. However, there is an inherent tension caused by the fact that there are financial bonuses attached to meeting an SLO under ProComp, and that there will eventually be negative consequences under annual LEAP evaluations if a teacher's students demonstrate little or no growth. The recognition that there will be stakes attached to SLO performance is one reason that many teachers expressed a desire for a process that would sacrifice a flexible SLO process tailored to the needs of individual teachers in favor of a more standardized process. For example, some teachers worried that if the choice of a course-specific trajectory were left at the discretion of each teacher, some teachers might be tempted to less challenging upper anchors, or to define many fine-grained trajectory levels to make it easier to demonstrate growth. Some teachers worried that if transitions across levels were allowed to take on different score values within and/or across schools, it could lead to outcomes that would be unfair. Finally, some teachers expressed the specific concern that the incentives attached to SLOs could cause other teachers to essentially game the system by placing students in the lowest possible LP level when setting baselines at the outset of the school year.

Recommendation 4.1: Establish some clear rules for the (a) number of acceptable levels in course-specific trajectories by grade and subject, (b) the values associated with student transitions across levels from the beginning to the end of a semester or school year.

In the long-run, if the SLO approach described in this report is to be scaled up, these rules will be necessary. We would recommend 5 levels and equal values of 1 point for each transition across levels as described in Chapter 2. However, over time it will be equally important for the district to establish a process for making changes and adaptations to these rules as teachers and schools gain more experience with the LP-based approach. For example, having teachers specify no more than 5 performance criteria in the current SLO template would serve as a good transitional step for moving into this direction. This rule is applicable to the current SLO process being rolled out by the district since similar to the levels within a course-specific trajectory, each of the criteria is being framed as a separate objective embedded in the SLO accompanied by its own set of assessment opportunities.

Recommendation 4.2: Put checks in place to ensure that teachers are accurately characterizing students according to the LP at the baseline and end of course.

SLO approvers should take care to look for signs that teachers are placing too many students in the lowest level of their course-specific trajectory. For example, if a realistic lower anchor has been defined to capture the lowest 1/3 of students a teacher typically sees from year to year, then if more than 1/3 of students are placed into this category it should raise a red flag. Each teacher will need to be able to make an empirically defensible case for how students were placed into baseline levels and then end-of-course levels to define growth along the LP. In the long-run, DPS should consider doing random annual audits of a small proportion of SLOs to ensure that they meet quality criteria.

Finding 5: Conventional Assessment Practices are Hard to Change

Much like the letter grades of A-F have acquired an ingrained sense of meaning in the American education system, the achievement levels of U (Unsatisfactory), PP (Partially Proficient), P (Proficient) and A (Advanced) have become internalized as a way to characterize student performance in DPS. Because a Learning Progression also defines levels, there was a tendency for teachers at our pilot sites to equate these levels to the U, PP, P and A achievement levels with which they were most familiar. At all math sites, we found that most teachers thought about growth and assessment in terms of proficiency as the only standard of interest for evaluating students and demonstrated performance that fell short of this standard was considered to be "below proficient". However, the LP approach requires thinking about growth in terms of how students develop an increased sophistication of knowledge and skills as they move along a developmental continuum, and thinking about assessment as a means for understanding student reasoning and locating students based on their responses on a continuum of knowledge and skills that lead toward mastery and higher levels. Table 1 outlines some of the shifts that we believe distinguish these two conceptions:

Table 1. Comparison of proficiency and learning conceptions of assessment

	Proficiency conception	Growth conception
Beliefs about the purpose of assessment	• To identify a student's proficiency level. • To provide feedback to students about their level of proficiency • To determine when re-teaching is necessary	• To determine where a student is located on a continuum of learning. • To provide feedback to students about what they have learned • To inform new teaching strategies.
Assessment practices		
When are assessments given?	Assessments are given after students are expected to have achieved proficiency on a particular objective	Assessment should be ongoing and integrated into learning activities
What is assessed?	Assessments are designed to assess learning objectives that are meant to reflect proficiency standards.	Assessments activities should be designed based on a developmental model and should cover multiple levels in the model (even those above and below where the student is hypothesized to "be")
How are results interpreted?	Assessment results should be interpreted in terms of "amount of proficiency achieved", using either the percentage of total points received or using a proficiency scale (i.e., partially proficient, proficient, advanced)	Observations should be interpreted to locate students on the developmental model
How is feedback given?	Students should be given feedback in terms of their proficiency level.	Students should be given qualitative feedback based on the developmental model
How do teachers use results?	Assessments results should be used to determine whether re-teaching or remediation is necessary	Assessment results should always guide instructional activities, based on the developmental model.

Recommendation 5.1: Continue to operationalize the assessment triangle using learning progressions.

Based on the lessons learned from this pilot, we believe this approach has many advantages. First, in the ideal scenario it can be based upon research-based best practices in both assessment and student learning. Second, it encourages teachers to set challenging goals by rewarding teachers for student growth rather than student status. Third, it encourages vertical

collaboration centered on discussions of student reasoning. It is our intention to deepen the implementation of this approach by focusing our efforts in year two on student reasoning, assessment quality and identifying instructional moves. For the Tier 3 elective areas, efforts will be made to ensure that score moderation activities are facilitated through the teacher leader workshops.

Recommendation 5.2: Recognize the shifts in thinking about growth and assessments that this work requires. Provide adequate time and resources to professional learning.

The shift in assessment conception described above requires PD time. In addition, helping teachers create and use LPs requires PD. For schools that use a research-based LP, teachers need training for how the LP itself, and the assessments based on the LP, can inform classroom instruction. For schools that begin with a generic LP and slowly flesh out the messy middle as they collaboratively examine student work, teachers need training on how to extract student reasoning from student work, and how to conceptualize this reasoning on a continuum of sophistication. We suggest that "adequate" resources for this work should include one or two full days of PD per year, along with reoccurring collaborative time (at least monthly), facilitated by a coach that is trained in this work.

Considering that teachers will vary considerably in their needs for support on various topics concerning the SLO process, different PDU workshops could also be structured and developed to address each topic in more depth. For example, PDU workshops could focus on refining tasks to meet quality standards or focus on rubric development since most teachers commonly use this scoring approach. These two topics were identified as areas of weaknesses with our participants, and based on input received from district staff members in ARE and the visual arts department, these two topics are also considered to be an area of common weakness found across many teachers at all schools. This would require coordinating efforts with the PDU office to strategically develop modules and supports for teachers wishing to participate in these PDU sessions.

Additional Reflections

The Value of Facilitating Teacher Collaborations

The interactions documented from both visual arts and math pilots illuminated the benefits of fostering collaborations across teachers for the SLO work, although the degree of success gained from these collaborations varied across sites.

For the math pilot, input collected from participants appears to generally point to positive responses received in collaborating through the score moderation process and from developing tasks together. At Blue, the number of grades represented in the math pilot proved challenging for vertical team collaborations since the professional development needs of lower elementary teachers differed from their peers located at the higher grade levels. In contrast to Blue, the smaller group of teachers at Spruce (4 teachers) and GB (9 teachers) and the narrower grade spans represented in these schools seemed to foster higher levels of facilitated collaborations since the professional development needs of all teachers in the group were relatively similar. As documented in Chapters 5 and 6, the teachers indicated their appreciation of collaborative work, including discussing student work with colleagues, working together to develop the LP and performance tasks, and engaging in the score moderation process (see pg. 79 and pg. 105).

For the two visual arts, although Spruce visual arts teachers had engaged in more collaborative work than the GB visual arts teachers, the small group dynamics and the high staff to teacher ratio made it difficult for them to acknowledge when they needed help and what areas were still unclear to them during each work session. As noted by the Spruce visual teachers, "we felt like we were being watched and observed by too many people and we didn't feel comfortable in that setting" (personal communication, November 12, 2013). In the case of the GB teachers, the collaborative dynamic was difficult to foster since one teacher was new to the profession and the other was a veteran teacher who also served as the mentor for the new teacher. Both teachers at GB interacted extensively with the facilitators and observer/participants (the visual arts program manager and the ARE data assessment coach) during the workshops, more so than with one another.

For the teacher leader arts pilot, the total number of teacher leaders collaborating was observed to work well by the visual arts program manager. This observation was confirmed by the highly positive input received from all 9 participants engaged with group activities together. The collaborative benefits noted by all participants make a strong case for pursuing a similar

model for the elective areas next year and to ensure that "singleton" teachers have structured and well facilitated opportunities to collaborate with one another.

Improving the Quality of Assessments Used with SLOs

In the visual arts, teachers were already accustomed to thinking about assessments as part of their daily routine since they continuously evaluate their performance of students in this performance-rich subject, and shared many examples about how they adjusted instruction based on reviewing student work. They were, however, accustomed to using formal pre- and post-tests for their SGOs, and believed that the district valued standardized assessments over classroom embedded assessments. Although all participating arts teachers welcomed the opportunity to integrate their performance/unit tasks into the body of evidence used to evaluate students on their SLO, ensuring that these tasks meet quality criteria is a critical step that will need to be taken during next year's pilot with elective teachers.

Similar to the visual arts teachers, math teachers at Spruce also indicated that the former SGO process had no "value for teacher-created artifacts" (see pg. 81) and appreciated using their own tasks to evaluate student growth on the SLOs. Although this perspective (i.e. valuing the use of teacher-created artifacts to support the SLO process and LPs) seemed to be shared across sites, the facilitators at Blue noted that more work needed to be done to help teachers understand how results from the assessments could be used to adapt classroom instruction (see pg. 61). For all sites, including elective areas, making stronger connections between assessments, student reasoning and instructional moves will be a focal area for the pilot work next year. Additionally, the math pilot sites, like the electives pilot next year, will also focus more on assessment quality to improve upon tasks being used or developed as part of the SLO body of evidence.

