The Learning Progression Project Year 2 Pilot Findings: MATHEMATICS

September 2015

Derek C. Briggs
Fred Peck
Raymond Johnson
Jessica Alzen

CADRE

Center for Assessment, Design, Research and Evaluation (CADRE)
University of Colorado Boulder

Table of Contents

Executive Summary **3**

1. Overview of Study **4**

2. Methods **7**
- Data 7
- Student Focus Sessions 8
  - *Phase I of Student Focus Sessions* 8
  - *Phase II of Student Focus Sessions* 10
- Use of Task and Assessment Analysis Tools 11

3. Blue Elementary School **13**
- Participants, LP Topic, and Background 13
- Timeline 14
- Results 15
  - *Student Focus Sessions and Discussions about Student Reasoning* 15
  - *Outcomes of the Project* 18
  - *Teachers' Perceptions of LP-based SLOs.* 21

4. Green Beachway High School **24**
- Participants, LP Topic, and Background 24
- Timeline 24
- Results 27
  - *Student Focus Sessions and Discussions about Student Reasoning* 27
  - *Outcomes of the Project* 29
  - *Teachers' Perceptions of LP-based SLOs* 34

5. An Emergent Finding: Shifts in Teachers' Conceptions of Learning and Assessment **35**

6. Recommendations **38**
- Recommendation 1: Take Up the LPF Approach to SLOs 38
- Recommendation 2: Make Time Available for Teachers to Work Collaboratively 38
- Recommendation 3: Provide Teachers with Pre-made LPs and Electronic Interface for Data Entry that is Linked to the LP 38
- Recommendation 4: Work with Teachers to Move Beyond a "Count up points" Conception of Learning and Assessment 39

7. References **40**

Appendix A. Task and Assessment Analysis Tools **41**

Executive Summary

SLOs often suffer from two major problems than can undermine their usage. First, because student growth targets are often defined operationally in terms of points on an exam, the conceptual sense in which students have demonstrated growth is often unclear. Second, because SLO results are to be used for teacher evaluation, the accountability purpose is likely to trump other aspirations for formative use. In response to these problems, we have developed a *Learning Progression Framework* (LPF) for SLOs. The LPF is organized around the use of a *learning progression* that describes students' development of an overarching concept or skill that is required by state content standards. Growth is represented in terms of movement across discrete levels of the learning progression, and assessment tasks are written to distinguish students relative to their location on this continuum. A key feature of the LPF is the use of Student Focus Sessions in which teachers collaboratively examine student work on a small number of assessment tasks. A Student Focus Session culminates with ideas for how to improve the quality of tasks used to monitor student progress, and with instructional strategies that can be used to move students from lower to higher levels of the progression.

This report examines the application of the LPF among math teachers in an elementary school and high school in a large urban school district during the 2014-15 school year. The teachers in these schools all participated in 5-8 professional development sessions (ranging from 1-4 hours) over the course of a year. Throughout the project, Student Focus Sessions were used to monitor an SLO that had been created on the topics of place value (elementary school) and algebraic manipulations (high school). During the first few professional development sessions, teachers were guided through the Student Focus Session protocol by members of our research team, and then asked to implement the protocol independently. Subsequent professional development sessions focused on task development and the alignment of tasks to the underlying learning progression.

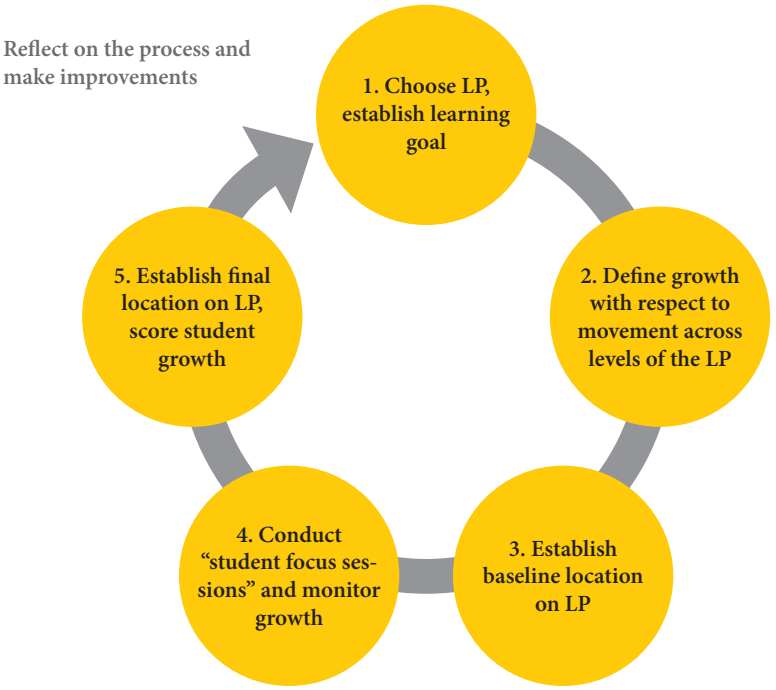
Teachers at both schools were able to independently conduct Student Focus Sessions, though not to the full extent that we had initially envisioned. Elementary school teachers in grades K-1 tended toward informal discussions of student work, and required more support in the form of user-friendly web-based supports for data entry than we had provided them. High school teachers implemented the approach with greater fidelity, but only implemented one of the two phases we had envisioned for Student Focus Sessions. Most teachers saw significant value in the LPF approach as a way to improve the quality of their classroom assessments, and for its emphasis on student reasoning. Many teachers reported changes to their classrooms practices as a result. The tradeoff to this perceived benefit was the recognition of the added cost in time and effort required to implement the approach, especially in the absence of the kind of expert facilitation and support provided by CADRE staff. An important emergent finding of this study was evidence of a shift in the way teachers were conceptualizing student growth. Prior to participation in the project, most teachers embraced a "count up the points" perspective about growth. Demonstrating growth was seen as a matter of students accumulating more points from a baseline period to and end period. By the end of this second year in the project, many teachers began to think of growth more developmentally with respect to movement across levels of a learning progression. Instead of just counting the points, these teachers were starting to ask deeper questions about what the points represent about student understanding.

1. Overview of Study

This report presents the results from the second year of a study to pilot a relatively novel approach for designing and implementing Student Learning Objectives (SLOs) according to a learning progression framework (LPF). The project took place in a large urban school district. Two schools participated in the project: Blue Elementary and Green Beachway (GB) High School[1].

The central components of the LPF, and the reason the use of this approach has some potential advantages in the context of SLOs, have been described in some detail in the CADRE Working Paper *Using a Learning Progression Framework to Assess and Evaluate Student Growth* (available at http://www.colorado.edu/education/cadre/learning-progression). 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). Figure 1 summarizes the implementation of an LPF approach to SLOs. The process begins in Step 1 with the choice of a topic around which a meaningful learning objective or objectives can be crafted. For the math teachers participating in the second year of this project at Blue Elementary and Green Beachway High School, the respective math content focus of SLOs was place value and algebraic manipulations. Both of these topics have a prominent position in the Common Core State Standards for Mathematics. In the first year of the project, teachers were given the chance to choose these topic areas on their own; if the process were to be applied on a larger scale, the range of possible topics may need to be constrained in order for a district to provide adequate support and relevant materials to all teachers in the district.

Figure 1. Stages of the Learning Progression Framework for SLOs

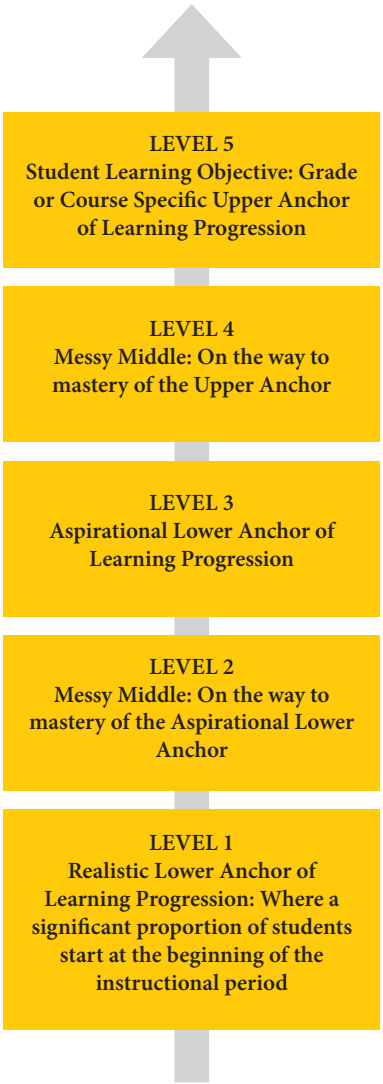


[1] All proper names are pseudonyms

Figure 2 gives an example of a generic learning progression (LP) with five levels that could be written by any teacher (or group of teachers) for students in a single grade or course in a given year. The top level ("Level 5") represents what is known as the "upper anchor" of the LP, and characterizes what students are expected to know and be able to do by the end of the instructional period. The middle level ("Level 3") represents the "aspirational lower anchor" of the LP, characterizing what knowledge and skills would be expected of students at the start of the instructional period if the students began the class fully prepared. The lowest level ("Level 1") represents the "realistic lower anchor" of the LP, characterizing the knowledge and skills observed among students who enter the class the most unprepared. The levels 2 and 4 of the LP represent levels of student understanding between the realistic and aspirational lower anchors, and the aspirational lower anchor and the upper anchor.

In the ideal implementation of the LPF, an LP has been established that crosses two or more grades or courses that have been sequenced under the assumption that students are becoming more sophisticated in their understanding of some big picture concept(s). This was the case in our work with math teachers as part of this two-year project. In Year 1, we worked with teachers to establish across-grade LPs for place value at Blue Elementary (spanning Kindergarten to Grade 5), and for algebraic manipulations at Green Beachway High School (spanning Algebra I to Pre-calculus). For details on the development of these LPs in Year 1 of the project, see *Learning Progressions Project: Documentation of Pilot Work and Lessons Learned in the 2013-2014 School Year* (available at https://www.colorado.edu/education/node/1797/attachment).

Figure 2. A Generic Course-Specific Learning Progression



The second step of the process depicted in Figure 1 is to define growth in terms of movement across levels of the LP. This involves a determination of the amount of movement necessary for a student to show "one year's growth." With respect to the generic LP shown in Figure 2, the default movement for one year's growth would be two levels (the distance from the aspirational lower anchor to the upper anchor). The third step is to establish each student's baseline location on the LP at the beginning of the instructional period. This is done by both referencing relevant information about student performance in prior grades/courses, and by administering baseline assessment items that are aligned to the LP. The fourth step is to monitor student growth over the course of the instructional period, and the fifth step is to establish student locations on the LP at the end of the instructional period. Although student growth monitoring (Step 4) is the longest step in the process, it is also the most important one if an SLO is expected to have an impact on teachers' instructional practices. In recognition of this, the focus on our project in Year 2 was to engage participating teachers more directly with their students' work over the course of the school year.

To accomplish this we introduced two new tools in our sessions. The first is a protocol for what we call *Student Focus Sessions*. The second is a set of *Task and Assessment Analysis Tools* that can be used to evaluate the quality of assessment items/tasks that are the basis for the student work shared among teachers as part of Student Focus Sessions. The purpose

of a Student Focus Session is to get teachers working collaboratively to make inferences about student reasoning and understandings based on analyzing student work, and to make connections between these inferences and their implications for student locations on the LP. The purpose of the Task and Assessment Analysis Tools is to evaluate the quality of the tasks and assessments comprised of these tasks that are being used to elicit student work. Details regarding the tools themselves, and the results from sharing these tools with participating teachers are described in the sections of the report that follow.

Before we proceed, some brief clarification is in order regarding the term "student reasoning." When we use this term, we intend the general definition offered by the National Council of Teachers of Mathematics (2009, p. 4) as "The process of drawing conclusions on the basis of evidence or stated assumptions." In other words, how a student goes from the evidence given in a task to an answer to the task. This requires teachers to pay careful attention to the way that students explain and/or justify their answers to questions. Such attention is central to a learning progression approach, as this information is what best establishes a student's location on the LP, and also consistent with the emphasis found in the Common Core State Standards. However, paying careful attention to student reasoning represented a shift in emphasis for many teachers participating in the project, many of whom focused exclusively on whether students could answer math questions correctly or not, rather than on the strategies students used to solve the questions.

Research Questions Relevant to Year 2 of Pilot Project

1. Student Focus Sessions
    a) How do teachers enact Student Focus Sessions?
    b) To what extent can Student Focus Sessions be conducted solely by teachers without outside facilitation?
    c) How do teachers discuss student reasoning in Student Focus Sessions?
2. Outcomes of the Learning Progression Project (LPP)
    a) Is the LPP associated with an improvement in the quality of assessment tasks?
    b) Is the LPP associated with changes in teachers' instructional practices?
3. Do teachers perceive differences between the LP-based SLOs and SGOs[2]?

[2] Prior to our project, SLOs were referred to as "student growth objectives" (SGOs) and were implemented without connection to an LP.

2. Methods

DATA

This report relies on five sources of data:

1. Video and audio recordings of professional development (PD) sessions held with the teachers from each school, including recordings of small-group conversations within and across grade-level teams.
2. Audio recordings of the Student Focus Sessions run independently by the teachers without members of our research team.
3. Audio recordings of a set of year-end one-on-one interviews from a subset of teachers at each site (three from Blue Elementary School and four from Green Beachway High School). These interviews, each about 20 minutes in length, were conducted either in person, remotely using an internet video conferencing service, or by email.
4. The assessment items and associated rubrics developed by teachers throughout the project.
5. Teacher's written responses to an anonymous online survey given at the beginning and end of the year which focuses on teachers' attitudes and beliefs about the SLO process.

