ASSESSMENT

Developing purposeful questions and analyzing student reasoning: Two tools

Frederick Peck, University of Montana & Jessica Alzen, Derek Briggs, and Raymond Johnson, University of Colorado

In this paper we introduce two tools to help teachers develop purposeful questions and collaboratively analyze student reasoning. We developed these tools during a two-year research-practice partnership between researchers from the Center for Assessment, Design, Research, and Evaluation (CADRE) at the University of Colorado, and elementary, middle, and high school math teachers in Colorado (due to logistical conflicts, the middle school teachers only participated for one year).

Together, we developed a framework for learning and assessment called the *Learning Progression Framework* (LPF). The framework has its roots in the National Research Council's (2001) report, *Knowing What Students Know*. This report introduced the concept of the "assessment triangle," consisting of three interconnected elements (represented as vertices) that should be the basis for any high quality student assessment: (1) the *cognition* vertex is a model of how knowledge develops, (2) the *observation* vertex is a method of collecting evidence about student cognition (e.g., tasks or other observable activities), and (3) the *interpretation* vertex is a method of making inferences about the observations with respect to the model of cognition.

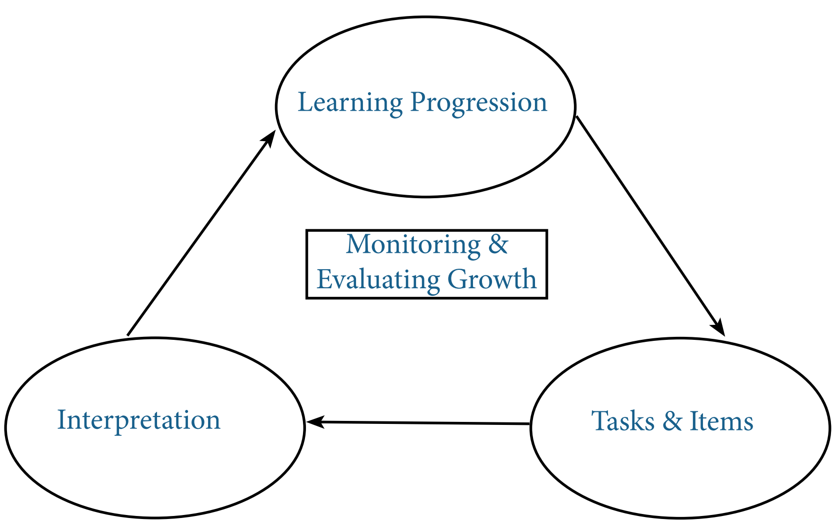


Figure 1. The learning progression framework (LPF).

In the LPF, we operationalized the assessment triangle using *learning progressions* (LPs; Anderson et al., 2012; Clements & Sarama, 2004; Daro, Mosher, & Corcoran, 2011) as our models of cognition, as shown in Figure 1. Learning progressions are "empirically supported hypotheses about the levels or waypoints of thinking, knowledge, and skill in using knowledge, that students are likely to go through as they learn mathematics" (Daro et al., 2011, p. 12). Learning progressions are often created by researchers in mathematics education, after years of careful study of how students learn a particular topic. In our collaboration, teachers combined these researcher-created progressions with the progressions inherent in the Common Core State Standards for Mathematics to create *conjectured learning progressions* in a single domain at each level: place value in elementary school, proportional reasoning in middle school, and algebraic manipulation in high school. As the arrows in Figure 1 make clear, in the LPF we do not consider these learning progressions to be fixed. Rather, they are conjectures about how learning happens, and as such they can be (and were) refined over time based on teachers' observations of student learning.

In this paper, our focus is on the other two pillars of the assessment triangle, so we will not discuss the process of creating or refining an LP further here. A complete discussion is available in the reports on the CADRE website, http://www.colorado.edu/cadre/learning-progressions-project.