A majority of teachers in the district use teacher-developed assessments to support their SGOs (Briggs, et al., 2014). In the math pilot, engaging all teachers in the process of developing performance tasks and with score moderation provided them with experiences that facilitated understanding of measurement concepts applying to both externally developed and teacher developed assessments, such as "alignment", "reliability" and "error". That is, these concepts were discussed to check on how well: 1) their tasks were aligned with the LPs developed; 2) the extent to which inferences about student growth can be supported based on the quality of data

being generated from the tasks; and, 3) the performance standards set for each task also supported information about the location of students along the LP.

In contrast to the math pilot, the visual arts pilot could not engage in score moderation activities since teachers used completely different tasks embedded in their own units of study. Without engaging in score moderation, there were no collaborative checks conducted to see whether the resulting student work matched the performance expectations set in the scoring rubrics. Further, when the SLO quality check tool was used, one clear area consistently observed by us, the visual arts program manager and the data coach was that teachers in both visual arts pilots experienced difficulty in developing performance descriptors. Developing strong descriptors in rubrics are critical for increasing the reliability of scores achieved on each performance task and locating students on the LP. The weak scoring criteria meant that there is a strong likelihood that the performance of some students may have been misclassified. As indicated earlier (see pg. 160) the electives pilot will focus on assessment quality as a priority topic to address next year.

For the visual arts pilot, although operationalizing the assessment triangle may not have been as successful as the math pilot, the larger and more critical need of providing a foundation for integrating curriculum with instruction and assessment was largely met as observed by the visual arts program manager and the teachers who develop their own curriculum for instructing visual arts. As recorded in a conversation with one teacher leader based at a High School, "…this learning progression will be so valuable to us because so many arts teachers just teach on the fly and don't have a real sense of why they're even doing the activities that they're doing with their students. This is really tough on high school arts teachers when we get students who are coming to us from different high schools or middle schools with teachers emphasizing different things based on their personal preference. If [a student] joining my [arts] class had a teacher with a ceramics background, then they'll come to [me] with little to no understanding of value, contrast, and lines. But if we have these progressions spelled out, especially if we can get to the point of building a [multi-grade LP], then there's a better hope of getting teachers to agree on common skills and experiences that all students should have in all schools…that would be a great thing!" [personal communication recorded on July 25, 2014].

Next Steps

As we move into the next school year, the nature of the pilots will change to accommodate the expansion of both math and visual arts pilots, and to address some of the challenges highlighted in this report. We discuss preliminary next steps to be taken in each respective pilot and highlight areas where lessons learned are being applied to improve upon the delivery of Tier 3 to additional sites or elective areas:

Mathematics

Currently, discussions are underway to identify one to two new sites for implementing the Tier 3 math work and to continue on with the Tier 3 pilot at the year one sites. CADRE will coordinate with the district's curriculum director of mathematics to ensure that the Tier 3 pilot also supports the long-term goals of the math department around their standards implementation work.

Four major changes to the structure of the professional development (PD) sessions for new Tier 3 sites based on the input received from year one sites will be to: 1) cut down the total number of direct PD sessions from 40 hours to approximately 15-17 hours; 2) provide the big topic area to the site and provide a multi-grade LP used in a year one site with lower and upper anchors defined; 2) provide teachers with tasks used at year one Tier 3 sites to administer to students before the first session; and 3) immediately engage teachers in the process of score moderation during the second session of the school year.

Changes to the PD structure are being made to respond to concerns articulated by all math Tier 3 year one sites that there was too much planning time taken away from them during the PD sessions. However, now that tasks and multi-grade LPs are available from year one of the pilot, new pilot sites benefit from being able to immediately engage in the work of understanding the relationship of a learning progression relative to the SLO process, and to test out the set of performance tasks used to help identify baseline performance prior to session two.

For all sites including the year one Tier 3 math pilot sites, CADRE plans to place more emphasis during all PD sessions on student reasoning, score moderation, identifying instructional moves, and assessment quality.

Visual Arts

Based on the work accomplished with arts teacher leaders in year one, the Center for Assessment was asked to expand the visual arts pilot to the performing arts and to physical education. PD sessions will be offered to a group of teacher leaders in those five elective areas, and all year one Tier 3 visual arts pilot teachers will have the opportunity to join in on teacher leader work sessions to be scheduled throughout the school year. Currently, there will be approximately 10 visual arts, 15 P.E., and 14 performance arts teacher leaders who will join the elective Tier 3 pilot. During the next school year, the Center for Assessment will:

- Provide ongoing PD sessions to the curriculum coordinators in each elective area and provide them with guidance and content to train their own teacher leaders;
- Assist curriculum coordinators with developing course-specific trajectories;
- Assist in workshops being led by curriculum coordinators during the school year and help facilitate during each workshop;
- Align the Tier 3 electives work (with few modifications) to the district's SLO process and template; and,
- Develop or enhance existing tools that can be used by both teacher leaders and teachers to support the SLO process (e.g., improve upon the current format of the assessment alignment mapping tool used in year one).

These changes to the overall design of the Tier 3 visual arts pilot for the electives pilot were made to distribute the responsibility of developing course-specific trajectories as a basis for developing SLOs in other elective areas, and to also respond to the request made by the 3 elective curriculum coordinators to allow them to fully participate and help lead this work with their teachers.

One concern for this larger scale approach to implementing this Tier 3 electives pilot is that the coordination work across elective areas and teacher leaders charged with training their teachers will be considerable. Planning work is currently underway with curriculum coordinators to identify the extent to which activities such as score moderation can be effectively executed with teacher leaders during the 2014-2015 school year.

Appendix A

In this report we examine four quantitative measures of assessment quality: Alpha, standard error of measurement (SEM), item-total correlations, and p-values.

Table B.1 and C.1 show the values of Alpha and SEM for each of the tests used at Spruce and GB respectively. Alpha is a measure of the test's reliability—that is, whether the results of a particular test would generalize to a parallel test. Higher values of alpha indicate more confidence in the generalizability. The SEM is measured in test-score units, and can be used to construct confidence intervals for a given test score. To construct a 95% confidence interval, one would add and subtract 2*SEM to each score. One can also use SEM to get a feel for the number of categorization levels that are appropriate for a given test. One heuristic for this is to multiply the SEM by 4 to get the "width" (in terms of score-range) of each category. For example, in a 24-point test, if the SEM is 2 then each category ought to be 8-points. This means that the test can reliably support about 3 categories.

Item-total correlations are a measure of whether individual items measure the same construct as the test. In general we want to see item-total correlations above 0.4. Item-total correlations are reported in Figures B.1 and C.1 for Spruce and GB respectively.

P-values (represented in Figures B.2 and C.2 for Spruce and GB respectively) express the mean proportion of total points earned for a particular item (for example, a p-value of 0.1 means that, on average, students received 10% of the possible points for that item). Thus, they are a measure of item difficulty—low p-values indicate difficult items and high p-values indicate easy items. In general, we want to see a bell-shaped distribution. This would indicate that the majority of items are at the level of most students, while there are still some items that are above and below this level.

Appendix B. Assessment statistics at Spruce

The four quantitative measures of assessment quality presented in this report are further detailed here for Spruce. Summary and reliability statistics for the four final assessments are presented in Table 5.1, the distribution of item-total correlations is given in Figure B.1, and the distribution of the p-values for all assessments is provided in Figure B.2.

Table 5.1 Summary statistics for final assessments

Course	Teacher	Number of students	Number of items	Max possible score	Mean score	SD	Alpha	SEM
6th-Grade	Dave	75	3	8	6.4	2.0	.78	.93
7th-Grade	Cora	88	5	16	9.7	4.6	.86	1.72
8th-Grade	Mike	90	5	17	10.0	4.8	.78	2.25
Newcomer	Lisa	13	20	32	21.8	8.2	.87	2.96

Table B.1 indicates that each assessment had relatively high reliability (alpha > 0.8). Alpha is a measure of the test's reliability—that is, whether the results of a particular test would generalize to a parallel test. Higher values of alpha indicate more confidence in the generalizability. Tests with alpha > 0.8 are generally considered to be reliable. Additionally, Table B.1 indicates that the standard error of measurement (SEM) for each of the tests at Spruce. The SEM is used to create a 95% confidence interval around a test score. This is an indication of how "sure" we are about a student's score on the collection of tasks. If the SEM is large, we have less confidence about the scores assigned to students.