We created content logs for all of the recorded data, and coded these logs to correspond to the research questions that motivated the project. We then sorted the coded segments, aggregated them by code, and analyzed the aggregated segments for emergent themes and patterns. In some cases, this involved sub-coding. For example, as we discuss below, we noticed that conversations about student reasoning tended to be either about specific students, or general summaries of multiple students. To explore this emergent finding further, we coded each "student reasoning segment" as either *specific* or *general.*

Table 1 shows the data sources we used to answer each research question.

Table 1. Data sources for each research question

RQ	Data sources
1a: How do teachers enact Student Focus Sessions?	Video of PD sessions, Audio of Student Focus Sessions
1b: Can Student Focus Sessions be conducted independently by teachers?	Video of PD sessions, Audio of Student Focus Sessions
1c: How do teachers discuss students reasoning in Student Focus Sessions?	Audio of Student Focus Sessions, Year-end interviews
2a: Quality of assessment tasks	Audio of Student Focus Sessions, Assessment tasks, Year-end interviews, Survey
2b: Effect on instructional practices	Video of PD sessions, Audio of Student Focus Sessions, Year-end interviews, Survey
3: Differences between LP-based SLOs and SGOs	Year-end interviews, Survey

STUDENT FOCUS SESSIONS

A major focus of Year 2 of the project was on Student Focus Sessions (SFSs). During these sessions, teachers were to have collaborative, structured conversations around student reasoning with two goals in mind. The first goal was to learn more about how students are reasoning about tasks, and to design instructional moves and classroom activities that are response to students' reasoning. In general, observations about student reasoning are most helpful when assessment tasks have been written in a constructed response format and aligned with the LP. Given this, a second major goal of SFSs was to improve the quality of assessment tasks. We defined improvement of assessment tasks in two ways:

- Improve the *validity* of the task by strengthening the connection between the task/rubric and the learning progression.
- Improve the *reliability* of task scores by writing tasks with unambiguous scoring rules.

In order to meet both goals of an SFS, we developed a protocol in which every teacher has one of three defined roles: lead teacher, recorder, or participant. Each SFS is facilitated by the lead teacher. This teacher selects tasks, representative student work on these tasks, and facilitates the meeting. The recorder takes notes that characterize the group's discussions related to the quality of the selected tasks, student reasoning, and responsive classroom activities. The recorder is also a participant. Participants score tasks and engage in discussion around scores and student reasoning. They do not have any "extra" responsibilities.

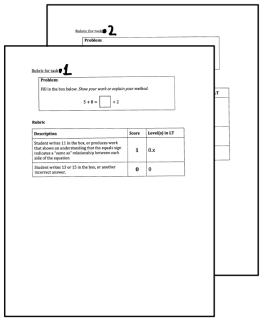
An SFS has two phases, each lasting approximately one hour. The phases can be completed in a single two-hour session or split up over two one-hour sessions. In Phase I, all participants score a sample of student work provided by the lead teacher on common tasks. They then discuss any discrepancies in their scores and arrive at a consensus score. Phase I culminates in a discussion of ideas for modifying the tasks and/or rubrics so that such scoring discrepancies across teacher can be minimized. Phase II ends with teachers offering qualitative descriptions of the strengths and weaknesses of each student as well as ideas for instructional strategies that could be used to help their students make further progress along the LP. Below we describe these phases in more detail. (The complete protocol can be downloaded from the CADRE website at https://www.colorado.edu/education/node/1791/attachment).

Phase I of Student Focus Sessions

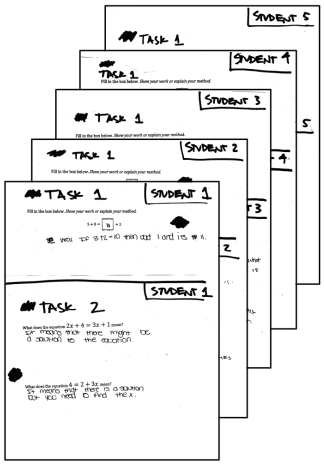
In preparation for Phase I, the lead teacher selects two tasks for a sample of five students. The lead teacher then provides anonymized copies of the student work as well as rubrics for each task in a participant packet. A sample packet is depicted in Figure 3. The lead teacher is expected to use the following criteria when selecting tasks and students:

- Choose tasks and students that will be interesting to discuss.
- Choose tasks that allow students to demonstrate the reasoning they used to arrive at an answer.
- Choose students that are representative of the range of performance on the tasks.

Figure 3. Sample SFS participant packet



Rubrics for both tasks



Anonymized student work (up to 5 students)

At the beginning of Phase I, the lead teacher introduces the tasks. She explains how the tasks are related to the LP, and she explains why she chose these particular tasks. Following this, participants independently score the student work in the packets using the provided rubrics. Scores are recorded in a central location visible to all participants (e.g., on a shared Google spreadsheet, projected on a board, or written on large paper at the front of the room) (See Figure 4 below for an example). All of this can take anywhere from 10 to 20 minutes, depending upon the complexity of the task, the length of the responses, and the nature of the scoring rubric.

Next, the lead teacher facilitates a discussion of the scores. The goal is to develop a consensus score for each student-task combination, which the lead teacher records, and to modify tasks and rubrics so as to minimize the possibility of future scoring disagreements. Discussion should begin with those task-student combinations that have many disagreements so that if time runs out, the most contentious task-student combinations have been addressed. In addition to discussion around factors which led the scores to differ, participants should also share any other modifications to tasks or rubrics that they think should be made. During this discussion, the recorder takes notes regarding attributes that cause disagreement in scoring, and the group's ideas for modifications to minimize future disagreements. This should all take approximately 30-45 minutes. If time runs out and consensus has not been reached on all student-task combinations, the lead teacher should record the modal score for that column as the consensus score.

Figure 4. Consensus spreadsheet

Few disagreements here – Save this for last

Many disagreements here – Choose this column first

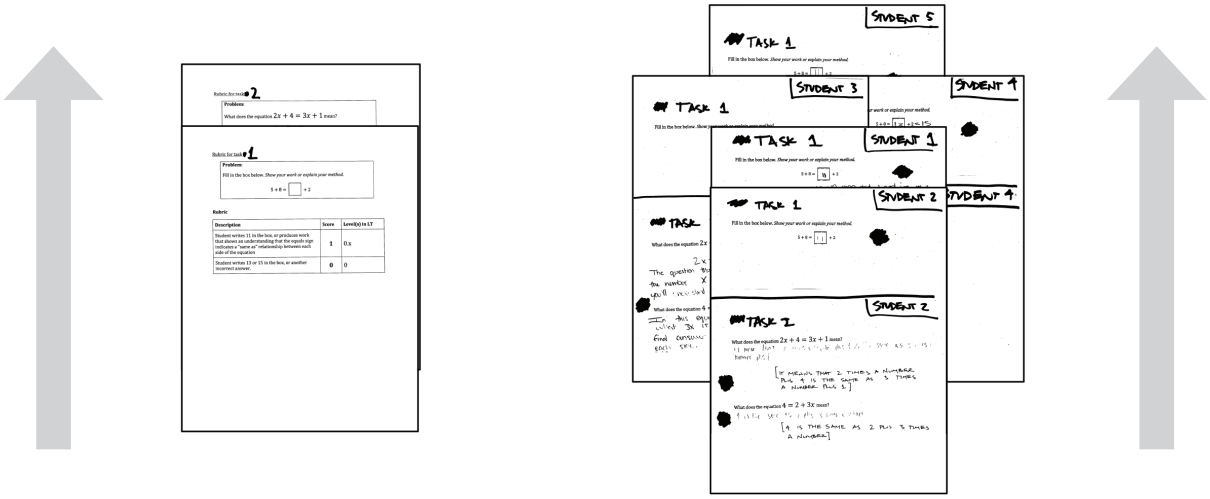
	Item 1					Item 2				
Rater	**Student 1**	**Student 2**	**Student 3**	**Student 4**	**Student 5**	**Student 1**	**Student 2**	**Student 3**	**Student 4**	**Student 5**
Bunk	0	0	1	2	2	0	0	1	2	2
McNulty	1	1	1	2	2	1	0	1	0	2
Greggs	1	0	2	2	2	0	0	1	1	2
Landsman	1	0	0	2	2	1	1	1	2	1
Presbo	1	1	2	0	2	1	0	1	1	2
Beadie	0	1	2	2	2	1	0	0	1	1
Sabotka	0	1	2	2	1	1	0	2	1	2
Cutty	1	1	2	2	2	0	1	1	2	1
Daniels	1	0	1	1	2	0	1	2	2	0
Mode	1	1	2	2	2	1	0	1	2	2
# of disagreements with mode	3	4	4	2	1	4	3	3	5	4
SD	0.50	0.53	0.73	0.71	0.33	0.53	0.50	0.60	0.71	0.73
Consensus rating										

Lead teacher writes consensus scores here

Phase II of Student Focus Sessions

To begin Phase II, participants formally map the tasks scored in Phase I to the LP underlying the SLO. That is, teachers should be able to indicate whether a given task would help to discriminate between students at level 2 vs 3 of the LP, or 3 vs 4, or 2 vs. 3 vs. 4, etc. This act puts teachers in the mindset of making connections between tasks, the LP, and students for the duration of the SFS. Teachers then take 10 minutes to examine student work and decide which task is more difficult than the other based on the LP. Next, they put the students in ascending order on the LP based on a holistic appraisal of the student's reasoning as it relates to levels of the LP (See Figure 5).

Figure 5. Beginning of Phase II



Participants share their rationales for ordering students and tasks. If there is not an opportunity for students to use multiple strategies in coming to a solution for a given task, participants discuss modifications that could be made so that it would be possible to observe multiple strategies in the future. If there is an aspect of the task that is causing differences in student responses that is not aligned to the learning progression (e.g., a vocabulary word), participants discuss ideas to modify the tasks to remove this "progression-irrelevant variance." During this discussion, the recorder takes detailed notes about what the participants notice in terms of student reasoning. The recorder also notes ideas for changes to tasks and rubrics.

Next, the lead teacher provides the "Consensus Score Sheet". A sample Consensus Score Sheet is show in Figure 6.

Figure 6. Sample Consensus Score Sheet

	Student 1	Student 2	Student 3	Student 4	Student 5	Mean scores
Item 1	1	0	2	2	2	70%
Item 2	0	0	1	1	2	40%
Total score	25%	0%	75%	75%	100%	

The consensus scores are automatically copied over

Points possible	
Item 1	2
Item 2	2

The lead teacher enters the maximum points possible for each item

Participants compare their holistic ordering of students with regard to their reasoning with the order given by the analytic consensus scores. In the example above, the tasks would be ordered: task 1 (easier), task 2 (harder). Students would be ordered in ascending order according to their analytic scores as follows: 2, 1, 3/4, 5. If the holistic and analytic orderings don't match, this becomes a topic for discussion because it indicates that are important distinctions in student reasoning related to the LP that are not captured by the rubric. In that case, participants discuss ideas to modify the rubric so as to capture these distinctions. If there are two students with the same score (like 3 and 4 in the example above), participants examine the student work to determine if the students are showing evidence of responses that are comparable in terms of the levels of the LP. If the participants agree that the students are not comparable in terms of the LP, they discuss ideas to modify the rubric to capture the distinctions. The recorder notes all ideas related to the task, rubric, and student reasoning.

Teachers spend the next 30 minutes of Phase II making inferences about students' understandings based on the tasks analyzed, and considering instructional strategies that are responsive to student understanding. To begin this conversation, participants write a 2-3 sentence narrative description of each student, capturing inferences about what the student understands and struggles with, and relating that to levels of the LP. Participants then discuss instructional strategies that might be helpful for both these specific students, and for other students who may have similar strengths and weaknesses. The recorder notes the instructional strategies discussed for each student.

By the end of Phase II, teachers will have discussed tasks, student reasoning, and classroom activities in detail. These discussions should be captured in the recorder's notes, which should include the following elements:

- Detailed ideas for how to improve each task and rubric.
- Detailed notes about the various forms of student reasoning that have been observed, including differences in strategies used to solve a task, and the participants' thoughts about the sophistication of these strategies.
- Narrative summaries of each exemplar student and ideas for instructional strategies that could be incorporated into subsequent classes.

USE OF TASK AND ASSESSMENT ANALYSIS TOOLS

In order to evaluate the quality of assessment tasks, our team developed Task and Assessment Analysis Tools to assess the quality of individual tasks as well as overall assessments written according to the LPF. We created these tools based on our sessions with teachers, as well as our informal analyses of teacher-created tasks and assessments. (The tools are given in Appendix A, and can be downloaded from the CADRE website at https://www.colorado.edu/education/node/1795/attachment).

Quality of tasks should be considered based on the following criteria:

- *Relevance to the learning progression:* the extent to which task and rubric are relevant to the LP.
- *Options for expressing understanding:* indicates if the task prompts students to reveal their understanding in only one way (such as with a closed-ended problem like multiple choice or fill-in-the-blank) or in multiple ways (such as with open-ended problems that ask for multiple representations of a solution, or a task that asks for a procedure with a justification, or other means of proving multiple options to express understanding).
- *Cognitive demand:* the extent to which students must make connections between processes and underlying concepts (based on the framework for cognitive demand given in Smith, Stein, Henningsen, & Silver, 2009).
- *Rubric reliability:* indicates if there is a high probability that the task could be scored reliably by any teacher in the respective area and grade level.
- *Rubric validity:* indicates that: (a) the rubric covers everything that students are asked to do (e.g., if the task asks students to "show work" the rubric gives guidance as to how to score the work), and (b) the rubric comprehensively covers the range of possible student responses. If there are multiple possible responses, the rubric gives guidance as to how to score likely or common responses.
- *Rubric specificity:* indicates that all adjectives and general statements (e.g., "shows understanding" or "solves problem correctly") in the rubric are accompanied by specific descriptors related to the problem. For example, if the rubric says "solves problem correctly" the correct answer(s) for the problem is given in the rubric.

