Using Assessment Frameworks to Inform the Design of Classroom Assessment[1]

David C. Webb, University of Colorado at Boulder, dcwebb@colorado.edu

Abstract

Goals for mathematics education are often communicated through a variety of policy documents and resources designed to illustrate how students and teachers might engage in mathematical activity. For example, the *Standards for Mathematical Practices* articulated in the *Common Core State Standards for Mathematics* (*CCSS-M;* National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010), provide a synthesis of reasoning goals, processes, and habits of mind that suggest how students should experience mathematics. To address these goals, teachers need to be mindful of how characteristics of mathematical tasks can elicit various aspects of student reasoning. The analysis of goals, tasks, and potential student reasoning is at the heart of classroom assessment design, which is described in the Mathematics Teacher Education Partnership's (MTE-Partnership) *Guiding Principles for Secondary Mathematic Teacher Preparation* (2014) under Guiding Principle 5: Candidates' Knowledge and Use of Educational Practices. Classroom assessment is more typically designed around familiar tasks that appear in instructional materials rather than designed to promote specific types of reasoning. To support productive formative and summative assessment design and use, a framework for what it means to assess student understanding is needed. The framework described and exemplified in this paper has been used by secondary mathematics teachers to analyze mathematical activities, interpret student work, and rethink approaches to scoring and grading summative assessments. This framework has also been used to analyze the reasoning goals addressed in exams used in the undergraduate calculus sequence.

Rationale

Education researchers repeatedly have asserted that to improve student learning, teachers need to give greater attention to their use of classroom assessment (Black & Wiliam, 1998; Wiliam, 2007). To effectively guide student learning in a manner that reflects more formative purposes and goals assessment, instructors must develop greater confidence in their own decision making and understandings of the role of classroom assessment. With respect to both secondary and tertiary mathematics education settings, the real benefit of understanding classroom assessment is not in the scoring and grading of quizzes, midterms, and final exams; rather, the value of developing professional knowledge about classroom assessment is that it informs: (a) the types of questions that are posed as part of instruction, (b) the ways in which student responses might be interpreted, and (c) the subsequent instructional moves based on student responses. The proposed assessment framework in this paper supports the use of tasks that are more open to students' ways of thinking, strategic competence, and student insight.

Guiding Principle 5 (MTE-Partnership, 2014) focuses on candidates' knowledge and use of educational practices, which includes the "knowledge, skills and dispositions needed to implement educational practices" (p. 4). Specific aspects of assessment are articulated in Guideline 5-C:

[1] This paper was supported, in part, by funding from the National Science Foundation under grant DUE-1624628. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

> Assessment and Reflection: The teacher preparation program ensures that teacher candidates can assess the ongoing learning of their students using both formative and summative assessments, and appropriately and ethically use data from those assessments to promote the success of all students as well as to reflect on their own professional growth. (p. 4)

The purpose of this paper, therefore, is to argue for more substantive experiences with classroom assessment in teacher education and outline a practical framework for assessing student understanding that can be applied in these various professional preparation contexts. This framework has been applied in teacher education and professional development for over 20 years in primary, secondary, and tertiary mathematics education, confirming that this framework is, indeed, practical and can be used to inform emerging conceptions of what it means to assess understanding. However, while this framework has been used in these mathematics education contexts (e.g., Shafer & Foster, 1997; Webb, 2009; Webb, Stade, & Grover, 2014), knowledge of this framework among mathematics educators is limited. Even though more recent attention has been given to distinctions between summative and formative assessment, as well as the role of feedback (e.g., Chappuis, Stiggins, Chappuis, & Arter, 2012; Wiliam & Leahy, 2015), teachers lack opportunities to critically examine how tasks, questions, and activities promote students' mathematical reasoning. To what extent do my classroom assessments represent expectations for mathematical reasoning and important mathematical goals? To what degree is the mathematical reasoning elicited in instructional materials, classroom activities, and graded summative assessments aligned? The application of this framework can expand how future mathematics instructors interpret curriculum, enact instructionally embedded assessment, and use students' ideas to inform instruction.

A Framework for Classroom Assessment in Mathematics

To restate the relevant MTE-Partnership guideline in the form of a problem statement: *What knowledge, skills, and dispositions are necessary for productive and equitable use of classroom assessment*? An important starting point is to re-examining one's dispositions toward assessment, through a critical analysis of prior experiences with classroom assessment, as a mathematics student and teacher, and the conceptions of classroom assessment that were reinforced by such experiences. Classroom assessment is often thought of as testing, and the design and use of summative assessment. This is often the more visible and emotionally palpable experience, so it is more memorable. For this reason, programs designed to support mathematics educators' classroom assessment practices must necessarily challenge conceptions of assessment and how these conceptions actually influence their understanding of the discipline they teach. Assessment practices often reify what counts and what is valued. This powerful interaction between goals, as articulated by assessment tasks, and instructional goals was described by Schoenfeld (2007) citing Burkhardt as, " 'What You Test Is What You Get' (WYTIWYG)" (p. 72). As Schoenfeld went on to describe:

> The WYTIWYG principle operates both at the curriculum level and at the individual student level ... and if the test focuses on skills, other aspects of mathematical proficiency tend to be given short shrift. … Similarly, students take tests as models of what they are to know. Thus, assessment shapes what students attend to, and what they learn (p. 72).