Figure B.1 Distribution of Item-Total Correlations for All Assessments

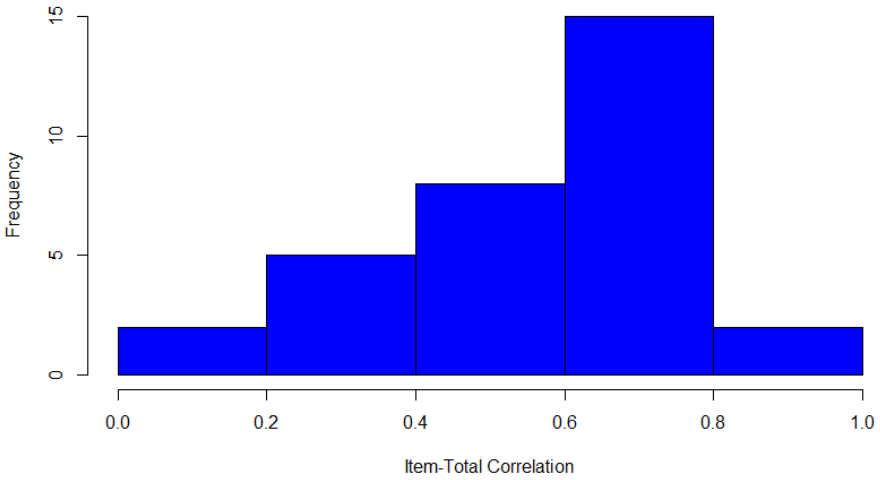


Figure B.2 Distribution of P-Values for All Assessments

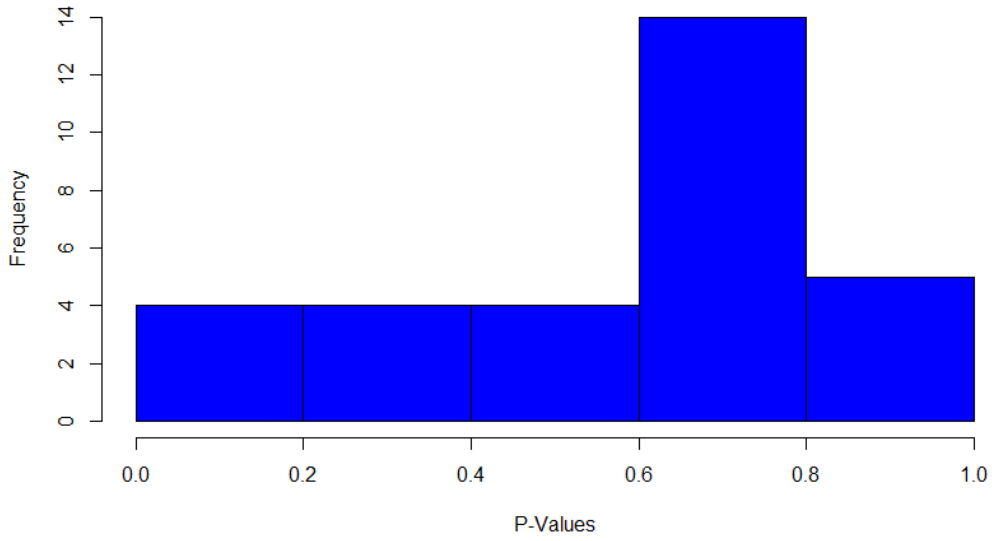


Figure B.1 shows a distribution of item-total correlation for all assessments (that is, all 32 items across the four tests). Item-total correlations are a measure of whether individual items measure the same construct as the test. They range from -1 to 1. In general we want to see item-

total correlations above 0.4. As shown, most items have an item-total correlation above between 0.4 and 0.8. Seven items have an item-total correlation below 0.4.

Figure B.2 shows the distribution of p-values for all items. P-values express the mean proportion of total points earned for a particular item (for example, a p-value of 0.1 means that, on average, students received 10% of the possible points for that item). Thus, they are a measure of item difficulty—low p-values indicate difficult items and high p-values indicate easy items. In general, we want to see a bell-shaped distribution. This would indicate that the majority of items are at the level of most students, while there are still some items that are above and below this level. As shown in Figure B.2, the distribution of p-values is skewed left, indicating that the tests contained too many easy items. However, the p-values span from 0 to 1, which shows that the test at least covered a full spectrum of difficulty levels.

Appendix C. Assessment statistics at GB

In this report we examine four quantitative measures of assessment quality: Alpha, standard error of measurement (SEM), item-total correlations, and p-values.

Table C.1 shows the values of Alpha and SEM for each of the five tests used at GB. As shown, three assessments had relatively high reliability (alpha > 0.8). The two assessments with assessments with lower reliability (alpha < 0.6) had only four questions each. These 4-question assessments do not meet the criteria that we outlined earlier (multiple questions at each expected student level, with some questions above and below these levels). The values of the SEM for each test suggest that the number of classification levels may have been too high. For example, consider Test B (categorization mapping shown in Figure 6.9 of the report, and statistics given in the second row of Table C.1). The SEM for Test A is 2.1, and the maximum score is 18. This means that the category width should be about 8 points, and that the test can reliably support 2-3 categories. As shown in Figure 6.9, at GB the test was used to create six categories, most of which had a width of 2-3 points each. Thus, even though Test B has a high reliability (alpha=0.83), it probably cannot support the classifications that we wanted to make. For this reason, the student-level results discussed in the report should be interpreted with caution.

Table C.1 Summary statistics for final assessments

Test ID	Course	Number of students	Number of items	Max possible score	Mean score	SD	Alpha	SEM
Test A	Algebra 1	30	11	20	10.9	4.9	0.80	2.2
Test B	Algebra 1	94	11	18	8.3	5.1	0.83	2.1
Test C	Algebra 2	35	17	24	6.9	5.7	0.87	2.0
Test D	Precalc	40	4	10	8.1	1.8	0.57	1.2
Test E	Precalc	66	4	9	4.8	2.5	0.52	1.7

Figure C.1 shows a distribution of item-total correlation for all assessments (that is, all 47 items across the five tests). Recall that we want to see item-total correlations above 0.4. As shown in Figure C.1, most items have an item-total correlation between 0.4 and 0.8. Four items have an item-total correlation below 0.4.

Figure C.1 Distribution of item-total correlations for all items

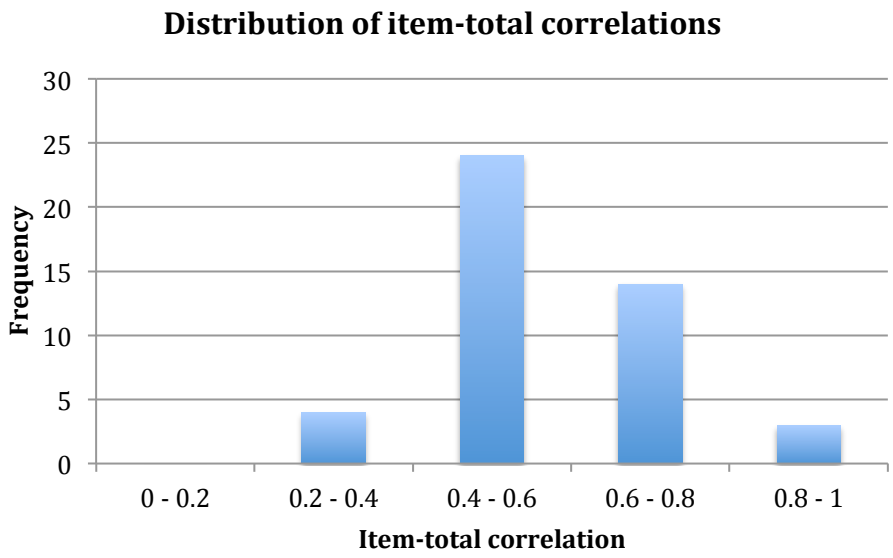
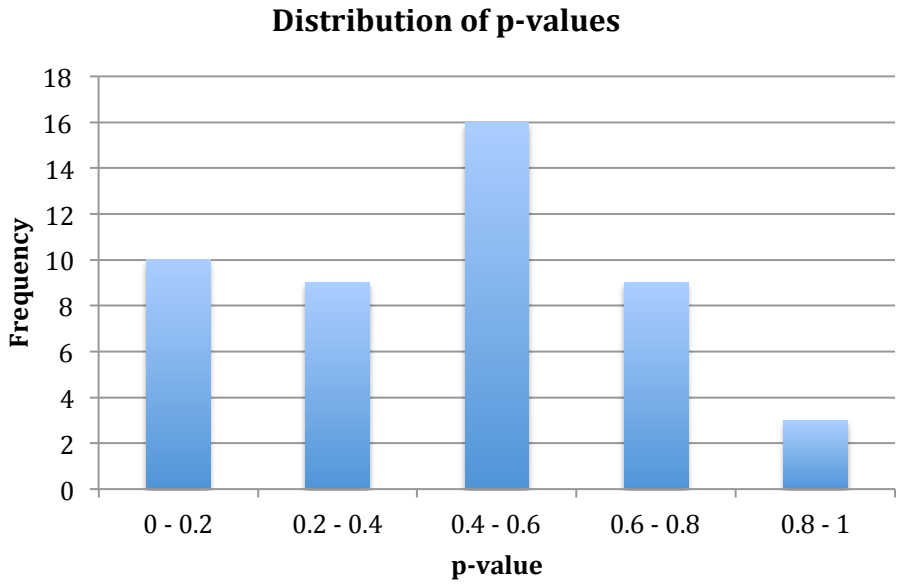


Figure C.2 shows a distribution of p-values for all 47 items. As shown, the distribution is approximately bell-shaped however, there were too many difficult items (items with p-values between 0-0.2).

Figure C.2 Distribution of p-values for all items



Appendix D. The across-grade LP at GB

4. (End of pre-calc): Students are able to distinguish between different types of manipulations for different types of equations/expressions, they flexibly choose strategies that are appropriate for the situation. This includes: (a) Students demonstrate that functions can be treated as objects by manipulating equations and expressions involving functions by treating functions as variables; (b) Students demonstrate understanding of inverse operations and inverse functions by using inverse operations & functions to manipulate equations (including logarithmic/exponential, and trig/inverse-trig); (c) Students use multiple strategies to write equivalent expressions, including factoring and expanding, rules of exponents and logarithms, substitution, and combining like terms.

- 3.3 Exponential and Log Equations
  - 3.3.3 Use *e* and natural log
  - 3.3.2 Solve log equations
  - 3.3.1 Solve exponential equations
- 3.2 Trig Equations
  - 3.2.2 Solve trig equations
  - 3.2.1 Verify trig equations using the identities
- 3.1 Quadratic equations
  - 3.1.1 Solve quadratic equations by completing the square

3. (end of algebra 2): Students show work when solving. Students solve quadratic equations over the set of complex numbers using multiple methods as appropriate (inspection, taking square roots, complete the square, factoring). They demonstrate understanding of positive, negative, and non-integer exponents by using them to solve exponential equations.