- *Fairness:* indicates if the material is familiar to students from identifiable cultural, gender, linguistic, and other groups; is free of stereotypes; can be reasonable completed under the specified conditions, and if students will all have access to resources necessary for task completion (e.g. Internet, calculators, spellcheck, etc.).
- *Clarity:* indicates if the wording in the task and instructions is clear and grammatically correct. The task and instructions are free of wordiness, irrelevant information, unusual words, and ambiguous words.

We used the Task Analysis Tool to rate the quality of assessment tasks, and we also see it as a valuable tool to be taken up by teachers during the SFS as well. The Task Analysis Tool should be used both during the creation of new tasks or consideration of existing tasks as well as during SFSs as the need for additional adjustments to tasks may become obvious through collaborative discussion.

In addition to considering tasks individually, assessment quality as a whole should be considered with respect to the following categories:

- *Alignment to the learning progression:* the bulk of the tasks cover the range of the LP where teachers expect most students to be, with some tasks below this range and some tasks above this range.
- *Options for expressing understanding:* the assessment includes tasks that have only one way to express understanding as well as those which have more than one way to express understanding.
- *Fairness:* assessment questions/prompts are marked with graphic or visual cues (e.g. bullets, numbers, in a text box, etc.) and the task format is consistent, visually clear, and uncluttered.

The process our team undertook for task rating followed the example set forth by Hallgren (2012) and included both a group training phase, an individual rating phase, and a group consensus phase. These phases proceeded as follows:

Training phase:

1. We used 24 tasks from Spruce Middle School (a school that participated in Year 1 of this project, but not Year 2) for training.
2. Using a random number generator, we selected four tasks as "practice tasks." From this set, we chose two tasks to practice with each from a different teacher.
3. For each practice task, the full group rated the quality of the practice task independently using the criteria listed above. We discussed our disagreements and came to a consensus rating. We then refined the Task Analysis Tool scoring criteria to reflect our shared interpretations.
4. After two practice tasks, the raters believed that the instrument was reliable enough for them to rate the tasks from Blue and Green Beachway.

Individual rating phase:

1. Two raters each rated all available tasks (baseline and final tasks from Blue and GB).
2. All discrepancies were tabulated.
3. Inter-rater agreement was computed.

Group consensus phase:

1. The full group discussed disagreements and came to a consensus on task ratings.

Our research team did not apply the Assessment Analysis Tool, but teachers at each campus used the tool as they prepared their final assessments.

3. Blue Elementary School

PARTICIPANTS, LP TOPIC, AND BACKGROUND

Blue Elementary participated in both Year 1 and Year 2 of our pilot study. Four kindergarten and four 1st grade teachers participated in Year 2. Six of these teachers had participated in Year 1 of the pilot, but each of the two grades had a first-year teacher who was new to SLOs and to the project in Year 2. The participation by only kindergarten and 1st grade teachers was a significant change from Year 1, where a single learning progression for place value was developed across grades K-5. The Blue Elementary principal and CADRE researchers agreed to reduce the number of grades in Year 2 to K-1 to address what was seen as the greatest assessment needs, as well as to alleviate some of the prior year's challenges related to working with a large number of teachers representing vastly different levels of student ability. None of the teachers in Year 2 claimed any particular expertise or specialized preparation to teach mathematics. For example, none claimed familiarity with the National Council of Teachers of Mathematics or early elementary strategies for teaching mathematics such as Cognitively Guided Instruction or Math Recovery.

In Year 1, the teachers at Blue Elementary focused their SLO on a learning progression that described the mathematical topic of place value. Table 2 shows this learning progression for Grades K-1.

Table 2 Place value learning progression for Blue Elementary Grades K-1

Level	Relative position / anchors	Description
7	1st aspirational upper 2nd aspirational lower 3rd realistic lower	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. 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.
6	1st 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
5	K aspirational upper 1st aspirational lower 2nd realistic lower	**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 nine ones. Students can understand that 10 can be thought of as a group of ten ones, called a "ten".**
4	K 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.**
3	K aspirational lower 1st realistic lower	**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.**
2	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.**
1	Pre-K aspirational lower K realistic lower	**Students can recognize numbers 1-3. Students can verbally count to 3. Students can accurately count up to 3 objects.**

TIMELINE

We met with K-1 teachers at Blue Elementary for five 4-hour sessions during the 2014-2015 school year, and we spent additional time in classrooms to study possible ways to assess kindergarten students. In an initial meeting with the Blue principals on August 15, we decided to have four, 4-hour sessions, one each in September, October, January, and March. A fifth session was added in May to give the group an opportunity to work with end-of-year data and to calculate student growth. All eight teachers attended each of the five SLO work sessions. A Blue Elementary principal or assistant principal attended each of the sessions, and often a second non-teacher from the district also attended, such as an SLO coach or district evaluator. These non-teachers participated in discussions but were not the focus of the professional development or data collection. Raymond Johnson was the primary facilitator and researcher working with the Blue teachers, with Derek Briggs, Fred Peck, and Jessica Alzen each helping co-facilitate one or more of the sessions. The dates and primary topics for the five sessions are in Table 3.

Table 3. Schedule of Year 2 Sessions at Blue Elementary

Session	Date	Primary Topic(s)
1	September 18, 2014	Exploring the use of video assessment data
2	October 30, 2014	Introducing *Realistic Mathematics Education* (RME). Planning for assessments and data collection
3	January 15, 2015	Placing students on the LP using baseline data
4	March 12, 2015	Student Focus Session with Kindergarten video data
5	May 14, 2015	Student Focus Session with 1st grade data

In addition to the 20 hours of SLO PD sessions, researchers visited Blue to observe teachers on two occasions. The first, on September 18, 2014, was to observe a kindergarten teacher's attempt to assess students in small groups using a pencil-and-paper test. The second, during several days around the second week of December, was to observe kindergarten teachers assess students one-on-one using the district's mid-year interim mathematics exam.

Briefly, the focus of each of the five work sessions was as follows:

- *Session 1:* Participants reflected on Year 1 and reviewed the approach to using learning progressions to measure student growth. This led into a more detailed discussion concerning the place value LP used at Blue and the kinds of assessments needed to measure student growth along the LP. Several key goals were set, including: (a) doing a better job meeting the needs of K-1 teachers, (b) shifting from creating the LP to using the LP as a basis for assessment, (c) improving assessment tasks and rubrics, and (d) focusing on how well students understand and reason with place value, beyond just improving their scores on place value tasks. As an example of an approach to use to improve assessments in K-1, the group reviewed several video examples from other projects in which students were assessed in small groups or one-on-one. Teachers were also introduced to *Realistic Mathematics Education* (Freudenthal, 1983, 1991; Treffers, 1987) as a model for how student reasoning becomes progressively more sophisticated over time and how classroom activities can be designed to promote this. Finally, these ideas were combined in a draft plan for assessment approaches for the rest of the year.

- *Session 2:* Teachers focused on inventorying tasks that were relevant to the LP and determining what tasks and resources might yet be needed. Teachers were given time to work in grade-level teams to organize baseline data from assessments earlier in the year and use that data to locate students on the LP.

- *Session 3:* Teachers participated in their first Student Focus Session. Using video data collected from the kindergarten interim exam in December, CADRE staff facilitated the Student Focus Session using video clips of three tasks performed by five students. This was followed by a discussion of the challenges of doing Student Focus Sessions in Grades K-1.

- *Session 4:* Teachers participated in their second Student Focus Session, this time using several tasks from 1st grade. The session was co-facilitated by a 1st grade teacher and a researcher, with the teacher taking sole responsibility for choosing the tasks and student work for the session. The end of Session 4 shifted to the use of the Task and Assessment Analysis Tools to gauge assessment quality, then finished with planning for collecting end-of-year data.

- *Session 5:* Participants focused on aligning end-of-year tasks to the LP and then using student scores from those tasks to place students on the LP. A preliminary calculation of student growth was made, with plans for final calculations to come in the weeks following the session as teachers administered their final tasks of the year.

RESULTS

Student Focus Sessions and Discussions About Student Reasoning

A key goal of the Learning Progression Project at Blue Elementary was to focus on student reasoning in the area of place value. The systematic process under which we attempted to do this was through the use of *Student Focus Sessions* (SFSs), in which teachers examined the work of five students in close detail, both to better understand student reasoning demonstrated on the tasks and to scrutinize and improve the quality of the tasks and rubrics themselves. Teachers at Blue Elementary participated in three SFSs. CADRE researchers facilitated the first two SFSs during the 3rd and 4th sessions, and kindergarten teachers independently conducted an SFS in May. Unfortunately, an audio recording of this final session was lost and the teacher who facilitated the session went on leave soon after the session was conducted, so relatively little is known of the happenings of this session. Therefore, findings related to teachers' focus on student reasoning comes mostly from the two SFSs in Sessions 3 and 4, with additional data from interview responses. The research questions answered in this subsection are:

1.a. How do teachers enact Student Focus Sessions?

1.b. To what extent can Student Focus Sessions be conducted solely by teachers without outside facilitation?

1.c. How do teachers discuss student reasoning in Student Focus Sessions?

During the first two SFSs, facilitated or co-facilitated by the researchers, teachers followed the phases and procedures described in the SFS handbook. Teachers recorded their scores of the student work in an online scoresheet and took notes in a shared document. Student work examined in the first session consisted of video clips of three tasks from five students that were recorded during the district midyear assessment administered in December, 2014. The collection of the video and the editing and selection of the video clips was all done by the researchers. Video data in a SFS provided a very rich source for examining student reasoning, but also provided certain limitations. For example, with the video data, teachers were unable to

Figure 7. Teacher's note-taking sheet for reviewing video of 1-on-1 assessment.

look across all videos simultaneously, such as one might do with five tasks on paper. This proved to not be a significant hindrance in analyzing student work. A teacher note-taking worksheet, illustrated with an image from the videos of each of the five students, helped teachers keep track of their observations as the group viewed the videos (see Figure 7). The student work examined in the second SFS consisted of written work from 1st grade students. These were chosen and copied by a teacher who served as co-facilitator.

Following the first SFS, which was facilitated by the researchers, the Blue Elementary teachers questioned their ability to conduct a SFS by themselves. Two factors influenced this doubt: one, a lack of experience collecting and editing video data, and two, unfamiliarity with using spreadsheets to collect and organize teachers' scores during the SFS. Researchers spent an estimated 10 hours collecting and editing the video for this SFS, time teachers are very unlikely to have if SFSs are to be conducted on a regular basis. Several kindergarten teachers experimented with collecting video with iPads during class activities, but without an over-the-shoulder view of students, it was difficult to see their work. Additionally, teachers had multiple problems organizing and sending video clips to the researchers. As for the spreadsheets, only one of the eight teachers claimed any experience in using them to do any kind of calculations. A simplified scoresheet was provided to the kindergarten teachers for their attempt to do an SFS without researcher facilitation. Afterwards, a teacher in an interview admitted that they never used it, preferring "just talking to each other," feeling that "it seemed to give us more time to discuss what we were seeing" and that "not having to deal with the technology aspect of it and somebody's computer isn't working or somebody hit the wrong button…it seemed [to] speed the process up for us and gave us more time to discuss what we were seeing." For teachers to regularly conduct SFSs on their own, time will need to be given to conduct the sessions and support in using any required technology. The second SFS, using written 1st grade tasks, gave teachers more confidence that they might conduct SFSs regularly and on their own if given time and support beyond that given in the pilot.

SFSs were successful in focusing teachers on student reasoning with place value. With none of the teachers having significant, specialized training to teach mathematics, observing student reasoning at this level of detail was new for most teachers. As one teacher said during her interview:

> This year was very different because we never went so deep into any other standards or topics. That was something that I never had before. All the process with the students, and the assessments, and the rubrics, and everything … it was something very new for me. That helped me to just go deeper into understanding the reasoning, the thinking of my students.

Also in an interview, another teacher reported developing her skill in noticing student place value reasoning:

> It's very interesting for me to see just how students think about place value, especially when you get out of teen numbers. It's like a whole other world for kindergarten when you get out of teen numbers and have to add an additional 10 because then you're putting in their ability to count by tens, which is something that a lot of students struggle with, and not only to count by tens but also to go back and count by ones to give you what else you had in the ones place.

The SFSs surfaced numerous examples of what student reasoning with place value can look like. Most prominent became observing students' recognition of ten as one quantity, rather than a quantity of ten units, a skill explicitly called for in the Common Core State Standards. This was seen in the difference between students who recognized the quantity ten in base-10 blocks or ten frames and those students who still needed to count them individually starting at one. For at least one teacher, use of these manipulatives in this way was new, as revealed in an interview:

> This year we just started using the [ten frames] and the manipulatives with the ones and tens. We didn't use those in the past. So this year we just ordered them for kindergarten because we never had those in kindergarten before. So this was new, and I feel like the kids did a very, very good job using the manipulatives and that helped them to just understand the concept.

The SFSs were especially effective in guiding teachers to focus on student reasoning *in relation to a particular task.* That is, teachers focused their attention to the strategies students used or potentially could have used on a given task. Somewhat less attention was given to comparisons of reasoning in relation to the LP. A key facilitation move became the stressing of *meaningful differences* in reasoning, particularly for two students who might be placed at the same

level of the rubric yet have noticeably different solution strategies. *Meaningful* was framed in relation to the LP, with a desire to represent different levels of the LP with different levels of a task's rubric. Some progress was made towards this goal, but teachers occasionally held onto prior practices such as making the number of mistakes in student work the key differentiator in rubric levels, or assuming two students with the same number of points earned on a task should automatically be placed on the same level of the LP, regardless of the strategy used to complete the task. The SFSs provided repeated opportunities to confront these issues and orient teachers towards more LP-aligned ways of thinking, but with a limited number of SFSs with the Blue Elementary teachers, more repetition and experience would be needed to bring all teachers into alignment with an LPF approach to measuring student growth. We discuss this further in Section 5.