Even though Schoenfeld was referring to standardized tests in this excerpt—and the need for greater innovation in format and design—this principle is still valid with respect to the design of classroom tests and other summative assessments, the more regular and common experience for mathematics teachers and students.

Beginning with Dispositions

Given the influence of assessment on practice, dispositions toward mathematics assessment deserve far greater attention in mathematics teacher education than simply acknowledging that somewhat negative experiences or limited dispositions about classroom assessment exist. As noted by Webb (2009):

> Facilitating change in teachers' assessment practice is not so much a resource problem as it is a problem of: (a) creating opportunities for teachers to reconceptualize their instructional goals, (b) re-evaluating the extent to which teachers' assessment practices support those goals, and (c) helping teachers develop a "designers' eye" for selecting, adapting and designing tasks to assess student understanding. (p. 3)

As a starting point, having pre-service (and in-service) mathematics teachers articulate their conceptions of what it means to understand mathematics, and how such understandings might be represented, is an important activity prior to introducing a new framework for assessment. Describing the various ways we are confident in what we know, and the extent to which this knowledge represents some degree of understanding, can help an instructor assess the degree to which various forms of reasoning are recognized. By situating the examination of a new framework with respect to one's prior experiences and conceptions, an instructor will have a better sense of the conceptual terrain and instructional experiences that should be included when discussing less familiar conceptions of classroom assessment.

Another way of thinking about dispositions related to classroom assessment is that we each develop a tacit framework for assessment by drawing upon our experiences as learners and teachers of mathematics. If most of that experience has been devoid of opportunities to engage in reasoning beyond recall to demonstrate success in mathematics, what it means to "get an A" on a test, then why should we expect any divergence from this self-developed, repeatedly reinforced framework? To invite the use of a different assessment framework, it is necessary to critique and reflect upon past practices and experiences to understand the need for a different approach that better exemplifies what it means to do mathematics.

The Dutch Assessment Pyramid

One assessment framework that was designed to support teachers' conceptualization of what it means to assess student understanding was based on an assessment pyramid, first developed in 1995 by researchers at the Freudenthal Institute to support the design of the Dutch National Option for the Third International Mathematics and Science Study (Dekker, 2007) and later adapted by Shafer and Foster (1997) for use in research studies in secondary mathematics classrooms in the United States (Webb & Peck, 2019; see Figure 1). The categories for the assessment pyramid also were used to define the mathematical competencies clusters in the initial version of the mathematics assessment framework for the Programme for International Student Assessment (OECD, 2003, pp. 41–49). The Dutch Assessment Pyramid illustrates the relative distribution for the different levels of thinking that are named: reproduction, connections, and analysis (described in more detail as follows)—the levels of thinking dimension. The other two dimensions are domains of mathematics, which is the range of mathematics content addressed, and the questions posed dimension, which can be thought of as level of difficulty. The dimension of domains of mathematics suggests that this framework for assessment is a domain general approach that does not integrate the notion of a learning trajectory to progressively develop students' understanding of specific domains of mathematics. That would call for a different, perhaps complementary, domain specific framework for how student reasoning in measurement, proportionality, solving equations, etc., develops over time (cf. Webb, 2017).

Before elaborating further about what is meant by connections or analysis, notice how levels of thinking and difficulty are not the same dimension. That is, a task that requires students to make connections between a problem context and possible solution strategies is not necessarily a more difficult task than a recall question. In fact, questions that involve the recall of multiple procedures (e.g., using trigonometry identities to solve a definite

integral) can be more difficult than a problem context involving minimization of area where students decide between using a visual and/or symbolic representation to model the situation. It is also important to note that even though recall tasks still form the majority of assessment questions in this model, recall tasks cannot be the only type of assessment task used with students if the goal is to assess student understanding. Teachers need to move beyond a battery of questions designed to assess student recall and include tasks that allow students to show that they can relate mathematical representations, communicate their reasoning, solve problems, generalize, and mathematize context problems.