- 2.4 Rational functions
  - 2.4.2 Add/subtract multiply/divide
  - 2.4.1 Simplify by factoring
- 2.3 Manipulate quadratic
  - 2.3.1 Factor trinomial ("standard form") with leading coefficient $\neq 1$ into two binomials ("factored form")
- 2.2 Solve exponential and radical equations
  - 2.2.4 Solve by using like-bases
  - 2.2.3 Solve multi-step equations with rational exponents
  - 2.2.2 Undo rational exponent with their reciprocal
  - 2.2.1 Undo integer exponents

- 2.1 Simplify exponentials and radicals
  - 2.1.3 Simplify exponential expressions
  - 2.1.2 Rewrite radical as exponent
  - 2.1.1 Combine like terms

2. (end of geometry): Students rearrange formulas to highlight the quantity of interest, including using the four basic operations, squaring and squarerooting.

1 (end of algebra 1). Students understand that the solution to an algebra equation is the value for the variable(s) that makes the equation true by interpreting the solution in context or by checking through substitution. Students translate between various forms of linear equations, and they analyze and explain the process of solving linear equations in one variable (including those where the coefficients are represented by letters) by showing work. Students demonstrate that they understand the inverse relationship between multiplication & division and addition & subtraction by using inverse operations to solve linear equations in one variable with rational coefficients. In addition, students demonstrate that they understand that algebraic terms can be treated like objects by manipulating expressions (including combining like terms and applying the distributive property) and solving linear equations with algebraic terms on both sides of the equals sign. Students manipulate quadratic expression and solve quadratic equations using multiple methods as appropriate.

- 0.4 Manipulate and solve quadratic expressions and equations
  - 0.4.6 Solve equations in standard form with leading coefficient of 1 by factoring
  - 0.4.5 Convert from standard to factored with leading coefficient of 1
  - 0.4.4 Convert from factored to standard
  - 0.4.3 Solve in factored form
  - 0.4.2 Convert from vertex to standard form
  - 0.4.1 Solve given vertex form
- 0.3 Manipulate linear functions involving two (or more) variables
  - 0.3.2 Manipulate linear function with coefficients that are variables
  - 0.3.1 Convert between forms of linear equations (slope-intercept, point-slope, standard)
- 0.2 Solve equations where all terms are linear in one variable, showing work.
  - 0.2.4 Solve by simplifying on one or both sides separately before you solve
  - 0.2.3 Solve multi-step equations with rational coefficients
  - 0.2.2 Solve proportions, focus on fractions as division, multiplication undoes division
  - 0.2.1 Solve equations with variables on both sides

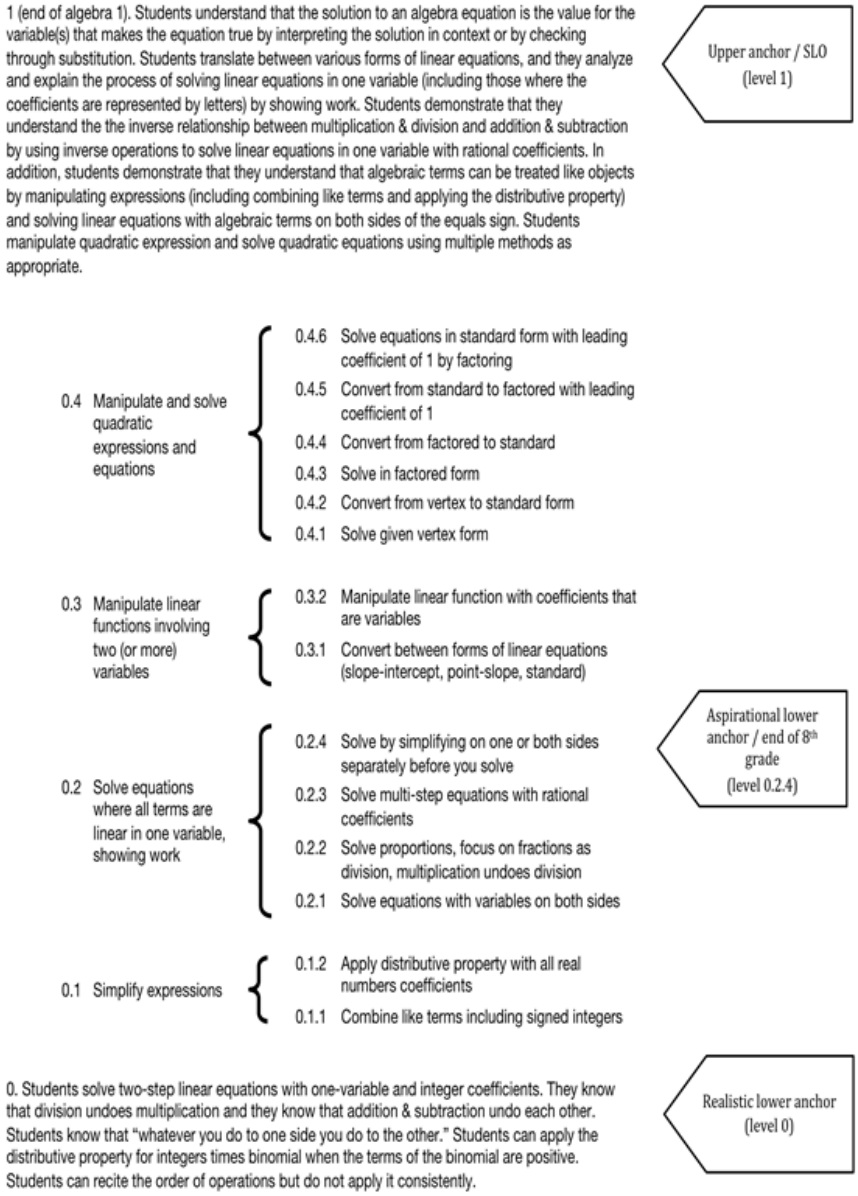
0.1 Simplify expressions
- 0.1.2 Apply distributive property with all real numbers coefficients
- 0.1.1 Combine like terms including signed integers

0. Students solve two-step linear equations with one-variable and integer coefficients. They know that division undoes multiplication and they know that addition & subtraction undo each other. Students know that "whatever you do to one side you do to the other." Students can apply the distributive property for integers times binomial when the terms of the binomial are positive. Students can recite the order of operations but do not apply it consistently.

(Note that throughout this progression, "solve" should be interpreted to mean "solve equation and check solution through substitution)

Appendix E. The course-specific trajectories at GB

Algebra I



3. (end of algebra 2): Students understand solving equations as a process of reasoning based on the assumption that the original equation has a solution, and they justify their solution steps. Students solve quadratic equations over the set of complex numbers using multiple methods as appropriate (inspection, taking square roots, complete the square, factoring). They understand positive, negative, and non-integer exponents, and they solve exponential equations using properties of exponents or by converting to logarithmic form and evaluating the logarithms.

Upper anchor / SLO (level 3)

- 2.4 Rational functions
  - 2.4.2 Add/subtract multiply/divide
  - 2.4.1 Simplify by factoring
- 2.3 Manipulate quadratic
  - 2.3.1 Factor trinomial ("standard form") with leading coefficient ≠ 1 into two binomials ("factored form")
- 2.2 Solve exponential and radical equations
  - 2.2.4 Solve by using like-bases
  - 2.2.3 Solve multi-step equations with rational exponents
  - 2.2.2 Undo rational exponent with their reciprocal
  - 2.2.1 Undo integer exponents
- 2.1 Simplify exponentials and radicals
  - 2.1.3 Simplify exponential expressions
  - 2.1.2 Rewrite radical as exponent
  - 2.1.1 Combine like terms

Aspirational lower anchor / end of geometry (level 2)

2. (end of geometry): Students rearrange formulas to highlight the quantity of interest, including using the four basic operations, squaring and squarerooting.

1 (end of algebra 1). Students understand that the solution to an algebra equation is the value for the variable(s) that makes the equation true. Students translate between various forms of linear equations, and they analyze and explain the process of solving linear equations in one variable (including those where the coefficients are represented by letters). Students use two main strategies for manipulating expressions and solving equations: balance and backtracking. They can explain and justify both strategies and they use both flexibly when solving and manipulating equations. Specifically, students understand that algebraic expression can be treated like processes, they understand the inverse relationship between multiplication & division and addition & subtraction, and they use these understandings to solve linear equations in one variable with rational coefficients (backtracking). In addition, students understand the algebraic terms can be treated like objects and they use this to manipulate expressions (including combining like terms and applying the distributive property) and solve linear equations with algebraic terms on both sides of the equals sign (balance). Students manipulate quadratic expression and solve quadratic equations using multiple methods as appropriate.