Two unexpected issues of teacher concern or interest surfaced during SFSs. One was a concern, primarily from one teacher, that analyzing the work of only five students was insufficient to understand how all of her students were performing on a task. In other words, this was a concern about sample size, or whether sampling itself was appropriate to use when teachers are accustomed to evaluating the work of every student. The second issue was a shared belief amongst a number of teachers that perhaps *student confidence* should be a differentiating factor in the LP and in task rubrics. This arose when watching the video clips during the first SFS and teachers noticed hesitation or a questioning tone when some students answered teachers' questions. As researchers we found this noticing of student behavior interesting but did not pursue including variations in confidence in the LP or in task rubrics. The LP should work best when measuring a single construct, in this case students' ability to reason with place value, and a focus on speed of response is unnecessary and potentially detrimental to students' learning of mathematics.

Summary: Brief answers to RQs 1a, 1b, and 1c
In summary, teachers at Blue Elementary conducted three SFSs, two of which were facilitated by researchers. Table 4 shows our answers to RQs 1a, 1b, and 1c for Blue Elementary based on the analysis of the two researcher-facilitated SFSs and interview data.

Table 4. Brief answers to RQs 1a, 1b, and 1c for Blue Elementary.

Research question	Brief answer
1.a. How do teachers enact Student Focus Sessions?	Teachers used SFSs to engage in issues related to student reasoning and task quality. SFSs were conducted with both pencil and paper data and video data of 1-on-1 assessment of students, the latter of which provided a rich view of student reasoning but required significant effort on the part of researchers to collect and edit the video.
1.b. To what extent can Student Focus Sessions be conducted solely by teachers without outside facilitation?	Teachers expressed doubt in their ability to conduct SFSs on their own. Reasons teachers cited included a lack of time in their current schedules to conduct the SFS and a lack of knowledge about using spreadsheets to use the scoring spreadsheets needed for the SFSs. In the SFS without researcher facilitation, teachers followed the SFS protocol loosely, preferring to use their time to talk about tasks and student work rather than use the scoring spreadsheets and consensus tools described in the SFS handbook.
1.c. How do teachers discuss student reasoning in Student Focus Sessions?	Teachers mostly discussed student reasoning with respect to a particular task, focusing less on the development of reasoning over time. Teachers' took an interest in how student reasoning was reflected in students' use of mathematical manipulatives, and wondered if students' confidence in answering questions reflected greater understanding.

Outcomes of the Project

In this section we address the following research questions concerning the Learning Progressions Project (LPP), the quality of assessment tasks, and impacts on instructional practice:

a) Is the LPP associated with an improvement in the quality of assessment tasks?

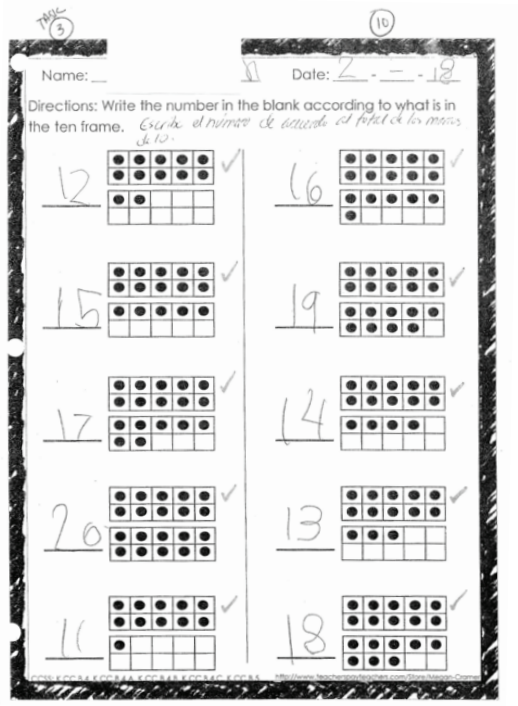
b) Is the LPP associated with changes in teachers' instructional practices?

Quality of assessment tasks.
A second goal of the Learning Progression Project was to improve the quality of tasks and rubrics used to assess student progress along the place value LP. While Task and Assessment Analysis Tools (See Appendix A) was developed to approach task quality in a systematic way, at Blue Elementary the need to find approaches to assessment that worked with the needs and abilities of K-1 students generally took precedence, particularly for the kindergarten teachers. The tasks and rubrics used at Blue needed to account for students who may not yet have the ability to read, write, or do either without some assistance. Without this, other measures of task quality would not be relevant.

In our initial session, some teachers expressed the feeling that they were "unqualified" to write assessments for their students. After reviewing a video of a teacher assessing a student one-on-one, teachers noted that such opportunities were rare for them, and perhaps some exploration of small group assessment might be possible. In September, kindergarten teachers attempted to administer a more traditional, pencil-and-paper assessment with groups of 8-12 students at a time. Students sat spaced apart on the floor, with a lap board, a pencil, and their assessment, while the teacher stood at the front and read each question. In general, the tasks were not of high quality, either because they were not relevant to the LP, they relied solely on memorization ("Write the number 7"), or the directions were unclear to students. The tasks were written without associated scoring rubrics.

During the assessment, students struggled to follow along and many looked at the papers of students near them. The pace of the assessment was also an issue, as one group took 17 minutes to complete a 7-question assessment, a duration difficult for kindergarten students working on their own. After attempting the assessment with two groups, the teacher opted to abandon the assessment, feeling that she was not collecting worthwhile data or that students' time was being used wisely. A brief meeting with the kindergarten team followed, and it was decided that such a traditional form of assessment via paper and pencil was not a good option for kindergarten, even in small groups. Either assessment data would have to be collected in a way that fit in the normal routine of classroom activity, or collected in a one-on-one setting. Opportunities to assess students one-on-one are rare for teachers, but teachers saw the possibility of leveraging the district's interim assessment as most promising, which was scheduled to take place in a one-on-one setting at three different points during the school year.

Figure 8. A kindergarten task administered in February 2015



Kindergarten teachers obtained an advanced copy of the midyear interim exam, to be administered in December, and reviewed it for tasks with relevance to the LP. Three such tasks were identified. One of the tasks, involving the counting of 17 cubes, was supplemented with having students construct the number 17 using base-10 blocks. The base-10 blocks helped teachers see which students recognized ten as a special quantity, versus those who needed to count the group of 10 blocks as if they were independent cubes. A supplementary rubric was created for this task to record some of the predicted strategies students might use on the task. After receiving permission to record students, researchers came to Blue Elementary to video record students completing these tasks. When selections of these videos were used at the SFS in January 2015, teachers noticed not only differences in

student reasoning about place value, but differences in how they administered the tasks and applied the rubrics. While perhaps not a universal occurrence, teachers in this discussion seemed to be wrestling with the important distinction between rubrics that reflect *a number of points earned out of a total versus rubrics that described levels of sophistication of reasoning as it related to the place value LP*. This is the other side of *meaningful differences* discussed in the previous subsection: where there are meaningful differences in student reasoning, the rubric should distinguish them, and therefore rubrics should be created with potentially meaningful differences in mind as they relate to the LP.

Task and rubric quality at Blue Elementary can be illustrated with several representative tasks used during the 2014-2015 school year. Figure 8 shows a kindergarten task from February of 2015. Although the task does involve place value and the use of ten frames, unless teachers closely supervised students completing the tasks, they would not know if students counted each dot starting at one, or knew a full ten frame represented 10 and counted on from there with the remaining dots. Without observing the student, the task could provide evidence that a student has progressed as far as Level 4 (kindergarten messy middle), because it gives students an opportunity to count up to 20 objects. By observing students performing this task, a teacher would have evidence that student had achieved the Level 5 (kindergarten upper anchor) because recognition of ten in the ten frame would be part of composing a teen number with ten and ones.

Figure 9 shows the rubric for the task shown in Figure 8. It is also problematic because it allocated points solely on the basis of correctness, and does not consider variation in student reasoning. For example, it does not consider whether students recognize a ten-frame as a unit of 10 dots, or whether students count each dot individually.

Figure 9. The rubric for scoring the task shown in Figure 8.

Task: Out of 10. Students count dots on ten-frame and write number:

Points awarded for each sub-part of task, with each part worth 1 point for being correct.

9-10 points: K upper

7-8 points: K messy middle

5-6 points: K lower

3-4 points: PreK messy middle

0-2 points: PreK lower

The most positive feature of the rubric in Figure 9 is that it connects the levels of the rubric with levels of the LP. However, it likely does not do so very carefully, as the LP is concerned with different levels of student understanding about place value, which may or not correspond to the proportion of items that a student has answered correctly. While there is likely a difference between a student who may get 3 of 10 correct versus a student who correctly answers all 10, without some further evidence of the strategies used by those students, it is not clear what evidence this task provides to locate *either* student on the LP.

As a second illustration of a task used at Blue Elementary, Figure 10 shows both a task and its rubric that was used by 1st grade in May of 2015.

Figure 10. A 1st grade task administered in May 2015.

3) Using 9 tens and 3 ones, compose the correct two-digit number.

_____

Explain how you got your answer:

_____

_____

Assessment Rubric

Item	Score	Scoring Guide
3a.		93 (1 point)
3b.		Example: I know that 9 tens in the same as 90 and then I added 3 ones to get 93. Or 9 tens and 3 ones equals 93. (1 point)

In the task shown in Figure 10, there is a greater effort to elicit student reasoning by asking students to explain how they got their answer. That effort is also reflected in the rubric by describing an example of reasoning that students might use, in particular the knowledge that 9 tens is the same as 90. However, the levels of the rubric are not aligned to levels of the LP, a step that would eventually have to be taken in order to use this task as evidence towards a student's location on the LP.

Members of the CADRE team applied the Task Analysis Tool to 17 tasks used in K-1 during the latter half of the 2014-2015 school year. Quantitatively, the analysis suggests work yet to be done to raise assessment quality as part of the SLO process. A summary of the analysis is listed in Table 5.

With a priority given to finding ways to administer kindergarten tasks, and only two SFSs, limited attention was given to improving task quality across the five sessions. Even when SFSs put greater focus on task and rubric quality, teachers sometimes resisted their role as learners in the assessment process. In the fourth session, a teacher stated, "We're not test writers or curriculum writers, we're teachers." This was followed by a discussion of trade-offs between an SLO process that is driven from the top-down versus one where teachers take more responsibility—and do more of the difficult work—to create assessments for their students. In year-end interviews, teachers noted the particular difficulty in writing quality rubrics, with one admitting that she hadn't created rubrics prior to this project. However, all teachers interviewed at the end of the year appreciated the opportunity to work and learn together. When the teacher who was new to writing rubrics was asked if she thought her skills had improved, she focused on the collective effort: "It really was a lot of teamwork with [names other teachers]. So, yeah, that was…that changed a lot. I mean from zero to ten."

Table 5. Summary of Task Analysis at Blue Elementary

Criterion	Spring 2015 %
1. Relevance to the learning progression	
Fully relevant	18
Partially relevant	82
Not relevant	0
2. Level of cognitive demand	
Level 4: Doing mathematics	0
Level 3: Procedures with connections	29
Level 2: Procedures without connections	24
Level 1: Memorization	47
3. Options for expressing understanding	
More than one way to express understanding	18
One way to express understanding	82
4. Rubric	
a. Rubric is reliable	67
b. Rubric is valid	0
c. Rubric is specific	67
5. Fairness	
a. Material is familiar to students for diverse backgrounds	100
b. Items are free of stereotypes	100
c. Students have access to required tools	100
d. Task can be reasonably completed in given amount of time	94
6. Clarity	
Clear and grammatically correct	76
Generally clear; only slight grammatical problems or wordiness	24
Barely comprehensible	0

Instructional practice.

Ideally, assessment should be integral to instruction, and it was hoped that the LPF approach to SLOs would have a positive influence in the classroom. Because this study did not include direct observations of teacher practice, curriculum diaries, access to district- or school-level observations, or the like, this section relies on self-reported data. In the year-end interviews of Blue Elementary teachers, all three were asked if and how the activities of the LPP had an influence on their classroom practice. A first-year teacher at Blue answered

> I noticed that putting [counting by both tens and ones] into our daily routine of counting the days of school really helped them. I know that [other first-year teacher] uses her tens and cubes inside of that chart and I thought that was a great idea of something to add even so students understand the concept more of place value and seeing it with those manipulatives that they use (Year-end interview).

Two teachers expressed an appreciation for how the use of district interim assessment items for the place value LP had brought increased alignment between teaching, district assessments, and the SFSs. For example:

> Yes, I think, you know we did a student focus session around the last task, the second to last task that we had given students, kinda dealing with 10 frames and decomposing numbers, and I think that … it helped us to see

> exactly what students were missing so to really look at, you know, what concepts they understand and what we need to hit back on and I think it also helped us kinda prepare for interims because then it let us see anything that kind of worked within the interim that we were giving in class as a task. It helped us see where their misconceptions were so we could hit on that before they had to take the interim. (Year-end interview)

Another teacher described how project activities have influenced her effort to seek out student reasoning in regular class activity:

> When I was delivering my lessons about [place value], I always tried to include, even maybe if we couldn't do it in writing because of time, at least orally to explain the thinking. And that was something that we did every single time, and then not only for place value, but for any other topic that we were reviewing in the classroom about math. So that was a big change for me, and for [students], I'm sure, because they were pushed to explain more all the time. (Year-end interview)

This same teacher, who worked with many Spanish-speaking students, also described how her attention to student thinking pushed her to work harder to distinguish differences between students struggling with language and those struggling with mathematics.

The third teacher interviewed, however, didn't feel the project created significant shifts in her practice:

> I feel like no, we have been working with those things, but these tasks were just to confirm how we're [students] doing, in which levels were they, so I think it was part of the work, because the kind of tasks that we chose where things that we have been working with and so I feel like those are not totally different things than the things we have been working with during the year. (Year-end interview)

Notably, although this teacher did not identify changes to her own instructional practices, she did identify significant improvements in the SLO process, including the focus on task and rubric quality.