For the observation and interpretation pillars, we developed and refined **two tools** during our collaboration: **A task and assessment analysis tool**, which is primarily focused on the observation pillar; and a protocol for collaborative, structured conversations of tasks and student reasoning—focused on both the observation and interpretation pillars—called **student focus sessions**. In this paper we describe these two tools, and discuss how teachers used them to create purposeful questions and engage in collaborative and purposeful analysis of student reasoning.

Task and Assessment Analysis Tool

The **task and assessment analysis tool** describes five *considerations* that emerged as being especially important for developing purposeful questions:

1. *Relevance to the learning progression:* the extent to which a given assessment task and its scoring rubric are likely to provide evidence relevant to the LP.

2. *Options for expressing understanding:* whether the task provides students with only one way to express their understanding (such as with a closed-ended problem like multiple choice or fill-in-the-blank, or tasks that ask for direct applications of routine procedures), or multiple ways to express their understanding (such as with open-ended problems that ask for multiple representations of a solution, or a task that asks for a mathematical procedure with a written justification).

3. *Cognitive demand required:* the extent to which tasks ask students to engage in high-level cognitive processes. There are four levels of cognitive demand (Stein, Grover, & Henningsen, 1996; Stein, Smith, Henningsen, & Silver, 2009):

- Level 1: Tasks that rely primarily on *memorization*.
- Level 2: Tasks that ask students to execute well-known *procedures without connections* to the underlying concepts.
- Level 3: Tasks that ask students to execute *procedures with connections* to underlying concepts.
- Level 4: Tasks that engage students in *doing mathematics*, which includes "framing problems, making conjectures, justifying, [and] explaining" (Stein et al., 1996, p. 464).

4. *Rubric quality:* including:

- Rubric reliability: Indicates whether 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.

5. *Accessibility*, including:

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

Teachers used this tool to analyze existing tasks, identify weaknesses or gaps, and take action to make improvements. For example, in a session in the beginning of the second year, the elementary teachers used an early version of the tool to analyze an assessment provided by the district. At first, many questions on the assessment appeared be aligned to the place value LP, including two tasks that asked students to create an addition expression equal to a given teen number (e.g., _____ + ______ = 19). However, as the teachers analyzed these questions, they found that the tasks were not well aligned to their learning progression and hence would not support related inferences about student knowledge and understanding (consideration 1). As they discussed the task, they realized that the key aspect from a place value perspective was decomposing the teen number into tens and ones, and that the "blank plus blank" task may not give teachers evidence about a student's ability to decompose a teen number in this way. A teacher explained to her colleagues:

> Our [learning progression] is composing and decomposing a teen number, breaking it into ten plus how many ones, whereas these are just blank plus blank. Do you know what I mean?

In this way, the first consideration helped teachers scrutinize tasks for particular mathematical content, and helped teachers make purposeful selections given their content objectives. Ultimately, the teachers found that none of the items on the district assessment were aligned to the place-value LP, so they examined other resources and found tasks that were more targeted.

In high school, teachers had created a bank of assessment tasks during the first year. These tasks were largely procedural, asking students to engage in routine—if often difficult—algebraic manipulations to solve for the value of a variable given an algebraic equation. They provided students with little opportunity to express understanding in more than one way or to make connections to underlying concepts, including properties of equality, properties of operations, and the meaning of solutions to algebraic equations. In the second year, the teachers used the task and assessment analysis tool to improve these questions by providing students with multiple ways to express understanding and by asking students to link the procedures with underlying concepts. For example, the teachers discussed single-variable equations with infinite or no solutions (e.g., $2x + 4 = 8 + 2x$, which has no solutions). They suspected that students often execute a solution procedure correctly, without understanding what the result of the procedure (e.g., $4=8$) means. To assess whether students could link the procedure to the underlying concept, they asked students to solve the equation, $2x + 4 = 8 + 2x$, and then explain the meaning of the solution.

As the teachers created questions that asked students to make connections in writing, they were concerned that analyzing and scoring student responses would be "too subjective." They wanted to analyze student reasoning, but they did not currently have a structure that enabled them to do so collaboratively. To address this, we developed a protocol for collaborative analysis of student reasoning called **student focus sessions.**

Student Focus Sessions

Student focus sessions are conversations that are structured to enable collaborative analysis of student reasoning. They are designed to be conducted by groups of teachers. Below, we outline the main features of student focus sessions. A reference guide written for teachers that describes the process in detail, is available at: https://www.colorado.edu/education/node/1791/attachment.