- 0.4 Manipulate and solve quadratic expressions and equations
  - 0.4.6 Solve equations in standard form with leading coefficient of 1 by factoring
  - 0.4.5 Convert from standard to factored with leading coefficient of 1
  - 0.4.4 Convert from factored to standard
  - 0.4.3 Solve in factored form
  - 0.4.2 Convert from vertex to standard form
  - 0.4.1 Solve given vertex form
- 0.3 Manipulate linear functions involving two (or more) variables
  - 0.3.2 Manipulate linear function with coefficients that are variables
  - 0.3.1 Convert between forms of linear equations (slope-intercept, point-slope, standard)
- 0.2 Solve equations where all terms are linear in one variable, justifying steps and explaining the meaning of the solution
  - 0.2.4 Solve by simplifying on one or both sides separately before you solve
  - 0.2.3 Solve multi-step equations with rational coefficients
  - 0.2.2 Solve proportions, focus on fractions as division, multiplication undoes division
  - 0.2.1 Solve equations with variables on both sides

Realistic lower anchor (level 0.2.4)

Algebra II

Pre-calculus

4. (End of pre-calc): Students are able to distinguish between different types of manipulations for different types of equations/expressions, they flexibly choose strategies that are appropriate for the situation. This includes: (a) Students demonstrate that functions can be treated as objects by manipulating equations and expressions involving functions by treating functions as variables; (b) Students demonstrate understanding of inverse operations and inverse functions by using inverse operations & functions to manipulate equations (including logarithmic/exponential, and trig/inverse-trig); (c) Students use multiple strategies to write equivalent expressions, including factoring and expanding, rules of exponents and logarithms, substitution, and combining like terms

Upper anchor / SLO (level 4)

3.3 Exponential and Log Equations	3.3.3 Use *e* and natural log 3.3.2 Solve log equations 3.3.1 Solve exponential equations
3.2 Trig Equations	3.2.2 Solve trig equations 3.2.1 Verify trig equations using the identities
3.1 Quadratic equations	3.1.1 Solve quadratic equations by completing the square

3. (end of algebra 2): Students show work when solving. Students solve quadratic equations over the set of complex numbers using multiple methods as appropriate (inspection, taking square roots, complete the square, factoring). They demonstrate understanding of positive, negative, and non-integer exponents by using them to solve exponential equations.

Aspirational lower anchor / end of Alg. II (level 3)

2.4 Rational functions	2.4.2 Add/subtract multiply/divide 2.4.1 Simplify by factoring
2.3 Manipulate quadratic	2.3.1 Factor trinomial ("standard form") with leading coefficient ≠ 1 into two binomials ("factored form")

Realistic lower anchor (level 2.3.1)

Appendix F. Multi-grade visual arts learning progression

Self-Expression Through the Creation of Independent Artwork

	Attributes	Colorado Academic Standards for Visual Arts		
Upper Anchor	c) Students can create a piece of artwork <u>independently</u> based on a clear self-developed concept AND be able to articulate their choices using appropriate terminology.	VA09-Gr.HS-S.2-GLE.2-EO.a Articulate and defend a personal philosophy of art using informed criteria (DOK 3-4) VA09-Gr.HS-S.3-GLE.2-EO.e Skillfully create and exhibit one's own works of art (DOK 1-4) VA09-Gr.HS-S.3-GLE.3-EO.b Interpret similarities and differences in artistic decision making (DOK 1-3) VA09-Gr.HS-S.4-GLE.2-EO.b Explain the personal influences shape the creation of functioning art (DOK 2-3)		
	b) Students are able to explain the how and why of their artwork in writing and verbally.	VA09-Gr.HS-S.2-GLE.3-EO.b Demonstrate fluency in using critique vocabulary to assess personal works of art and the others' works of art (DOK 3) VA09-Gr.HS-S.2-GLE.3-EO.d Interpret how meaning in works of art is related to the materials and processes chosen by the artist (DOK 1-3)		
	a) Students are able to visually communicate meaning and message through mastery of a variety of media.	VA09-Gr.HS-S.3-GLE.1-EO.a Articulate ideas that can be expressed in traditional, advanced, and evolving media (DOK 1-2)		
		Activities	**Skills**	
		artwork that tells a story, a concentration of concepts	• Mastery of Art Elements • Mastery of Principles of Design • Mastery of Media – combining media • Composition	**Drawing & Painting**
			• Mastery of All Techniques – use of kiln, loading/unloading • Mastery of Research – self-reflection and use of inspiration to make art	**Ceramics**

	Attributes	Colorado Academic Standards for Visual Arts		
Level 3	c) Students are able to reflect on the reasons we make art and reflect on the self to create artwork about social personal problems.	VA09-Gr.HS-S.1-GLE.3-EO.b Research and document how the visual arts are manifested in contemporary society (DOK 1-3) VA09-Gr.HS-S.2-GLE.1-EO.c Compare and contrast the technical proficiency of artists to communicate ideas, emotions, and information (DOK 2-3)		
	b) Students are aware of the environmental impact of art.	VA09-Gr.HS-S.3-GLE.3-EO.a Compare and contrast the analytical processes used to interpret works of art and images of mass media (DOK 2-3) VA09-Gr.HS-S.3-GLE.3-EO.c Discuss and debate how society has come to value some works of art over others (DOK 3-4)		
	a) Students understand visual arts through expressions beyond 2D designs using a variety of media.	VA09-Gr.HS-S.2-GLE.2-EO.b Document and apply investigations into a range of traditional and nontraditional studio practices to personal expression (DOK 1-4) VA09-Gr.HS-S.3-GLE.1-EO.b Investigate and document a wide range of traditional, advanced, and evolving media used in creating images that communicate ideas (DOK 1-3) VA09-Gr.HS-S.3-GLE.1-EO.c Create works of art representing traditional subject matter that use new media (DOK 3-4) VA09-Gr.HS-S.3-GLE.1-EO.d Create works of art representing personal narratives that use new media (DOK 1-4)		
		Activities	**Skills**	
		self-portrait (mixed media), wire sculpture, block print, plaster sculpture in the round, vans shoes national design contest (mixed media), wall murals (acrylic paint/marker), wood collage, public art (school parking lot trash cans) (airbrush, acrylic paint)	• Application of Art Elements in 3D Contexts • Application of Principles of Design in 3D Contexts • Application of Media – mixed media, wire, block print, plaster, acrylic paint, marker, wood, airbrush	**Drawing & Painting**
		research project, artist statement and critiques, personal narrative	• Application of Techniques – stains, mishima, underglazes, colored slip, new glazing techniques • Application of Critique Research – personal issues and how those issues influence art, inspiration, and flow	**Ceramics**

	Attributes	Colorado Academic Standards for Visual Arts	
Level 2	c) Students are able to create artwork that tells a story.	VA09-Gr.HS-S.2-GLE.1-EO.a Explain the process of critique using the progression of description, analysis, interpretation, and evaluation (DOK 1-2) VA09-Gr.HS-S.2-GLE.1-EO.b Communicate and defend reasons for opinions about the intentions (successful or not) of a work of art (DOK 1-4) VA09-Gr.HS-S.3-GLE.2-EO.d Create works of art that speak to personal artistic opinion in response to cultural contexts (DOK 3-4)	
	b) Students examine how history and culture relate to art creations.	VA09-Gr.HS-S.1-GLE.1-EO.c Connect and compare visual art characteristics and expressive features of art and design in cultural contexts (DOK 1-3) VA09-Gr.HS-S.1-GLE.2-EO.b Discern the complexity of art and historical paradigms in cultural context (DOK 3) VA09-Gr.HS-S.1-GLE.2-EO.c Debate divergent art histories and traditions (DOK 3-4) VA09-Gr.HS-S.1-GLE.2-EO.e Analyze innovation in art through historical and cultural context (DOK 2-4) VA09-Gr.HS-S.1-GLE.2-EO.f Analyze the reasons to avoid replication, interpretation, or illustration of images or icons that are culturally sensitive such as kachina dolls, and Navajo sand painting (DOK 2-3) VA09-Gr.HS-S.1-GLE.3-EO.a Investigate and articulate the aims of disparate art practices and traditions (DOK 1-3) VA09-Gr.HS-S.1-GLE.3-EO.c Interpret the purposes of art across time, culture, and diversity, and be mindful of cultural sensitivities (DOK 4) VA09-Gr.HS-S.4-GLE.1-EO.b Discern the value of works of art based on historical significance, craftsmanship, cultural context, and originality using appropriate, domain-specific terminology that honors sensitive traditions (DOK 3-4) VA09-Gr.HS-S.4-GLE.3-EO.a Compare and contrast the roles of artists and designers in historical and contemporary context (DOK 2-3)	
	a) Students discover which technique would work best to create a successful work through guided self-choice.	VA09-Gr.HS-S.1-GLE.1-EO.b Investigate and articulate the value of the characteristics and expressive features of art and design in diverse and disparate works of art (DOK 1-3) VA09-Gr.HS-S.3-GLE.2-EO.b Discern and articulate the quality of personal works of art using a variety of reflective processes (DOK 1-4)	
		Activities	**Skills**
		self-portrait (pencil), primitive drawing (marker), still life (oil pastel), transformation (color pencil), watercolor painting, photo reproduction (pencil), perspective drawing (color pencil), dream state (mixed idea), abstraction (coffee, ink), interpretive self-portrait (color pencil), pointillism (marker), commission work	• Application of Art Elements • Application of Principles of Design • Application of Media – pencil, marker, oil pastel, color pencil, watercolor, mixed, coffee, ink • Art History – primitive, pre-Renaissance, Impressionist, Surrealist, Contemporary • Cultures – Native American, European, Mexican, Asian, American **Drawing & Painting**
		metamorphosis, social issues figure series	• Application of Techniques – hand-building, carving, throwing • Critique – problems and solutions students have learned on their own **Ceramics**