Summary: Brief answers to RQs 2a and 2b.
In this section we summarized apparent effects of the LPP activities concerning task quality and instructional practice. A task analysis reveals teachers had particular difficulties writing valid rubrics, generally because the rubric did not make connections to the learning progression. Several representative examples of tasks used at Blue Elementary reveal some of the progress that has been made and continuing challenges teachers face as they construct tasks and rubrics to assess place value. Self-reported impacts on instructional practice appear broadly positive, however, as teachers feel a sense of progress and integration of the LPP into their teaching. Table 6 shows our answers to RQ 2a and 2b for Blue Elementary, based on these analyses.

Table 6. Brief Answers to RQs 2a and 2b for Blue Elementary

Research question	Brief answer
2.a. Is the LPP associated with an improvement in the quality of assessment tasks?	Teachers, particularly kindergarten teachers, experimented with different forms of assessment that might work best for students unable to read and write for a traditional assessment. Task and rubric quality remained mixed throughout the year.
2.b. Is the LPP associated with changes in teachers' instructional practices?	Teachers reported a number of changes in their practice due to LPP activities, including greater use of manipulatives, more integration of place value tasks into regular classroom activities, and an increased desire to have students explain and justify their thinking.

Teachers' Perceptions of LP-based SLOs.

A driver behind the LPP was to improve the district's approach to measuring student growth, and so it is important that teachers view and use SLOs differently from the SGO process used previously. Teachers were asked for their perspective in the year-end interviews, which helps answer the research question:

3. Do teachers perceive differences between the LP-based SLOs and SGOs?

K-1 teachers at Blue Elementary made comparisons between the LP-based SLOs, their non-LP-based SLO used for literacy, and SGOs. Some of the differences were superficial, such as differences in terminology between the processes, while others indicated potential misunderstandings about growth and accountability. Teachers also suggested some ways to improve SLOs, and expressed value in the opportunities to improve their assessment creation skills during the LP-based SLO work.

A teacher focused on differences in terminology said this in a year-end interview:

> In the past we were using proficient, or unsatisfactory, like in the past with the SGOs. So now I feel like we have different [terminology] for…using upper and messy middle and, so we have like 4, a range of 4 different levels. (Year-end interview)

When asked about ways to improve the project, another teacher focused on struggling with terminology as a learner of English herself:

> It would be really useful for us as second language learners—teachers—because I'm a second language learner… to make the process a little more friendly for us, because one of the things, I mean, I'm not just struggling with the math concepts, the content, but also the language. … We have many teachers, like me or [names other Spanish-first teacher], we need more waiting time, maybe more one-on-one. I don't know. It's just, maybe I'm just in the position that, it's my reality. (Year-end interview)

Teachers' perspectives on using SLOs for teacher accountability also yielded some interesting answers. The first-year teacher was enthusiastic:

> I think if you're looking at growth, and you're really taking the data that teachers are collecting, and really analyzing it and seeing maybe the student didn't make it to a proficient level, or, you know, to the point that [the district wants them at], if that student is showing that they're making growth that's something to be celebrated, and I think they need to take that into [account] when you're talking about teacher effectiveness. (Year-end interview)

The other two teachers gave responses that indicate that they believed teacher effectiveness was tied to teacher performance in SLO-related activities, such as task and rubric creation. This might be a direct result of the way the district phased in SLOs, but it might be a sign that teachers are still not well-informed how student growth will be used as a part of a determination of teacher effectiveness. When asked what could make the process more fair, a teacher replied:

> Having more time as a team to create the tasks, or to make changes to the tasks, and to the rubrics that we have been working with this year…so to get a little bit better and more confident on just creating those rubrics and tasks. (Year-end interview)

Similarly, another teacher said this about using this process to measure her effectiveness as a teacher:

> I think this is a good start but it's a beginning and I don't really feel confident that I can do rubrics and tasks and tests, design them, and you know, all the organization around them. So, my growth and…if somebody comes and scores me about my performance and all the tasks that we did in the process, I will say that I'm in the kind of beginning, middle, messy middle part. [laughs] (Year-end interview)

Perhaps most importantly, several teachers made key distinctions between their district-based literacy SLO and their LPP-based math SLO. One kindergarten teacher appreciated the structure of the district-based literacy SLO and in several sessions expressed some discomfort with the more exploratory approach taken by CADRE and the pilot of the LPP version of SLOs. However, that approach might have helped teachers see the math SLO as more than a compliance or accountability exercise, as explained by a 1st grade teacher at the end of Session 5:

> I think for me, in hindsight, I think of this as less of a How To Do SLOs Training and more of how to develop assessments and how to create an understanding of where your students start and different places where they go. So I feel like, almost, we're using the same terminology for two totally separate things. … So I feel like this process, in my perception, has been more about developing assessments using a trajectory as a focal point. And I'll say in hindsight that we used it less than we should have, to create [assessments]. I feel like in our literacy

> SLO training it's very specific to, 'This is how you do an SLO process.' So in terms of how you enter the data, what you need to put in these blocks, what it needs to look like in the beginning, middle, and end. So this [LPP SLO process] seems to be the functional back-end of it, while the stuff we're doing in literacy seems to be more like the introduction for how you input the data and less about actually how you do any of the tracking [of growth].

Summary: Brief answer to RQ 3.

In this section we analyzed the interviews and Session 5 data to identify the extent to which teachers viewed LP-based SLOs as different from SGOs. Table 7 shows our answer to RQ 3 for Blue Elementary, based on these analyses.

Table 7. Brief answer to RQ 3 for Blue Elementary

Research question	Brief answer
3. Do teachers perceive differences between the LP-based SLOs and SGOs?	Teachers compared the LP-based SLO with the district's regular SLO process they used in literacy. Teachers cited differences in terminology, and some comments indicated teachers felt they were being held accountable for assessment quality rather than student growth. While the structure of the literacy (non LP-based) SLO was appreciated, the LP-based SLO used in math did more to focus teachers on quality assessment and measuring student growth.

4. Green Beachway High School

PARTICIPANTS, LP TOPIC, AND BACKGROUND

Green Beachway was a pilot site in both Year 1 and Year 2. In both years, the entire math department participated in activities of the Learning Progression Project (LPP). In Year 2, there were seven teachers in the math department, five of whom had participated in Year 1, and two who were new to the school and the project.

In Year 1, the teachers chose to focus on algebraic manipulation, and they created an across-grade LP that covered all math courses offered at GB. Figure 11 shows an excerpt of this LP. As shown, the LP had "course-level anchors" (e.g,. levels 3 and 4), along with "micro-levels" in between the course-level anchors. The course-level anchors were narrative statements that described course expectations—including skills, strategies, and understandings—for algebraic manipulation based on Common Core standards. The micro-levels were discrete algebraic skills, ordered according to the sequence in which they were taught at GB. As discussed below, we continued to use this LP in Year 2, but modified it slightly.

Figure 11. An excerpt from the Year 1 across-grade LP at GB.

4. (End of pre-calc): Students understand solving equations as a process of reasoning and are able to explain the reasoning. They 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, and they explain the connections between strategies. This includes: (a) Students know that functions can be treated as objects, and they manipulate equations and expressions involving functions by treating functions as variables; (b) Students understand inverse operations and inverse functions, and use 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 Trig Equations
  - 3.3.2 Solve trig equations
  - 3.3.1 Verify trig equations using the identities
- 3.2 Exponential and Log Equations
  - 3.2.3 Use *e* and natural log
  - 3.2.2 Solve log equations
  - 3.2.1 Solve exponential equations
- 3.1 Quadratic equations
  - 3.1.1 Solve quadratic equations by completing the square

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.

TIMELINE

We conducted nine PD sessions with the teachers throughout the school year that ranged in length from 75 to 105 minutes. The teachers engaged in five student focus sessions: the first was led by CADRE facilitators, and the remaining four were independently conducted by the teachers. Table 8 shows the dates, times, and content of each LPP activity.

PD sessions 1, 2, and 3 were pivotal in terms of setting the focus for the year. In PD 1 we agreed that the focus of Year 2 of the project would be on interpreting student reasoning and using student reasoning to guide instruction. In this session, we agreed

to modify the LP to remove the micro-levels, and simply refer to the levels between the course-level anchors as the "messy middle." We did this because the micro-levels in the LP were defined by teachers based on the sequence in which algebraic skills were taught at GB, rather than though an analysis of how student reasoning develops over time. As such, the LP was more like a teaching progression than a learning progression (For more detail, see *Learning progressions project: Documentation of pilot work and lessons learned in the 2013-2014 school year* available at https://www.colorado.edu/education/node/1797/attachment). In Year 2, we hypothesized that, by leaving the "messy middle" open, we could fill it in over the year as teachers engaged in purposeful analyses of student reasoning. Figure 12 shows an excerpt of the LP after PD 1. As shown, the course-level anchors are the same as in Year 1 (Figure 11), but the micro-level progression of skills has been replaced by an undefined "messy middle."

As shown in Table 8, the remaining PD sessions were largely devoted to writing, interpreting, and improving assessment activities.

Table 8. Summary of LPF activities with GB Teachers

Session	Date	Hours	Content summary
PD 1	8/19/2014	1:15	Agree on Year 2 goals: focus on understanding student reasoning and using student reasoning to guide classroom practice. Agree to remove ordered skills from LP.
	9/2014		Teachers give and score baseline assessment.
PD 2	9/24/2014	1:45	Introduce "realistic mathematics education" (RME) to guide interpretations of student reasoning and design of classroom activities.
PD 3	10/21/2014	1:30	Re-negotiate goals and project activities to focus more on task development and less on classroom activities. Decision to drop RME and return skills (unordered) to LP. Create and modify interim tasks for SFSs.
	10/2014		Teachers give and score interim tasks.
PD 4 / SFS 1	11/6/2014	1.5	Introduction to *Student Focus Sessions* (SFSs). CADRE leads an SFS on Algebra I, using a GB teacher's student work.
SFS 2	11/18/2014	0:30	Teachers conduct SFS on Algebra II. (Phase 1)
PD 5	1/5/2015	2	Discuss SFSs, map baseline scores to LP.
PD 6	2/13/2015	1.5	Introduction to *Task Analysis Tool*, create new interim tasks for SFSs.
	2/2015		Teachers give and score interim tasks.
SFS 3	2/18/2015	0:30	Teachers conduct SFS on Algebra I. (Phase 1)
SFS 4	2/27/2015	0:50	Teachers conduct SFS on Algebra II. (Phase 1)
SFS 5	3/6/2015	0:50	Teachers conduct SFS on Geometry. (Phase 1)
PD 7	4/23/2015	1.5	Discuss SFSs, work on tasks for final assessment.
PD 8	4/28/2015	1.5	Introduction to *Assessment Analysis Tool*, select and finalize tasks for final assessment.
	5/2015		Teachers give and score final assessment.
PD 9	5/19/2015	1.5	Map final scores to LP. Measure growth.

In order to help structure teachers' analyses of student reasoning and to help teachers use student reasoning to guide instruction, CADRE introduced a domain-specific instructional theory, *Realistic Mathematics Education* (Freudenthal, 1983, 1991; Treffers, 1987) in PD 2. In particular, we introduced the notions of "progressive formalization" (Webb, Boswinkel, & Dekker, 2008) and "emergent modeling," (Gravemeijer, 1999) which provide a framework for interpreting student reasoning—including the strategies and representations students use—and a theory for how this reasoning becomes more formal over time.

In PD 3, however, the teachers made it clear that the changes to Year 2 were too far of a departure from Year 1. The teachers asked that we bring back the micro-level sequence of skills to the LP, and asked to keep the PD sessions focused on assessment, rather than on how to interpret student reasoning or use student reasoning to guide

classroom practice. In this session, we re-negotiated our goals for the year. CADRE agreed to focus largely on writing, improving, and interpreting assessment activities. At the same time, teachers agreed to include a focus on student reasoning and classroom practice, but they did not want any further training on either of these. Therefore, we dropped reference to the *Realistic Mathematics Education* framework from subsequent PD sessions. Figure 13 shows an excerpt of the LP after PD 3. As shown, the course-level anchors are unchanged, but the micro-level skills have been returned to the messy middle. The messy middle also includes an open area to capture common ways of reasoning.

Figure 12. An excerpt of the LP after PD 1.

4. (End of pre-calc): Students understand solving equations as a process of reasoning and are able to explain the reasoning. They 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, and they explain the connections between strategies. This includes: (a) Students know that functions can be treated as objects, and they manipulate equations and expressions involving functions by treating functions as variables; (b) Students understand inverse operations and inverse functions, and use 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 (Learning Pre-calc): Undefined "messy middle"

3. (end of Algebra II): 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.

Figure 13. An excerpt from the Year 2 LP

4. (end of Pre-calc): Students understand solving equations as a process of reasoning and are able to explain the reasoning. They 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, and they explain the connections between strategies. This includes: (a) Students know that functions can be treated as objects, and they manipulate equations and expressions involving functions by treating functions as variables; (b) Students understand inverse operations and inverse functions, and use 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 (messy middle, learning Pre-calc):

Students have mastered some combination of these *skills*	Common ways of *reasoning*, including pre-formal representations & strategies, and everyday conceptions
Trig equations • Solve trig equations • Verify trig equations using the identities Exponential and logarithmic equations • Use e and natural log • Solve log equations • Solve exponential equations Quadratic equations • Solve quadratic equations by completing the square	

3. (end of Algebra II): 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.

RESULTS

Student Focus Sessions and Discussions about Student Reasoning

The teachers at GB conducted four independent SFSs. As shown in Table 8, each session was devoted to a single course (e.g., Algebra I), but all of the teachers participated in every session. Thus, the SFSs were opportunities for across-grade collaboration. Each SFS was audio-recorded. In this section, we discuss our analysis of these recordings, along with data from teacher reports in PD sessions and interviews, to address the following research questions:

1a. How do teachers enact Student Focus Sessions?

1b. To what extent can Student Focus Sessions be conducted solely by teachers without outside facilitation?

1c. How do teachers discuss student reasoning in Student Focus Sessions?