Student focus sessions have three goals: (1) to learn more about how students are reasoning about tasks, (2) to design instructional moves and classroom activities that are responsive to student reasoning, and (3) to improve the reliability and validity of assessment tasks and rubrics. In a student focus session, teachers examine approximately five examples of student work on two tasks from a common assessment. Although there is no hard-and-fast rule about the quantity of student work, we found five students and two tasks was a sufficient amount of student work to represent a range of diverse responses, while being small enough to enable deep discussions about each student's reasoning.

Student focus sessions have two phases, each lasting about one hour. They can be held in a single two-hour session, or they can be broken into two one-hour sessions in order to fit into the one-hour meeting times that are common in many schools. Again, there is no hard-and-fast rule about the timing, but in our experience this timing worked well.

Phase I

The goal of Phase I is to improve the reliability of task scores by revising tasks and rubrics so as to minimize ambiguity in scoring rules. In this phase, all participants score the student work on common tasks. They then examine any instances where there is substantial disagreement in their scores. They discuss these disagreements, focusing closely on student reasoning, and arrive at a consensus score. They then discuss ways to modify the tasks and/or rubrics so that such scoring discrepancies can be minimized.

Teachers' discussions in this phase often centered on clarifying vague terms used in rubrics. For example, the high school teachers discussed the task and rubric shown in Figure 2.

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

Figure 2. A high-school task.

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

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.

Of particular interest here is the way that the teachers, in searching for consensus, do more than clarify an ambiguous term like "good algebra." In addition, they clarify for themselves what, exactly, they are looking for in the problem. This was a common occurrence in student focus sessions, and at the end of the project many teachers commented on how student focus sessions helped to make tasks more targeted. A high school teacher explained:

You really need to ask yourself, 'what are you trying to understand about their [students'] understanding?' Because you can change a task in the most- in such a small way, and suddenly you're addressing a totally different issue.

Phase II

Phase II has three goals: (1) to improve the validity of the tasks by strengthening the connection between the task and the learning progression, (2) to generate a deep understanding of each student's reasoning, and (3) to develop responsive classroom activities.

First, participants qualitatively analyze students and tasks with respect to the LP. They place students in order with respect to the LP based on a holistic analysis of each student's work, and they place tasks in order of difficulty with respect to the LP based on a holistic analysis of the student reasoning on each task. After coming to a consensus ordering of both students and tasks, they compare this ordering to the ordering inherent in the quantitative scores from Phase 1. If the orderings do not match, this likely indicates that there are important distinctions in student reasoning that are apparent to the teachers, but which are not being captured by the rubric. Participants discuss ways to improve the validity of the task and rubric by making sure that the rubric captures these distinctions.

Participants then focus on understanding each student's reasoning. For each student, they analyze the student's work on both tasks and use this analysis to create a narrative summary of the student. They then use this summary to devise instructional strategies that build on the reasoning and understandings that the student demonstrates in order to help her move along the learning progression. In this way, the instructional strategies gain nuance and go beyond simple decisions to "re-teach or move-on." As one veteran high school teacher explained at the end of the project:

> I started looking more directly at their [students'] 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.'

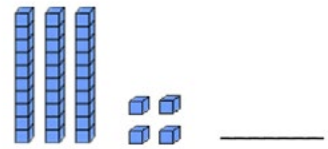
Similarly, an elementary teacher explained how student focus sessions helped to focus her instruction:

> [W]e 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.

As described by the teachers above, the conversations in student focus sessions help prompt teachers to focus on student *reasoning*, as opposed to simply determining whether an answer is correct or incorrect. In our experience, some of the best conversations happened when teachers had to confront two students who both had the correct answer or both had an incorrect answer, but whose reasoning was qualitatively different. For example, the elementary teachers discussed the task shown in Figure 3.