	Attributes	Colorado Academic Standards for Visual Arts		
Level 1	c) Students are able to engage in in-depth investigation and problem solving of visual and design issues.	VA09-Gr.HS-S.1-GLE.1-EO.a Demonstrate skills that utilize the characteristics and expressive features of art and design to communicate meaning (DOK 1-3)		
	b) Students focus on understanding the fundamentals of visual arts through the use of art elements and principles of design.	VA09-Gr.HS-S.1-GLE.2-EO.a Analyze visual arts traditions (DOK 2-4) VA09-Gr.HS-S.4-GLE.2-EO.a Use sketches, plans, and models to create and/or design a functioning work of art (DOK 3-4)		
	a) Students use and experiment with tools while focusing on a specific technique.	VA09-Gr.HS-S.3-GLE.2-EO.a Skillfully use a variety of techniques and media to create works of art (DOK 1-4) VA09-Gr.HS-S.3-GLE.2-EO.c Demonstrate collaboration to create works of art (DOK 1-2)		
		Activities	**Skills**	
		self-portrait (pencil), ink drawing (pen and ink), 3-D lettering (marker), outside space perspective (pencil), photo reproduction (pencil), landscapes (watercolor), product advertisement (color pencil), object drawing (charcoal), cartooning (color pencil), warm/cool painting (acrylic paint), distortion grid (pencil), formal evaluation gallery walk	• Art Elements – line, shape, value, color, texture, space, form • Principles of Design – movement, pattern, balance, rhythm, unity, balance, contrast • Techniques – line techniques and variety (cross hatch, stipple, value scale) and perspectives • Media – pencil, pen and ink, marker, watercolor, color pencil, acrylic paint	**Drawing & Painting**
		one project per technique (introduce/demo, have students experiment, students identify which technique is successful), slab box with carving tool, biomimicry, artist/image slideshow	• Techniques – coil, slab, pinch, throwing • Art History – contemporary art	**Ceramics**

	Attributes	Colorado Academic Standards for Visual Arts	
Lower Anchor	c) Students have displayed artwork and only have a classroom experience of art.	VA09-Gr.8-S.3-GLE.2-EO.b Create works of art that are display-ready (DOK 1-3) VA09-Gr.8-S.4-GLE.2-EO.b Create and display works of art created to enhance or commemorate an event (DOK 3-4)	
	b) Students have an understanding of basic skills (line, shape, etc.) and experience working with a variety of materials without much depth.	VA09-Gr.8-S.1-GLE.3-EO.c Analyze basic concepts in art such as balance, symmetry, integrity, authenticity, originality, "form follows function," and "thinking outside-of-the-box" (DOK 2-3)	
	a) Students have a desire to describe what they see or what they want to see.	VA09-Gr.8-S.3-GLE.1-EO.a Create innovative works of art (DOK 3-4)	
		Activities	**Skills**
		N/A	Drawing & Painting
		N/A	Ceramics

Appendix G. Course-specific visual arts learning progressions

Figure G.1. Course-Specific Progression for the Characteristics and Expressive Features of Art in Beginning Drawing & Painting

Level	Attributes	Activities	Skills
4 **upper anchor/ SLO**	Students will be able to engage in an in-depth investigation of the characteristic and expressive features of art.	• Charcoal Drawing of an Object • Warm and Cool Acrylic Painting	– advanced applications of line , value, texture, form, and space – advanced applications of color, shape, and space
3 **"messy middle"**	Students will focus on understanding fundamentals of art through the art elements and principles of design.	• Product Advertisement • Cartooning • Distortion Grid	– pattern, color, contrast, message – color, space, movement – line, space
2 **"messy middle"**	Students will be able to experiment with a variety of tools/media while focusing on a specific technique.	• Watercolor Landscape • 3D Lettering • Perspective of Outside Space • Pen and Ink • Self-Portrait in Pencil	– line, value, texture, form, space – line, value, texture, form – space, perspective, texture, line – pattern, contrast, color – shape, space, proportion
1 **aspirational lower anchor**	Students will learn content vocabulary for each experience.	• Vocabulary List/Test for each assignment • Exit Slip (definition of terms) • Cornell Notes for assignments	– understanding of key vocabulary used in visual arts
0.5 **almost ready for course**	Students have displayed art work and have had minimal classroom experience in art.		
0 **not ready for course**	No exposure to the arts.		

Figure G.2. Course-Specific Progression for Applications of 3D Art in Advanced Drawing & Painting

Level	Attributes	Activities	Skills
3 **upper anchor/ SLO**	Students will develop awareness of the environmental impact of art as it relates to the application of 3D and can use this awareness to create personal intention in a piece.	• Wall Murals and Public Art • Wood Collage	– permanent display in building color (acrylic painting) – shape, color, defined space
2 **"messy middle"**	Students are able to reflect on the reasons we make art and reflect on the self to create artwork about social personal problems.	• Artist Journal • Self-Grading on Assignments • Content Vocabulary Tests • Self-Portrait	– written, visual, and verbal communication
1 **aspirational lower anchor**	Students will understand visual arts through expressions beyond 2D designs using a variety of media.	• Plaster Relief Sculpture • Vans Contest (Shoe Design) • Wire Sculpture • Block Printing	– 3D fore-, mid-, and background – 3D image based on theme – 3D shape and form with line – imagery with color, +/– space
0.5 **almost ready for course**	Students understand principles of design and art element but are weak on application.		

Figure G.3. Course-Specific Progression for Characteristics and Expressive Features in Beginning Ceramics

Level	Attributes	Activities	Skills
3 **upper anchor/ SLO**	Students will be able to engage in an in-depth investigation of the characteristic and expressive features of 3D art.	• Metamorphosis Figure/Sculpture • Written Reflection of Metamorphosis Sculpture	– demonstrating intent with art elements – developing display-ready pieces – communication of "why" and "how" pieces were created (providing a rationale)
2 **"messy middle"**	Students will focus on understanding fundamentals of art through the art elements and principles of design.	• Creature Head • Tea Set • Slab Box	– application of 3 different ceramic handbuilding techniques – identifying and applying principles and elements of art – wax resists, stains
1 **lower anchor**	Students will be able to experiment with a variety of tools/media while focusing on a specific technique.	• Two-Sided Vase • Pinch Pot • Coil Pot • Experimenting with Glazing and Surface Decorating Techniques	– wedge, pinch, slipping, blending, scoring – assessing form – glazing, color theory, pattern – texture, application of wire cutters, culture – photo transfer, mishima, underglazes, sgraffito – ceramic vocabulary
0.5 **almost ready for course**	Students have displayed art work and only have a minimal classroom experience in art.		N/A
0 **not ready for course**	No or minimal exposure to the arts.		N/A

Figure G.4. Course-Specific Progression for Artwork that Tells a Story in Intermediate Ceramics

Level	Attributes	Activities	Skills
3 upper anchor/ SLO	Students will be able to create artwork that tells a story while applying appropriate techniques and providing rationale for these techniques in their original artwork.	• Final Research Paper and Sculpture • Written/Oral Critique	– confident and proficient in building and glazing techniques – sketching detailed drawings with notes; piece visually represents the sketch
2 "messy middle"	Students will be able to self-reflect and decide which techniques and glazing they feel are most appropriate for the assignment.	• Research Paper Proposal and Sketches • Written/Oral Critiques • Collaborative Project • Personal Narrative • Biomimicry • Metamorphosis	– application of building and glazing techniques – sketching detailed drawings with notes; piece visually represents the sketch
1 lower anchor	Students will be able to experiment with various building techniques and glazing techniques to be able to make their own decisions, which would work best for them personally.	• Simple Models (focusing on practice of ceramics techniques) • Simple Models (focusing on practice of ceramics tools)	– sketching, note-taking – glazing, building – use of specific tools
0.5 almost ready for course	Students will be able to use various tools and specific techniques to demonstrate problem solving of visual and design issues.	• Discussion of Tools • Discussion of Design in Ceramics • Discussion of Building Technique	– design – hand-building, application of building techniques – experiment and use of tools, learning tool names
0 not ready for course	No exposure to the arts.	N/A	

Appendix H. SLO assessment quality check tool[7]

Part 1: Clarity and Focus
The assessment : check all that apply (Note: include as much information as possible to provide a clear picture of the assessment) • Addresses an essential question/issue, big idea, or key concept or skill of the unit/course • Clearly indicates what the student is being asked to do **(Student directions & assessment task/prompt** – what does the student see/use? • Assesses what is intended to be assessed – will elicit what the student knows and can do related to the chosen learning objective, and accompanying standards and benchmarks • Includes what will be assessed individually by the student (even if it is a group task) • Is linked to ongoing instruction (e.g., evaluates students as they progress through a unit of study or course) • **Other:**
Based on the content evaluated by the task or the set of items reviewed, explain what purpose the assessment serves (e.g. diagnostic, report card grades, interim, etc.).
Part 2: Content
• Is clearly aligned to my learning objective • Is clearly aligned to specific Content Standards (or intended GLEs in the content standards)
Identify DOK levels assessed. For example, an essay would mostly assess DOK 3 (via weighting in the rubric), but some DOK 2 items might also be included. You would check "most" for DOK 3 and "some" for DOK 2 below: DOK 1: recall; memorization; define/locate a term/principle; perform routine operation (e.g., graph points, evaluate an expression, measure, follow rule to round a number or check spelling); complete a sentence or label a diagram (____most of test/ ___some of the test/ ___none of the test) DOK 2: summarize; sequence events; identify main idea; make basic inference; classify/organize data; show/explain relationship; determine fact-fiction. There is a correct answer involving multiple concepts or multiple decision points. (____most of test/ ___some of the test/ ___none of the test)

[7] © 2013 Modifications by Diaz-Bilello, E., SLO Assessment Quality Check Tool
© 2012 Modifications by J. Theompson; High Quality Assessment Review Tool
© 2009 Hess, Karin K., Local Assessment Toolkit: High Quality Assessment.
Permission to reproduce is given when authorship is fully cited.

DOK 3: support thinking or reasoning by citing references from text or other sources/data; go beyond the text/data to analyze, generalize or connect ideas. Requires deeper understanding beyond what is presented. Items may require testing conjectures/conducting investigation, abstract reasoning, interpretations, and application/transfer of prior knowledge or text support for an analytical judgment about a text. More than one possible correct answer/solution.
(____most of test/ ___some of the test/ ___none of the test)

DOK 4: Requires use of multiple sources/texts/data sets/concepts; complex reasoning, planning, and developing of concepts. Usually applies to an extended task or project. Examples: evaluate works by the same author, critique an issue across time periods or research topic/issue from different perspectives; design & conduct longer investigations or research projects applying science and math concepts or developing alternative solutions.
(____most of test/ ___some of the test/ ___none of the test)

- Is aligned with intended rigor of the learning objective.