The independent SFSs had similarities and differences from our design. The biggest adaptation that teachers made was in terms of the phases of the focus sessions. Recall that SFSs have two phases. Phase I is devoted to scoring tasks to improve task validity and reliability, whereas Phase II is devoted to understanding student reasoning and using student reasoning to guide instruction. In line with the teachers' preference for focusing on tasks rather than student reasoning or classroom practice, GB teachers only conducted Phase I in their independent SFSs. They did not conduct Phase II for any of the independent sessions.

The teachers' enactment of Phase I closely matched our design. Different teachers supplied the student work for each session. In the sessions, teachers first independently scored each piece of student work, and then the lead teacher led a discussion in order to resolve disagreements on scores. The same teacher took the role of lead teacher each time. He generally followed the protocol and kept the discussions focused on the task and the student work. In his year-end interview, he explained that the knowledge that they were being recorded helped to keep the group focused.

In their discussions, teachers modified the tasks and rubrics in an attempt to improve their validity and reliability. Discussions related to validity were generally oriented around modifying task prompts to elicit specific skills from students. For example, in SFS 3, teachers discussed the task shown in Figure 14.

Figure 14. Task discussed in SFS 3.

Josh and Angela are manipulating equations to see which ones are equivalent. Josh says A and B are the same and Angela says A and C are the same. Manipulate the equations using algebraic properties to determine which equations are equivalent. Circle the equivalent equations. Show your work.

A. $y=3x+2$

B. $-3x-y=2$

C. $y=3(x-4)+14$

The teachers who wrote the task described how their intent was for students to manipulate both equations B and C, to determine which was equivalent to A. It was important to these teachers that students manipulate both equations. However, the teachers noticed that students were only manipulating one equation. In SFS 3, the teachers discussed various changes to the task prompt, and agreed to change the prompt to "Show why Josh is wrong and Angela is correct, by manipulating both equations using algebraic properties." In end of year interviews, three of the four teachers noted that modifications to the wording of tasks to elicit something specific from students were a common outcome of SFSs. This sort of modification increases the validity of the task because they help to ensure that the task is eliciting the intended skill from students.

Teachers' discussions about reliability often centered on clarifying vague terms used in rubrics. For example, in SFS 5, teachers discussed the task and rubric shown in Figure 15.

Figure 15. Task discussed in SFS 5 (the rubric did not include a descriptor for level 0)

Task:

Solve for b_1:

$$\frac{(b_1 + b_2)}{2} H = A$$

Rubric:

Description	Score
Completely and correctly solves for b_1.	2
Generally appropriate strategy, however b_1 may not be completely solved for or there may be algebraic mistakes.	1

Notice that the description for score level 1 in the rubric includes the term "generally appropriate strategy." The teachers' discussed the need to clarify the term, "generally appropriate strategy" (in their conversations, the teachers used the term, "good algebra"), as shown in Segment 1 below:

Segment 1 (SFS 5)

Teacher A: What would you define as "good algebra?"

Teacher B: In a multiple step problem, multiple steps… I mean, I don't-

Teacher C: It's impossible to define

Teacher B: Yeah

Teacher A: Right, but like, what mistakes could they make to get a one?

Teacher D: I think the one I described, where they put it all over *h* (referencing an earlier part of the discussion)

Teacher E: So we just need to define it better in the rubric. And show what mistakes are okay. (crosstalk) It IS a common mistake that they divide the whole thing by *h*, not just the 2*a*, but 2*a* minus b_2 over *h*. That's a reasonable mistake that they're gonna make. So I think we take out the words 'good algebra' and say these are the- this is what we're looking for.

The teachers discussed the rubric for approximately 15 minutes. Ultimately, they agreed that the key step involved multiplying by 2 correctly. They agreed to update the rubric to state that the student "must multiply by 2 correctly, there may be small errors in the remainder of the problem" (from the recorder's notes). Note that the updated rubric still lacks specificity: neither "correctly" nor "small errors" is defined. However, the updated rubric is more reliable than the original.

To describe the way that teachers discussed student reasoning in SFSs, it is helpful to contrast it with the way that teachers discussed student reasoning in PD sessions. In PD sessions, teachers' statements about student reasoning were often about students or classes in *general*, exemplified below:

- Statement 1, summarizing a class's response to a particular task: "We tried to at least ask them [students] what it meant, they had no clue" (PD 8)
- Statement 2, summarizing a teacher's experience with students' struggles with negative numbers in general: "It's like negatives are these foreign objects that they [students] are not allowed to have" (PD 3)

These statements are characterized by the use of plural pronouns (e.g., them, they) and/or the use of the generalized present tense (e.g., as in Statement 2). As shown in both statements, the general statements were almost always deficit oriented, in that they were related to what students can't do, or what they don't understand. Finally, these statements were not usually contested or scrutinized by other teachers.

In contrast, conversations related to student reasoning in SFSs were almost always focused on a *specific* student's work on a specific task, and the conversations were composed of multiple turns as teachers worked together to understand the student work. These features are exemplified in Segment 2, as the teachers work together to follow a student's work.

Segment 2 (SFS 4)

1. Teacher A: (reading a student's work on a task that involved the quadratic formula) Negative seven, plus or minus square root of nine, that's correct
2. Teacher B: Right, they're only wrong in their final answer
3. Teacher C: Well, they didn't use 2*a*, they just divided by 2

As shown, the teachers used plural pronouns in these conversations, but here the pronouns function as gender-neutral singular. Using the pronoun this way, the teachers' talk is focused on a specific student's work on a specific problem. These conversations tended to be focused on particular *steps* in a problem that a student did correctly (e.g., Teacher A in turn 1) or incorrectly (e.g., Teacher C in turn 3). Furthermore, because the teachers only discussed students for whom there were discrepancies in scoring, the specifics of a student's work were coordinated with the rubric in an attempt to give the work a consensus score.

This is precisely the sort of conversations that we expected teachers to have during Phase I of Student Focus Sessions—which, recall, are focused on resolving discrepancies in scoring, and using the resolutions to improve tasks and rubrics. However, the focus on specific steps for scoring purposes may obscure *patterns* of reasoning and understandings behind the steps. To understand this distinction, consider the difference between determining which steps in a formula a student did correctly on the one hand, and exploring the different strategies and representations that students use to solve a quadratic equation and making conjectures about what these different ways of solving reveal about student understanding on the other. Whereas the former is the sort of conversation we expect to see in Phase I (and which we did see), the latter is what we expected to see in Phase II of the SFS, in which teachers focus specifically on patterns of student reasoning and make conjectures about student understandings. As discussed above, teachers did not enact Phase II in any of their independent SFSs, and hence, the conversations in SFSs at GB were predominantly focused on reconstructing a student's step-by-step solution. Whether teachers' conversations in Phase II would have been different is a question that we cannot answer.

Summary: Brief answers to RQs 1a, 1b, and 1c
In summary, teachers at GB conducted four independent SFSs, all of which included Phase I only. Table 9 shows our answers to RQs 1a, 1b, and 1c for GB, based on our analysis of these sessions.

Table 9. Brief answers to RQs 1a, 1b, and 1c for Green Beachway.

Research question	Brief answer
1a: How do teachers enact SFSs?	Teachers focused on Phase I. They generally followed the protocol we developed, and used SFSs to modify the tasks that they discussed.
1b: Can SFSs be conducted independently by teachers?	Yes. However, teachers adapted the SFSs to focus more on tasks and less on deeply understanding student reasoning.
1c: How do teachers discuss students reasoning in SFSs?	Teachers predominantly discussed specific student work on specific problems, focusing on steps that students did correctly or incorrectly, and coordinating these with the rubric.

Outcomes of the Project

In this section, we discuss the outcomes of the project with respect to the quality of assessment tasks and instructional practice. We address the following research questions:

2. Outcomes of the Learning Progression Project (LPP)

a) Is the LPP associated with an improvement in the quality of assessment tasks?

b) Is the LPP associated with changes in teachers' instructional practices?

Quality of Assessment Tasks

As discussed above, improving assessment tasks was a major focus of the LP project activities in Year 2. Teachers began the year with a "baseline task bank" of 36 tasks that they had created in Year 1. Teachers selected tasks from this bank to create baseline assessments in the beginning of Year 2. Throughout Year 2, teachers created and modified tasks, culminating in a "final task bank" of 29 tasks, from which they again selected tasks to create final assessments. Overall, teachers used 25 baseline tasks and 27 final tasks in their assessments. To determine the quality of these tasks, we rated them using the Task Analysis Tool described in Section 2. In addition, we examined the final assessments for reliability, coverage, and discrimination.

As discussed earlier, the Task Analysis Tool includes six criteria, listed in the first column of Table 10. The right columns of Table 10 show the ratings of the baseline and final tasks for each criterion.

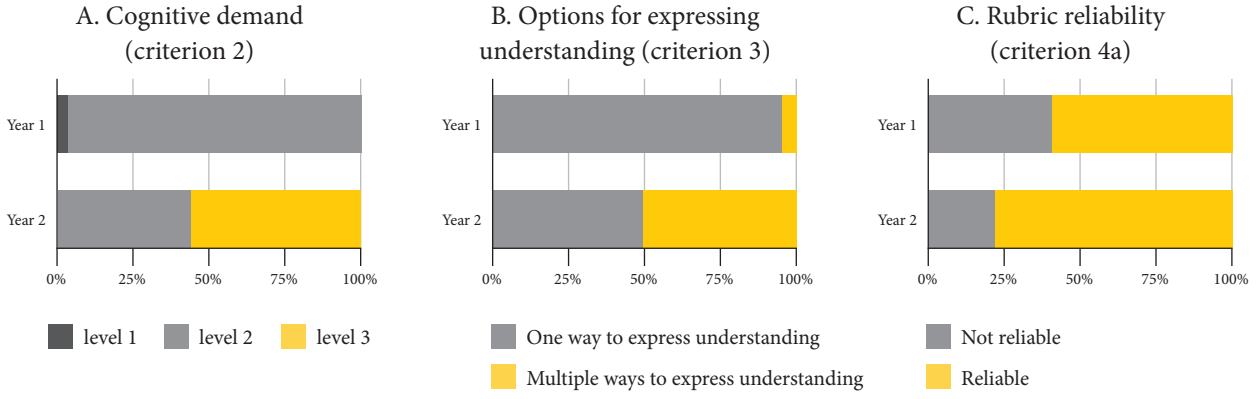
As shown, both the baseline and final tasks were strong with respect to the following criteria: Relevance to the LP (criterion 1), fairness (criterion 5), and clarity (criterion 6). Both sets of tasks were weak in terms of rubric validity (criterion 4b) and rubric specificity (criterion 4c). With respect to validity, we found that the majority of the rubrics did not provide adequate guidance for the full range of possible student responses. With respect to specificity, the majority of rubrics included general descriptors that were not defined in terms of the specific tasks. For example, many rubrics included general terms such as "solved correctly" or "minor errors," without defining "correct" or "minor error" in terms of the problem.

Three criteria, including cognitive demand (criterion 2), options for expressing understanding (criterion 3) and rubric reliability (criterion 4a), showed notable improvement throughout the year, as shown in Figure 16 and elaborated below.

Table 10. Summary of task analysis

Criterion	Baseline %	Final %
1. Relevance to the learning progression		
Fully relevant	100	96
Partially relevant	0	0
Not relevant	0	4
2. Level of cognitive demand		
Level 4: Doing mathematics	0	0
Level 3: Procedures with connections	0	56
Level 2: Procedures without connections	96	44
Level 1: Memorization	4	0
3. Options for expressing understanding		
More than one way to express understanding	4	50
One way to express understanding	96	50
4. Rubric		
a. Rubric is reliable	60	78
b. Rubric is valid	16	22
c. Rubric is specific	32	22
5. Fairness		
a. Material is familiar to students for diverse backgrounds	100	100
b. Items are free of stereotypes	100	100
c. Students have access to required tools	100	100
d. Task can be reasonably completed in given amount of time	100	100
6. Clarity		
Clear and grammatically correct	80	85
Generally clear; only slight grammatical problems or wordiness	20	15
Barely comprehensible	0	0

Figure 16. Tasks improved in terms of cognitive demand, options for expressing understanding, and rubric reliability.



Cognitive demand: Baseline tasks primarily required students to execute known procedures to solve for unknown variables given an algebraic equation. None of the baseline tasks required students to make strong connections between these procedures and the concepts that underlie these procedures (for example, concepts of additive and multiplicative inverses, properties of equality, and conceptions of variables and solutions to equations), even though these conceptual connections are part of the course-level anchors in the LP (see, e.g., levels 3 and 4 in Figure 13). As shown in Panel A of Figure 16, 96% of the baseline tasks were rated as "Level 2: Procedures without connections." The final tasks also required students to execute known algorithms, but over half of the tasks required students to provide explain the connections between the mathematical procedures and underlying concepts (Level 3).

Options for expressing understanding: In line with our agreement to focus on student reasoning in Year 2, CADRE facilitators encouraged teachers to write tasks with multiple ways for students to express understanding. Whereas only one baseline task included multiple ways for students to express their understanding, half of the final tasks included multiple ways to express understanding (Panel B of Figure 16).

The majority of tasks that increased in cognitive demand and options for expressing understanding did so because teachers added a writing prompt. For example, Figure 17 shows a task in which both the baseline and the final version ask students to solve an algebraic equation for which there is a well-known solution procedure. The final version also includes a writing prompt that asks students to explain how their answer relates to original equation. The baseline version was rated as having a cognitive demand at Level 2 (procedures without connections to concepts), and as having only one way to demonstrate understanding. Including the writing prompt in the final version connects the procedure with underlying concepts—increasing cognitive demand from Level 2 to Level 3—and it provides students with multiple ways to express their understanding. The addition of a writing prompt was a common way that tasks changed from baseline to final.