Task:

a) What number am I?



b) What number would I be if there were 7 more ? ____

Student responses:

Randy:
a) 34
b) 8

Salvador:
a) 34
b) 42 *(including drawing 8 cubes onto the figure in part a)*

Figure 3. An elementary school task.

As shown, two students, who we'll call Randy and Salvador, each wrote 34 for part (a). Randy wrote 8 for part (b), while Salvador drew 8 cubes onto the figure in part (a), and wrote 42 for part (b). Both of the students had an incorrect answer for part (b), and using the original rubric—which focused solely on whether the students' answers were correct—both students had the same score. However, the student focus session prompted teachers to look closer at each student's reasoning. Even as some teachers argued that the score on the tasks should be based entirely on correctness, they all agreed that Salvador showed more sophisticated understanding of place value (for example, Salvador correctly grouped 10 ones into one ten, and accurately adjusted digits in both the tens and ones places). Furthermore, even if the teachers disagreed about whether this distinction should be captured in the score, they all agreed that this sort of analysis of student reasoning was important for instructional purposes. During the discussion, one teacher captured the sentiment in the room:

> So I think that, what the student was thinking and us being able to look at these two students, as a teacher and have that direct my instruction, I'm able to say, okay, I know that Salvador has a better understanding of this than Randy. So when I group my students I'm going to group them differently and my instruction is going to look

> different for these two students. But as far as my data tracker goes, I guess I'm not sure how that is going to look when they're both wrong answers.

Over the course of the project these conversations started to have an effect on grading practices. One high school teacher explained the effect of student focus sessions on grading practices in the math department:

> 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.

Similarly, an elementary teacher described how she struggled between scoring a task based on correctness vs. the sophistication of student reasoning. Ultimately, she scored the task based on the student's reasoning:

> 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?'

Student focus sessions are powerful because they give teachers an opportunity to collaboratively engage in analysis of student reasoning. They also support *Principles to Actions* (National Council of Teachers of Mathematics, 2014) mathematical teaching practice: "Elicit and use evidence of student thinking. Effective teaching of mathematics uses evidence of student thinking to assess progress toward mathematical understanding and to adjust instruction continually in ways that support and extend learning." These conversations lead to nuanced understandings of students and more responsive classroom instruction, and they seem to have an effect on teachers' grading practices. However, they require dedicated and repeated time throughout the year. Each session takes two hours, and the process should be completed multiple times over the year. We found that it was unrealistic to expect teachers to conduct these sessions unless they were provided with dedicated time and support to prepare for, conduct, and follow-up with the sessions. In some cases, this support may include having a math coach act as a facilitator for the session.

Conclusion

In this paper, we introduced two tools that can help teachers create more purposeful questions and collaboratively analyze student reasoning. The *task and assessment analysis tool* describes five considerations that help to make tasks more purposeful. *Student focus sessions* allow teams of teachers to have structured conversations about student reasoning, leading to improved assessment tasks, deeper understanding of students, and more-responsive classroom activities. Together, these tools can help teachers create assessments that are grounded in the assessment triangle, and create stronger links between learning and assessment in their classrooms.

Both tools are ready to be used by other teachers, and both are available on the CADRE website: http://www.colorado.edu/cadre/learning-progressions-project.

Acknowledgements

We gratefully acknowledge the teachers who dedicated hundreds of hours to our collaboration. Although they have to remain anonymous, their contributions to the project are manifest throughout this paper.

References

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

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

Daro, P., Mosher, F. A., & Corcoran, T. (2011). *Learning trajectories in mathematics: A foundation for standards, curriculum, assessment, and instruction*. CPRE Research Report #RR-68. Philadelphia: Consortium for Policy Research in Education.

National Council of Teachers of Mathematics. (2014). *Principles to actions: Ensuring mathematical success for all*. Reston, VA: NCTM.

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

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building Student Capacity for Mathematical Thinking and Reasoning: An Analysis of Mathematical Tasks Used in Reform Classrooms. *American Educational Research Journal*, *33*(2), 455–488.

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