The assessment provides for ownership and decision-making, requiring the student to be actively engaged; uncovers thinking.

- Is authentic. Reflects a real-world situation or authentic application.
- Other:

Part 3: Scoring

Scoring Guide to be used with the assessment:

- Generalized Rubric (e.g., for persuasive writing, for all science labs, for all art tasks, etc.)
- Task-specific Rubric (only used for the particular task)
- Scoring Guidelines (e.g., checklist with score points for each part)
- Answer key, scoring template, computerized or machine scored
- Teacher Observation Sheet/Observation Checklist

The rubric/scoring guide used with the assessment:

- Assesses all intended parts of the learning target related to the SLO.
- Has clear performance descriptors that differentiate students in distinct performance levels. The scoring guide should be useful in determining what the student knows AND does not know, not simply yield a score. (e.g., what does a score of 25 really mean? What additional or next steps in instruction does the student need?)
- Has exemplars (from student work) or anchor papers available for the task that illustrates expectations aligned to the learning objective.

Part 4: Fairness
(the areas below should be discussed relative to the needs of ELLs, gifted and talented students, and students with disabilities)

Is fair and unbiased in language and design:

- Material is familiar to students from identifiable cultural, gender, linguistic, and other groups
- The task (context/texts used) is free of stereotypes
- All students have access to resources (e.g. Internet, calculators, spellcheck, etc.)
- Assessment conditions are the same for all students or flexible enough not to change what's being assessed (e.g., reading a passage aloud may be fine for interpreting, but not for decoding words)
- The task can be reasonably completed under the specified conditions
- The rubric or scoring guide is clear for different response modes (oral, written, etc.)

Adheres to the principles of Universal Design for Learning (UDL):

- Instructions are free of wordiness or irrelevant information
- Instructions are free of unusual words (unusual spellings or uses) that the student may not understand
- Instructions are free of ambiguous words
- There are no proper names that students may not understand (e.g., because they have never seen them before in instruction)
- Questions/prompts are marked with graphic or visual cues (bullets, numbers, in a text box, etc.)The task can be reasonably completed under the specified conditions
- The test format is consistent
- Formatting and layout is visually clear and uncluttered

Accommodations are commonly categorized in five ways: presentation, response, setting, timing and scheduling, and linguistics. Consider these (if needed) to ensure that accommodations are provided to ensure that English Learners and/or Students with Disabilities can fully access the content represented by the task or set of items reviewed.

Presentation Accommodations *– Allow students to access information in ways that do not require them to visually read standard print. These alternate modes of access are auditory, multi-sensory, tactile, and visual.*	
Response Accommodations*—Allow students to complete activities, assignments, and assessments in different ways or to solve or organize problems using some type of assistive device or organizer.*	
Setting Accommodations*—Change the location in which a test or assignment is given or the conditions of the assessment setting.*	

	Timing and Scheduling Accommodations—*Increase the allowable length of time to complete an assessment or assignment and perhaps change the way the time is organized.*	
	Linguistic Accommodations—*Allow English language learners (ELLs) to access academic content measured by reducing the linguistic load of an assessment. The accommodation is based on an ELL's limited English language proficiency, which is different than an accommodation based on a student's disability or a cognitive need.*	

*Please reference "Defining Features of Academic Language in WIDA's Standards"

Appendix I. Visual arts tasks and rubrics with feedback from the original pilot

Object Reproduction Task and Rubric for Drawing & Painting Course

Grading Rubric
Level 1 Object Reproduction

100-80: A complete object reproduction. The drawing demonstrates **consistently accurate** shading and value. The drawing is an exact reproduction of the image of that covers the surface. The image demonstrates consistently accurate use of light source. The drawing is **creative in composition and content**. Student uses an object for inspiration with changes if needed. Work is completed and turned in on due date.

79-70: A complete object reproduction. The drawing demonstrates **accurate** shading and value. The drawing is a close reproduction of the image of that covers the surface. The image demonstrates accurate use of light source. The drawing is **creative in composition and content**. Student uses an object for inspiration with changes if needed. Work is completed and turned in on due date.

69-60: An object reproduction. The drawing has an image that is shaded and covers no less then 75% of the surface. The image demonstrates **basic** use of light source. The drawing may be a **basic** reproduction of an object. Work is turned in on due date.

59-0: An incomplete drawing that uses **less than** 75% of the surface. The drawing has an image that **is not shaded appropriately**. The image demonstrates **inaccurate or no** use of light source. **No** attempt is made to do the assignment.

Standards Addressed:
1. Observe and Learn to Comprehend
Use the visual arts to express, communicate, and make meaning.
2. Envision and Critique to Reflect
Articulate and implement critical thinking in the visual arts by synthesizing, evaluating, and analyzing visual information.
3. Invent and Discover to Create
Generate works of arts that employ unique ideas, feelings, and values using different media, technologies, styles, and forms of expression.
4. Relate and Connect to Transfer:
Recognize, articulate, and validate the value of the visual arts to lifelong learning and the human experience.

_____, _____; Use of charcoal to create detailed realistic shading
_____, _____; Creative composition, and content
_____, _____; Consistent use of light source
_____, _____; Likeness of object
_____, _____; Choice of image
_____, _____; Effort and assessment

EDB 4/22/14 11:18 AM
Comment [1]: Unless it's mandated that you use a point scale from 0-100, I'd recommend that you use performance level descriptors rather than a scale. For example this highest level of 80-100 could be classified as "mastery" or something to that effect. Otherwise, it's tough to explain the difference between an 81 from an 82 or a 90. If you want to add bonus teacher criteria, then you can simply bump up a student from one performance level to the next and make a note as to why you did this (i.e., here, I'm thinking about that one special needs students who put in a lot of effort to work on the most difficulty replication approach based on fabric. Now if you absolutely MUST translate these descriptors to a 100 point scale, then let's talk about this.

EDB 4/24/14 4:20 PM
Comment [2]: Is it important to note that 75% of the image has been shaded as a key criteria? What if it were only 70% of the image – why is it important to emphasize this?

EDB 4/28/14 3:12 PM
Comment [3]: Is this a checklist? To me, these elements can be used to form a strong basis for a good rubric to evaluate student work on this unit. See the second document I've attached.

Proposed Revised Rubric for Object Drawing from PD Facilitator

Take a look at the below sample rubric. In this example, I've taken the core elements of the aspects that are critical to your object reproduction part of the assignment and captured it below. This is an example of how a few rubrics can be developed to capture different aspects of what you're assessing. The difference between the two is that the first one reflects a rubric where the point values are lower because the criteria are not as important as the criteria being evaluated under the second rubric. You'll also notice that I have self-assessment evaluated without reference to effort, which is part of your original criteria. The reason for this is that effort is being assessed in all other criteria.

Criteria	Below Basic (1 point)	Basic (2 points)	Proficient (3 points)	Advanced (4 points)
Image of object selected	Does not meet the deadline for selecting one image for reproduction	Selects the most simple image to reproduce (metal)	Selects an intermediate image to reproduce (glass)	Selects a difficult image to reproduce (fabric)
Self-assessment	Does not submit the self-assessment	The self-assessment is submitted with many supportive details missing to justify rating given	The self-assessment is submitted with rationale provided to justify rating given	The self-assessment is submitted with provided to justify rating given

Criteria	Below Basic (2 points)	Basic (4 points)	Proficient (6 points)	Advanced (8 points)
Realistic representation	Picture does not bear any resemblance to the object or fails to attempt the reproduction work	Picture bears a slight resemblance to the object	Picture resembles the object, but with some nuanced details missing from the reproduction and some inaccuracies with proportion	Picture is an exact replica of the object with nuanced details of the image included and with accurate proportions represented
Application of value	Does not apply any of the shading techniques learned	Uses basic techniques learned but does not apply the techniques accurately to define the image	Uses basic and some intermediate techniques learned with some accuracy reflected to define the image	Uses a combination of basic, intermediate and more advanced techniques to accurately define the image
Effective representation of light source on image	Does not capture any effects of lighting on the object	Captures effects of lighting on the object, but with great inaccuracies reflected	Captures the main effects of lighting on the object, but with some inaccuracies reflected	Uniformly captures the main and subtle effects of lighting on the object with little to no inaccuracies reflected
Use of entire surface	No background shading is applied to make use of the entire surface of the paper	Some background shading is applied, but considerable white space is still seen	Background shading is applied with almost full use of the entire surface of the paper	Background shading is applied to the entire surface of the paper to complement the placement of the object and to emphasize the object's placement in space

Ceramics 2/Advanced
Spring Research paper proposal

Name_____
Period_____

Artist 1-Ceramic Artist:

Name_____
Medium_____
Years active_____
Where is this artist based? (California, Paris, Denver?)_____
What surface decorating technique do they use? (mishima, underglazes, stains, etc)
Who/what influences the artist?

What is the public perception of this artist? Is the work liked or hated? Why? (You will need to look up a critique from a critic for this. Jerry Saltz is a NY Times art critic)

Why are you choosing this artist? What about their artwork inspires you?

Insert an image of the artwork from Artist 1 that inspires you most here. Include the title, medium, and date

How are you going to use this inspiration in your piece?

Artist 2:

Name_____
Medium_____
Years active_____
Where is this artist based? (California, Paris, Denver?)_____
Who/what influences the artist?