Figure 17. A task that increased in cognitive demand and options for expressing understanding

Baseline task

Solve the equation below for *x*. Leave your answer in fraction form. *Show your work or explain your method.*

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

Final task

Solve the equation below for *x*. Leave your answer in fraction form. *Show your work or explain your method.*

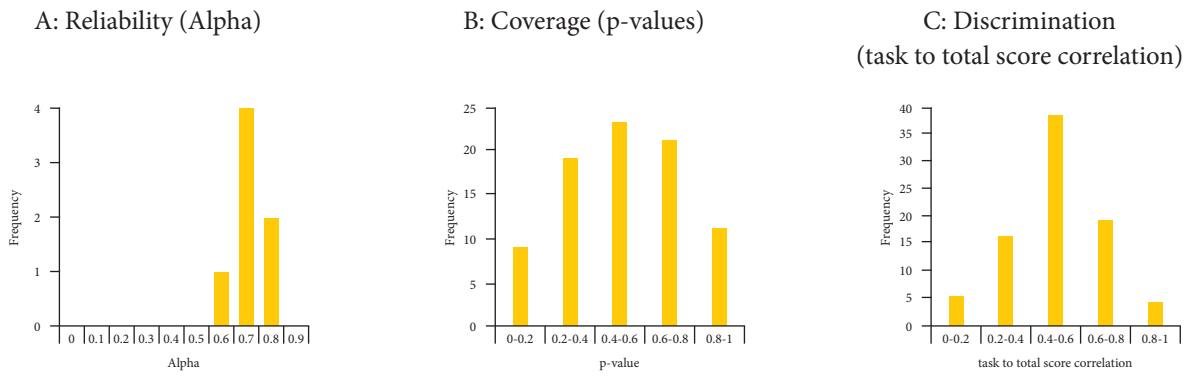
$$\frac{2}{3}x + 5 = 4 - 2(x - 6)$$

What does your solutionmean in terms of the original equation?

Reliability of the rubric: As shown in Panel C of Figure 16, sixty percent of the baseline rubrics were identified as reliable, but 78% of the final rubrics were identified as such. In year-end interviews, teachers credited this improvement to SFSs, and referenced conversations such as the one exemplified in Segment 1 above. Recall that this conversation improved the reliability of the rubric, even though the rubric still lacked some specificity. This explains the somewhat puzzling result that rubrics were largely rated as reliable but not specific.

Teachers assembled the final tasks into seven different assessments: One each for Algebra I/Geometry, Honors Geometry, Pre-calculus, Honors Pre-calculus, and Calculus, and two for Algebra II. To determine the quality of these assessments, we examined the assessments with respect to their reliability, difficulty, and discrimination, as shown in Figure 18, and elaborated below.

Figure 18. Final assessments had high reliability, appropriate coverage, and adequate discrimination.



Reliability: Reliability captures the extent to which the results of a particular test would generalize to another test with a parallel set of items. We measured reliability using Cronbach's alpha (Cronbach, 1951). Alpha ranges from 0 to 1, with higher values indicating higher reliability. For classroom assessments we want to see alpha>0.7. As shown in Panel A of Figure 18, the final assessments had adequate-to-high generalizability, with 6/7 assessments having alpha>0.7.

Difficulty: Assessments should be written so that the majority of the tasks are aligned to the LP levels where the majority of students are located, with a few tasks above this level (i.e., "difficult" tasks) and a few tasks below (i.e., "easy" tasks) to capture students above and below the class norm. We define a task as being "at the level" of a particular student if the student has a 50% chance of getting the tasks correct. We analyzed assessment "coverage" by examining the distribution of p-values. P-values express the mean proportion of total points earned for a particular tasks (for example, a p-value of 0.1 means that, on average, students received 10% of the possible points for that task). Low p-values indicate difficult tasks, high p-values indicate easy tasks, and tasks with a p-value of 0.5 are "on the level" of most students. In general, we want to see a bell-shaped distribution centered on 0.5. This would indicate that the majority of tasks are at the level of most students, while there are still some tasks that are above and below this level. As shown in Panel B of Figure 18, the distribution of p-values is bell-shaped and centered on 0.5, indicating that the final assessments had adequate coverage.

Discrimination: We analyzed the ability of tasks to discriminate among students using the correlation between task scores and the combined scores on all other tasks from the same assessment. The task to total score correlations can range from -1 to 1. In general, items with acceptable levels of discrimination should have correlations above 0.4. As shown in Panel C of Figure 18, 80% of the tasks on the final assessments had item-total correlations above 0.4.

Instructional Practice

We did not conduct classroom observations, so we don't have direct data on teachers' classroom practice. Therefore, in this section we will summarize teachers' self-reports when we asked them about changes to classroom practice in PD sessions, year-end interviews (4 teachers), and year-end survey (4 teachers responded).[3] All four teachers who we interviewed at the end of the year reported changes to classroom practice that they attributed to LP project activities, and all four respondents to the survey reported the same. For example, on the final survey, teachers explained:

> [The process] connected directly to better teaching strategies, has immediate connections to student learning and best practices [and is] helpful and authentic for both the students and the teacher. (Final survey).
>
> [Compared to SGOs,] this is a better way to guide lesson plans to keep covering material throughout the year. (Final survey).

In particular, teachers identified three aspects of their instructional practice that changed based on the LP project activities: (1) increased focus on algebraic manipulation; (2) increased focus on writing, explaining, and justifying; and (3) increased use of student reasoning to guide instruction. Below, we elaborate on the latter two aspects.

[3] Because the surveys were anonymous, we do not know the overlap between the teachers who were interviewed and those who responded to the survey.

Increased focus on writing, explaining, and justifying: As discussed above, the teachers added writing prompts in an attempt to elicit students' understanding of the concepts that underlie algebraic manipulation procedures. As students engaged with these writing prompts, the teachers realized that students had never explicitly been taught how to explain or justify in writing in a math context. Thus, the teachers reported that they incorporated both *activities* and *instruction* on explaining and justifying into their courses. One teacher explained that having students explain their reasoning had a positive effect on students' math skills and understanding:

> I have noticed that kids have definitely shown better understanding when they can explain what they're doing instead of just doing the math. When they start doing their 'creative math' and realize they have to explain it, they realize they're screwing up. (PD 9)

Increased use of student reasoning to guide instruction: Most commonly, teachers explained how they used student responses to specific tasks to guide their instruction. For example, one teacher explained how she adapted her instruction after giving students a task in which they had to explain the meaning of a negative exponent in context:

> We asked the kids to give us a situation first and then explain what 2^{-x} meant in terms of that situation—or, 2^{-3}. And we didn't get a lot of good situations to begin with, but then we thought that's probably because we never asked them to do that before. This was the first time we had asked them. So that influenced what we did after the task, and then we went back and talked about different situations, you know, that were exponential and what they would mean, and then the kids could give us some better situations after that. (Year-end interview)

One teacher, who has been teaching for many years, reported a more general change in his use of student reasoning to guide instruction. He explained that, because of the LP project, he was able to hone his instruction based on what students understand and struggle with, rather than re-teaching an entire concept:

> I started looking more directly at their work again. I mean I did that a long time ago, but what this has helped me do when I look directly at their work I don't teach a whole concept, I say 'okay this is where I notice a lot of kids are stumbling.' So 'you guys know a lot more than you give yourself credit for, so keep doing what you're doing, and that's where you've got to get a little more focused.' (Year-end interview)

Summary: Brief answers to RQs 2a and 2b

In this section we summarized apparent effects of the LP Project activities with respect to task quality and instructional practice. We analyzed task quality using our Task Analysis Tool, and we analyzed the effect on classroom practice based on teacher self-reports. Table 11 shows our answers to RQ 2a and 2b for Green Beachway, based on these analyses.

Table 11. Brief answers to RQs 2a and 2b for Green Beachway

Research question	Brief answer
2a: Quality of assessment tasks	Baseline and final tasks rated highly on relevance to the LP (criterion 1), being fair and unbiased (criterion 5), and having correct grammar (criterion 6). Both sets of tasks were weak in terms of rubric validity (criterion 4b) and rubric specificity (criterion 4c). Tasks improved from baseline to final with respect to options for expressing understanding (criterion 2), cognitive demand (criterion 3), and rubric reliability (criterion 4a). Final assessments had sufficient reliability, coverage, and discrimination.
2b: Effect on instructional practices	Teachers reported three effects on classroom practice: (1) increased focus on algebraic manipulation; (2) increased focus on writing, explaining, and justifying; and (3) increased use of student reasoning to guide instruction.

Teachers' Perceptions of LP-based SLOs

In this section, we discuss teachers' perceptions of LP-based SLOs compared to the district SGO process that teachers had used prior to their participation in the LPP. Drawing on teacher reports in year-end surveys and interviews, we address the following research question:

> *3. To what extent do teachers see the LP-based SLO process as different from the SGO process?*

Teachers reported three features that differentiate LP-based SLOs from SGOs: (1) LP-based SLOs connect more directly to classroom practice, (2) LP-based SLOs are focused on growth rather than status, and (3) LP-based SLOs are more time consuming, but ultimately worthwhile. In the previous section we described the connections to classroom practice that teachers reported. Below, we elaborate on the remaining two ways that teachers perceive the LP-based SLO process to be different from SGOs.

LP-based SLOs are focused on growth rather than status: All of the teachers that completed a final survey mentioned the focus on growth in the LP-based SLO process. For example, one teacher explained the difference between LP-based SLOs and SGOs as follows:

> SLOs are intended to show growth over the year. All students can show growth, no matter where they come in at. (How we measure the growth is a little fuzzy and needs more work.) SGOs have only the end goal in mind. They expect every student to finish at the same level, no matter where they come in. (Year-end survey. Question 19)

SLOs are more time consuming, but ultimately worthwhile: The notion that the LP-based SLO process is time-consuming relative to SGOs came up during PD sessions, year-end interviews, and in the final survey. When we asked teachers to identify the most time consuming parts, they often replied that creating the learning progression was the most time consuming step (this happened in Year 1 of the project). In terms of Year 2 activities, teachers mentioned the need to have dedicated time to conduct Student Focus Sessions, and the need for dedicated time to follow-up on SFSs, for example, planning lessons and writing or modifying tasks based on the conversations in SFSs.

Despite the perceived time commitment, teachers generally agreed that the process was worthwhile. For example, teachers explained:

> The beginning of the SLO process is very labor intensive. However, once you have a learning trajectory, good tasks and good rubrics that align with the trajectory, it is interesting to see the results from the students. (Year-end survey)

> It's a lot of hard work that people won't like at first until they see the benefit. (Year-end interview)

Summary: Brief answer to RQ 3.
In this section we analyzed the year-end survey and interviews to identify the extent to which teachers viewed LP-based SLOs as different from SGOs. Table 12 shows our answer to RQ 3 for Green Beachway, based on these analyses.

Table 12. Brief answer to RQ 3 for Green Beachway

Research question	Brief answer
3: Differences between LP-based SLOs and SGOs	Teachers reported three features that differentiate LP-based SLOs from SGOs: (1) LP-based SLOs connect more directly to classroom practice, (2) LP-based SLOs are focused on growth rather than status, and (3) LP-based SLOs are more time consuming, but ultimately worthwhile.

5. An Emergent Finding: Shifts in Teachers' Conceptions of Learning and Assessment

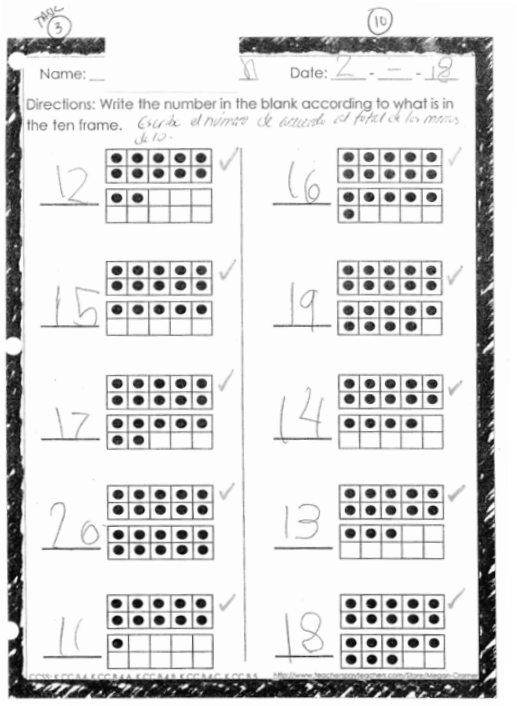
A notable development we began to see in teachers, and something we see as an important outcome of the LPP, is a shift in teachers' conceptions of learning and assessment. During the course of the project, we identified two qualitatively distinct conceptions of learning and assessment, shown in Figure 19.

Figure 19. Two conceptions of learning and assessment.
Note: KSA="knowledge, skills, and abilities"

Conception 1: "Count Up Points"		**Conception 2: Developmental Progression**	
Learning / growth defined			
Operationally, solely as a function of points on tests that are available		With respect to a (possibly implicit) Developmental Progression	
Learning/growth measured using			
Same test pre- and post, growth is difference in scores		All assessment activities are aligned to Developmental Progression, growth is difference in location	
Tasks scored based on			
Amount of correctness		Type of reasoning evident	
Describes students and tasks based on			
%age of points with arbitrary/ historical cutoffs (e.g., 90-80-70)	Amount of mastery of grade-level standards (e.g, prof, pp.) with levels undefined, or defined using arbitrary quantitative cutoffs	Qualitatively described, ordered levels of KSA (could still be prof, pp, etc. but each is defined in terms of KSA)	Qualitatively described ordered levels of qualitative differences in student reasoning
Rubrics			
Generalized rubric, applicable to multiple tasks		Rubrics are specific to tasks and aligned to levels	
Differentiation			
Teach/reteach based on numerical cutoffs		Adapt instruction based on observations of KSA or reasoning.	

We call the first conception a "count up points" count up the points conception. In this conception, tasks are scored based solely on the amount of correctness, regardless of the type of reasoning used. Learning is defined and measured operationally, based on solely on points. For example, as discussed in Section 3, one group of teachers from Blue Elementary used the task shown in Figure 20 below. The task contains 10 different prompts, each with dots arranged in ten-frames. For each prompt, the task is to write the numeral that corresponds to the number of dots in the associated ten-frames. Teachers scored this task by giving students 1 point for each correct numeral, such that the task is scored solely on the basis of how any numerals are correct. Score is not associated with qualitative descriptions of students' knowledge, skills and abilities, nor with patterns of student reasoning.

Figure 20. 1st grade task from Blue Elementary



In contrast, the quote below, from a teacher at Blue describing how she scored a similar task, reveals a "Developmental Progression" Developmental Progression conception of learning and assessment. For this task, students were asked to write the numeral corresponding to "dots in ten-frames" figure, and to explain their reasoning. The student in question wrote the wrong numeral, but reasoned on the basis of 10s and 1s, which was an important concept in the teachers' learning trajectory for place value. The teacher stated the following:

> I struggled with do I give this student two full points for their explanation or 0? I ended up giving him 2 because I think he explained using 10s and 1s. He just explained the wrong number. […] I was like "can he show the concept that I'm asking? That he understands the concept?"

Notice in this quote that the teacher is scoring the task based on the type of reasoning evident, and that this score is correlated to a Developmental Progression. In both examples, the teachers allocated points to tasks. The key question that distinguishes the examples—and the related conceptions of learning and assessment—is, "do the points have meaning with respect to student understanding?" From a Count up the points conception, points define understanding, and learning is defined as changes in points. From a Developmental Progression conception, understandings define points, and learning is defined as changes in understanding.

In Year 1, the Count up the points conception was dominant. We summarized this observation in our final report:

> We found that most teachers thought about growth and assessment in terms of proficiency as the only standard of interest for evaluating students (Year 1 final report, p. 81).

Some of these sentiments continued through Year 2. For example, consider the quote below, from a Green Beachway teacher at the end of Year 2, explaining why she wanted to give the identical pre- and post-tests:

> We wanted to give the same—we understand your guys' argument about not giving the exact same things. But we're going to give some of these same things to see if they grew. (Green Beachway, PD 8)

This quote came at the end of Year 2, as teachers were preparing their final assessments. The teacher's instance that the only way to measure growth is to give "some of the same things" reveals a Count up the points conception, in which learning is defined operationally, solely by points on a test. From this conception, it is essential to use the same test pre- and post-instruction in order to measure learning.

By the end of the project, however, this teacher was in the minority of the teachers at Green Beachway. Most teachers at Green Beachway embraced Developmental Progression conceptions of learning and assessment. For example, the

teacher below from Green Beachway articulates a Developmental Progression conception:

> [The LPF] allow us to evaluate the growth of our students in a way that is authentic to our classes. It also shows growth from a variety of measures, rather than through pre-post test percentages. (Green Beachway, Year-end Survey)

As teachers generally shifted from Count up the points to Developmental Progression conceptions, they began to focus more on student reasoning when scoring tasks:

> I think we've all kinda gotten past the point of right and wrong answers, versus, observing, you know, what– not so much common mistakes, but different thinking kids have through the problem. (Green Beachway Teacher, Year-end interview)

The focus on student reasoning that this teacher describes is a key aspect of a Developmental Progression conception. Notably this teacher is not just talking about himself, but rather is summarizing a shift that he sees across the teachers in the math department at Green Beachway. Teachers applied this sort of thinking to their own teacher-created assessments, as well as district interim assessments:

> You have to think about "what is her understanding?" Kids might score high on an interim, but when I actually look at the work they did on the interim, I might score differently. (Blue Elementary, PD 3)

This teacher at Blue specifically notes that the points that may be awarded based on the district interim assessment aren't the only source of information for her. Instead, she looks at the student's work, and that work may cause her to score the student differently. The student work enables the teacher to understand more about what her students know and can do.

Although there is still some evidence of the Count up the points conception at the two pilot schools, there is also evidence that teachers are shifting away from seeing growth solely as only the difference between two scores. These teachers are interested in student reasoning and what their students understand and are able to do. They see growth in terms of changes in student location along a learning trajectory rather than just the difference between two numbers. A student's growth has qualitative meaning related to the course goals rather than arbitrary percentages completely divorced from content.

6. Recommendations

RECOMMENDATION 1: TAKE UP THE LPF APPROACH TO SLOS

The LPF offers a coherent approach to the development and use of SLOs for the formative assessment of students. Teachers are able to measure growth for each of their students in a straightforward manner that is clearly defined with respect to an underlying LP with anchor levels tied to content standards. Additionally, through SFSs, the LPF approach is directly linked to student learning and instructional practice. SFSs enable teachers to gain a better understanding of student reasoning, and to collaboratively develop specific classroom activities that are responsive to students' locations on the LP. This connection to classroom instruction helps the SLO process to be more than just a compliance-based activity. Instead it is something that teachers are more likely to see as valuable and helpful for their instructional practices.

RECOMMENDATION 2: MAKE TIME AVAILABLE FOR TEACHERS TO WORK COLLABORATIVELY

In order to do this work well, adequate time must be allocated for the LPF process, in particular SFSs, which are the most time-consuming component of the process. SFSs require dedicated and repeated time throughout the year. Each SFS takes two hours, and the process should be completed multiple times over the year. It is unrealistic to expect SFSs to be completed above and beyond other data inquiry requirements without additional time allocation. Teacher at Blue Elementary School expressed that they found value in SFSs, but said they would need the district or their principal to be explicit about which of their current responsibilities or routines they would stop doing in (as part of regularly scheduled "data team" meetings, for example) order to make time for them. Green Beachway teachers made similar comments about the process being labor-intensive but also benefitted from the process. It was not uncommon to hear teachers discuss concern regarding the time commitment for this work at either site on multiple occasions during the year.

RECOMMENDATION 3: PROVIDE TEACHERS WITH PRE-MADE LPS AND ELECTRONIC INTERFACE FOR DATA ENTRY THAT IS LINKED TO THE LP

As evidenced in this report, our team provided tailored tools for teachers during each session. We see the specificity of these tools as key to successful implementation of SLOs. First, district should provide pre-made learning progressions that teachers can choose to adopt. As we found in Year 1, learning progression development is time intensive. Furthermore, as we documented in the Year 1 report, we did not feel that developing the LPs benefitted teachers. We therefore recommend that the district create LPs and provide them to teachers at the start of the year. This allows teachers to focus on task development, student reasoning, and instructional practice.

In addition, we recommend that the district provide tools for teachers to assist teachers in (a) writing tasks aligned to the LP, (b) collecting banks of LP-aligned tasks, and (c) recording observations about students (e.g., scores on tasks) in a way that is aligned to the the LP. Specific directions regarding the process and

standardized tools for completing the process was a common theme of discussion at Blue Elementary in particular. During the final session at Blue Elementary, one teacher stated her concerns as such:

> [The SLO process should] have the same formula or way of filling out some sort of spreadsheet that's kind of the organizer through the whole program. We haven't had a true organizational guide as to what I'm doing and when. (Blue Elementary. Session 5 Content Log)

RECOMMENDATION 4: WORK WITH TEACHERS TO MOVE BEYOND A "COUNT UP POINTS" CONCEPTION OF LEARNING AND ASSESSMENT

Based on the way we saw teachers shift their approach to conceptions of growth, we recommend that implementation of the LPF approach to SLOs intentionally present the Developmental Progression conception of learning and assessment from the beginning of the process. Teachers should be asked to constantly think about and evaluate their students' levels of understanding and relative growth within the year with respect to their changing locations on the LP rather than as a difference of raw points on assessments. Thinking of learning and assessment from a Developmental Progression conception gives more merit to measures of growth and connects growth directly to instructional practice.

7. References

Buckley, K., & Marion, S. (2011). A survey of approaches used to evaluate educators in non-tested grades and subjects. *Dover, NH: National Center for the Improvement of Educational Assessment. Retrieved February*, *21*, 2012.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334.

Freudenthal, H. (1983). *Didactical phenomenology of mathematical structures*. Dordrecht, The Netherlands: Reidel.

Freudenthal, H. (1991). *Revisiting mathematics education: China lectures*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Goe, L., & Holdheide, L. (2011). Measuring Teachers' Contributions to Student Learning Growth for Nontested Grades and Subjects. Research & Policy Brief. *National Comprehensive Center for Teacher Quality*.

Gravemeijer, K. (1999). How emergent models may foster the constitution of formal mathematics. *Mathematical Thinking and Learning*, *1*(2), 155-177.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in quantitative methods for psychology*, *8*(1), 23.

National Council of Teachers of Mathematics. (2009). *Focus in highschool mathematics: Reasoning and sense making*. Reston, VA: NCTM. http://doi.org/10.5951/mathteacher.106.8.0635

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

Stein, M. K., Smith, M. S., Henningsen, M., & Silver, E. A. (2009). *Implementing standards-based mathematics instruction* (2nd ed.). New York: Teachers College Press.

Treffers, A. (1987). *Three dimensions: A model of goal and theory description in mathematics instruction—The Wiskobas Project*. Dordrecht, The Netherlands: Reidel.

Webb, D. C., Boswinkel, N., & Dekker, T. (2008). Beneath the tip of the iceberg: Using representations to support student understanding. *Mathematics Teaching in the Middle School*, *14*(2), 110–113.

Appendix A. Task and Assessment Analysis Tools

TASK ANALYSIS TOOL

Relevance to learning progression

(Check the level that applies. Try to write the task to be as relevant as possible)

	Relevant: All parts of task and rubric are relevant to LP, and the relevance of each part of the task and rubric to the LP can be explicitly articulated.
	Partially relevant: Some parts of the task and rubric are relevant to the LP, but others are not relevant to the LP.
	Not relevant: No part of the task or rubric is relevant to the LP

Cognitive demand[1]

(Check the level that applies. Try to write the task at the top two levels of cognitive demand)

	Doing mathematics: No predictable way to solve, may involve some level of anxiety for the student due to the unpredictable nature of the solution process.
	Procedures with connections: involves some use suggested pathways or algorithmic thinking (perhaps implicitly), but requires students to engage with underlying concepts, for example, using procedures to **deepen connections** to underlying concepts. Explanations involve "why" rather than "what."
	Procedures without connections: Either algorithmic, with little ambiguity about what has to be done or how to do it, OR **little connection** to underlying concepts. Explanations, if present, involve "what" rather than "why."
	Memorization: Explicitly calls for an exact reproduction of previously-seen facts, rules, formulae, or definitions with **no connection** to underlying concepts or meaning

Options for expressing understanding

(Check the level that applies. In a complete assessment, some tasks should have multiple ways to express understanding, and others should have one way to express understanding)

	More than one way to express understanding: Multiple strategies can be used to solve the problem, and the scoring rubric takes student reasoning into account.
	One way to express understanding: The scoring for the problem is dichotomous (right or wrong), student reasoning is either not solicited or not taken into account in the scoring

Rubric quality

(Check all that apply. Try to write the rubric so that all of these can be checked.)

☐ *Reliable:* Either:
- There is some evidence that the rubric can be used reliably by others (e.g., the rubric has been tested and modified in a student focus session) OR
- There is a high probability that the task could be scored reliably by a math teacher at this level.

☐ *Valid:* The rubric is aligned to the task. This means that: (a) the rubric covers everything that students are asked to do (e.g., if the task asks students to "show work" the rubric gives guidance as to how to score the work), and (b) the rubric comprehensively covers the range of possible student responses. If there are multiple possible responses, the rubric gives guidance as to how to score all possible responses (within reason).

☐ *Specific:* All adjectives and general statements (e.g., "shows understanding" or "solves problem correctly") in the rubric are accompanied by specific descriptors related to the problem. For example, if the rubric says "solves problem correctly" the correct answer(s) for the problem is given in the rubric.

[1] Adapted from Stein, M. K., Smith, M. S., Henningsen, M., & Silver, E. A. (2009). *Implementing standards-based mathematics instruction* (2nd ed.). New York: Teachers College Press.

TASK ANALYSIS TOOL

Fairness[2]

(Check all that apply. Try to write the task so that all of these can be checked.)

- ☐ 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.)
- ☐ The task can be reasonably completed under the specified conditions

Clarity[2]

(Check level that applies. Try to write the task so that the highest level applies.)

	The wording in the task and instructions is clear and grammatically correct. The task and instructions are free of wordiness, irrelevant information, unusual words, and ambiguous words.
	The task and instructions are generally clear, but contain slight grammatical or wordiness problems.
	The task and instructions are barely comprehensible due to grammatical errors and wordiness.

[2] Adapted from Diaz-Bilello, E., Thompson, J., & Hess, K. K. (2013). *SLO Assessment Quality Check Tool.* Denver, CO: National Center for the Improvement of Educational Assessment.

ASSESSMENT ANALYSIS TOOLS

Alignment to learning progression

(Try to construct the assessment so that the bulk of the items cover the range of the LP where you expect most students to be, with some items below this range and some items above this range)

Write the range of levels in the LP where you expect most students to be:

☐ — ☐

Write the range of levels in the LP that this assessment covers:

☐ — ☐

Distribution of items:

	Number of items
Below expected range of students	
In expected range of students	
Above expected range of students	

Options for expressing understanding

(Write the number of items on the assessment in each category. Try to construct the assessment so that there are both types of items)

	Number of items
More than one way to express understanding	
One way to express understanding	

Fairness[3]

(Check all that apply. Try to write the task so that all of these can be checked.)

- ☐ Material is familiar to students from identifiable cultural, gender, linguistic, and other groups
- ☐ All tasks (context/texts used) are 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 assessment can be reasonably completed under the specified conditions
- ☐ The rubric or scoring guide is clear for different response modes (oral, written, etc.)
- ☐ 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 assessment format is consistent
- ☐ Formatting and layout is visually clear and uncluttered

[3] Adapted from Diaz-Bilello, E., Thompson, J., & Hess, K. K. (2013). *SLO Assessment Quality Check Tool.* Denver, CO: National Center for the Improvement of Educational Assessment.