What is the public perception of this artist? Is the work liked or hated? Why? (You will need to look up a critique from a critic for this. Jerry Saltz is a NY Times art critic)

Why are you choosing this artist? What about their artwork inspires you? BE SPECIFIC

Insert an image of the artwork from Artist 2 that inspires you most here. Include the title, medium, and date

How are you going to use this inspiration in your piece? BE SPECIFIC

Rajendra Chattergoon 5/5/14 9:24 AM
Comment [1]: It is unclear why this assignment is assessed using only completion criteria (writing in complete sentences). In our conversation, you mentioned that this paper proposal is used for planning the sculpture and the in-class critique. These skills should be part of your assessment strategy for the assignment.

Rajendra Chattergoon 5/5/14 9:07 AM
Comment [2]: We also talked about making sure that the assignment is accessible to all students. Although all students have access to the school library and the schools computers, can students get help with technical aspects of word processing such as inserting images? Are students able to print an image and paste it onto a paper copy?

Rajendra Chattergoon 5/5/14 10:11 AM
Comment [3]: What does it mean to be "specific?" What criteria are you looking for in their responses other than complete sentences? Is there a distinction between your expectations for Artist 1 and Artists 2 and 3. Artist 1 does not have the "be specific" criteria. I've attached a sample rubric that you might want to use as you think about the criteria that you would like to assess from this project.

Artist 3:

Name_____
Medium_____
Years active_____
Where is this artist based? (California, Paris, Denver?)_____
Who/what influences the artist?

What is the public perception of this artist? Is the work liked or hated? Why? (You will need to look up a critique from a critic for this. Jerry Saltz is a NY Times art critic)

Why are you choosing this artist? What about their artwork inspires you? BE SPECIFIC

Insert an image of the artwork from Artist 3 that inspires you most here. Include the title, medium, and date

How are you going to use this inspiration in your piece? BE SPECIFIC

Sketch your sculpture(s), include measurements for height, width, etc. BE AS SPECIFIC AS POSSIBLE. Using colored pencil color your sketch how you would like to glaze it.

Describe your work and how you are using each of the 3 artists as inspiration in your artwork. **How** will you create this? **What** is your method? **How** are you going to glaze this? **What** do you need from the studio to create this? *BE VERY CLEAR AND SPECIFIC.*

Are the first set of categories and criteria as important as the second set? If not, you might want to consider using points instead of letters to assign scores to your students. For example, for the first set of categories you could use the following scheme: advanced (5), proficient (4), secure (3), developing (2), unsatisfactory (1). For the second set, you could use: advanced (10), proficient (8), secure (6), developing (4), unsatisfactory (2).

Rubric for Research Project-Sculpture

Categories and Criteria	Advanced A	Proficient B	Secure C	Developing D	Unsatisfactory F
Tools and materials were used properly and efficiently. (All media and tools)	All media (clay, paint, canvas, found objects, etc.) was applied properly and used in appropriate amounts. Media was chosen with planning and intent. The use of the materials contributed to the overall aesthetics of the final piece.	All media (clay, paint, canvas, found objects, etc.) was applied properly and used in appropriate amounts. Media was chosen with *some* planning and intent. The use of the materials contributed to the overall aesthetics of the final piece.	Most media (clay, paint, canvas, found objects, etc.) was applied properly and used in appropriate amounts. Media was chosen with *some* planning and intent. The use of the materials contributed to the overall aesthetics of the final piece.	Some media (clay, paint, canvas, found objects, etc.) was applied properly and used in appropriate amounts. Media was chosen with *little* planning and intent. The use of the materials only somewhat contributes to the overall aesthetics.	None of the media (clay, paint, canvas, found objects, etc.) was applied properly or used in appropriate amounts. Media was chosen with *little or no* planning or intent. The use of the materials does not contribute to the overall aesthetics.
Maintenance of Studio Space	Workspace was always left free of materials and mess. Tools were always put away properly and cleaned properly.	Workspace was almost always left free of materials and mess. Tools were almost always put away and properly cleaned.	Workspace was free of materials and mess only about half the time. Tools were sometimes not put away or cleaned.	Workspace was often left messy. Tools were often not put away or left unclean.	Workspace was always left messy. Tools were never put away or left unclean.

Criteria and Mastery Levels	Advanced A	Proficient B	Secure C	Developing D	Unsatisfactory F
Sculpture(s) is original, and draws inspiration from 3 artists (with one of the artists being a ceramics artist)	Artist drew inspiration from 3 different artists, one being a ceramics artist. Appropriate detail from each artist is captured and it is evident that the artist was influenced by these artists.	Artist drew inspiration from 3 artists, but none of them work in ceramics. Most appropriate detail is captured from inspiration from the 3 artists..	Artist drew inspiration from 2 artists. Some evidence of inspiration from these artists are captured.	Artist used 1 artist for inspiration. Very little detail is included and the inspiration is a little unclear.	Artist did not use inspiration from any other artists.

Take a look at the criteria that you've listed and double check that they are all equally weighted. For example, is craftsmanship as important as the size or glazing of the sculpture? Can some of your criteria be combined? For example, if glazing and size are less important than craftsmanship perhaps you might want to think about combining these criteria.

Proposed Revised Rubric for Research Proposal from PD Facilitator

Please take a look at the following sample rubric for your research paper proposal. I've connected your task more explicitly to the learning targets and skills that you're assessing. You placed this task at the upper anchor of your micro-LP for your Level 2 ceramics class and the criteria in the rubric clearly reflect this alignment. Although your task was aligned to the LP, it was unclear how your assessment strategies matched the learning targets. What I've done is extract the core elements of your task and attempted to capture it below. You'll notice that each criterion is equally weighted and the performance descriptors are sufficiently detailed to enable distinction between the categories. I've keep the five categories that you used in your rubric for the research project sculpture, but you might want to consider combining the unsatisfactory and below basic categories to have four performance categories to make the connection between your assessment strategies and SLO placement more explicit.

Criteria	Advanced	Proficient	Basic	Below Basic	Unsatisfactory
Sketching Detailed Drawings	The sketch is a detailed drawing with more than two notes. It includes measurements for height, width, etc. Colored pencils are used to show glazing strategies.	The sketch is a detailed drawing with one or two notes. It includes some measurements for height, width, etc. Colored pencils are used to show glazing strategies.	The sketch is a simple drawing with no notes. It includes some measurements for height, width, etc. or colors are used to indicate glazing strategies.	The sketch is a simple drawing without notes and lacking details. It does not include measurements for height, width, etc. It does not indicate glazing strategies.	The sketch is incomplete, or unclear.
Self-Reflection on Techniques	Students clearly describe how each of the 3 artists inspire the planned piece. The method for creating the piece is clearly described with reference to specific handbuilding and glazing techniques. Students specify what they need from the studio.	Students describe how 2 of the artists inspire the planned piece. The method for creating the piece is described with some reference to specific handbuilding and glazing techniques. Students describe what they need from the studio in general terms.	Students describe how 1 of the artists inspire the planned piece. The method for creating the piece is described in general terms without reference to specific handbuilding and glazing techniques. Students do not describe what they need from the studio.	Students describe inspiration for the planned piece, but do not reference any of the artists. The method for creating the piece is described generally techniques. Students do not specify what they need from the studio.	The self-reflection components of the proposal are incomplete or unclear.
Effective Demonstra-tion of Research Skills	Information about each of the three artists is accurate and detailed. All responses are in complete sentences.	Information about each of the three artists is accurate and detailed, but there are some inconsistencies or inaccurate information. All responses are in complete sentences.	Information about each of the three artists is general and lacks the detail requested in the prompts. Some information may be incorrect. Most responses are in complete sentences.	Information about each of the three artists is general. Much of the information about the artists is inaccurate. Responses are not in complete sentences.	The information about the artists is incomplete or unclear. Responses are not in complete sentences

Appendix J. Sample rubric from teacher leader pilot

	Advanced 4	Proficient 3	Partially Proficient 2	In Progress 1
ELG 4 Know and apply principles of design	Pushed the expected outcomes of the assignment while developing a strong final composition from detailed sketches Superbly manipulated plane's space Text integrated into composition purposefully	Demonstrated a clear understanding of the assignment and developed a final composition from sketches Manipulated plane's space well Text related to composition purposefully	Unclear understanding of the assignment desired outcome Developed a final composition from sketches The plane's space not used or manipulated effectively Purpose of text seemed unrelated to composition	Could not demonstrate a clear understanding of the assignment final composition weak or incomplete Ineffective manipulation of the plane's space Purpose of text seemed unrelated to composition
ELG 5 Know and apply basic techniques and elements	Superior use of lines to generate a complex and precise perspective environment Masterful used of value to define form	Effective use of lines to create a personalized perspective environment. Value enhances the 3-d forms within the environment	Lines used were not effective to create an accurate perspective environment Shading lacked the value needed to create the illusion of 3-d form	Lines lacked the precision to accurately represent perspective Little or no value in shading
ELG 5.1 Know and apply art materials	Strictly adhered to prescribed methodology Precise transformation of 2-d shape into 3-d form	Adhered to prescribed methodology Successful transformation of 2-d shapes into 3-d forms	Lacked an adherence to prescribed methodology Perspective off, resulting in a flawed 3-d forms	Ignored or didn't adhere to the prescribed methodology Composition flat, reads as 2-d shapes not 3-d forms
Manual Core Values	Scholar exemplified the Manual Core Value of Honor through integrity by leading through his/her actions, words and thoughts	Scholar exhibited the qualities of the Manual Core Value of Honor through integrity through his/her actions, words and thoughts	Scholar tried to extoll qualities of the Manual Core Value of Honor through integrity through his/her actions, words and thoughts	Scholar failed to extoll qualities of the Manual Core Value of Honor through integrity through his/her actions, words and thoughts