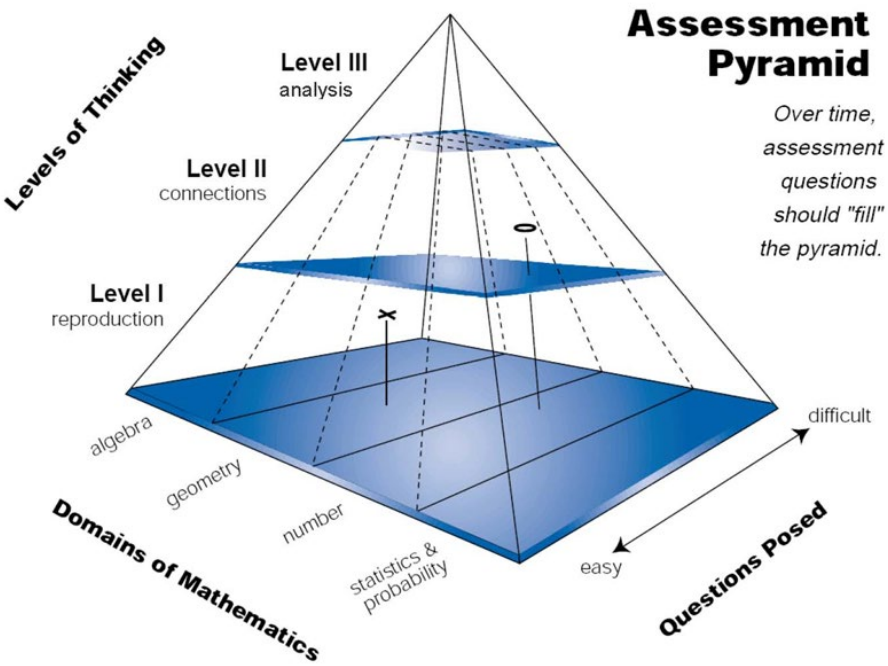


Figure 1. Dutch Assessment Pyramid (from Shafer & Foster, 1997, p. 3).

Using an Assessment Framework to Develop Knowledge and Skills

Assessment frameworks such as this are designed, in part, to serve as a lens for both approximating the potential reasoning elicited from tasks and interpreting students' responses to the tasks. Students reveal their understanding of mathematics when there are opportunities to do so in the tasks and activities that are selected that allow them to demonstrate their understandings. Even though this seems like an obvious statement, consider the extent to which there are opportunities in mathematics classes for students to demonstrate a range of reasoning in the questions asked during instruction or the tasks that are included on quizzes and tests.

To develop a *task-reasoning* lens and a greater sense of the similarities and differences among these three different types of reasoning, after an initial overview of each category and an explicit discussion of how higher level reasoning categories do not necessarily imply the task is more difficult (a different dimension in the pyramid), it is useful to allocate an extended period of time to review and discuss various exemplar tasks that represent each of the categories. In addition to categorizing each task's level of reasoning, which can often be stated as Level I, Level II or Level III, students should be prepared to justify the reasoning level they select. The

process of reviewing mathematics tasks to identify the type of student reasoning that is likely elicited is an ideal activity for pairs or small groups so that different perspectives and opinions can be shared and deliberated. Since the Programme for International Student Achievement (PISA) mathematics framework (Organisation for Economic Co-operation and Development [OECD], 2003) includes many examples of tasks aligned with these three reasoning goals for secondary mathematics, after each description of the reasoning level, released items from the 2003 PISA are discussed to exemplify features of each category.

Reproduction (Level I). The thinking at this level essentially involves reproduction, or recall, of practiced knowledge and skills. This level deals with knowing facts, recalling mathematical objects and properties, performing routine procedures, applying standard algorithms, and operating with statements and expressions containing symbols and formulas in standard form. These tasks are familiar to teachers, as they are found in the most commonly used tasks on standardized assessments and instructional materials. These types of tasks also tend to be the types of tasks teachers have an easier time creating on their own, which is one of the reasons they are so prevalent on summative assessments in mathematics.

Mathematics Example 5
Solve the equation 7x-3 = 13x + 15

Mathematics Example 6
What is the average of 7, 12, 8, 14, 15, 9?

Mathematics Example 7
Write 69% as a fraction.

Mathematics Example 8
Line m is called the circle"s: _____

Mathematics Example 9
1 000 zed is put in a savings account at a bank, with an interest rate of 4%.
How many zed will there be in the account after one year?

Figure 2. Examples of reproduction items from PISA (OECD, 2003, p. 43).

Figure 2 includes several examples of tasks that require recall of mathematical terms and skills. Even though all of these tasks exemplify Level I reasoning in the assessment pyramid, the level of difficulty is not the same. Example 8 involves recall of the definition for diameter (easier), while solving the equation in Example 5 requires recall of procedures, computation with integers, and multiple steps to find a solution. Consider how any of these items could be made easier or more difficult, by adding more numbers, more complex values, or additional steps to the prompt. Regardless of the shift in difficulty, the task will still be assessing the same type of reasoning.

Connections (Level II). The reasoning for this category can involve connections within and between different domains in mathematics. Tasks that elicit this type of reasoning motivate students to use different representations and strategies according to situation and purpose, without being asked to use a particular approach. Tasks that assess connections often involve a realistic context to assess connections made between the context and relevant mathematics. Tasks that reflect characteristics of this category often require students to

make decisions about, and choose from, various strategies and mathematical tools to solve problems (e.g., strategic competence). Therefore, these tasks tend to be more open to a range of representations and solution strategies.

In the seal problem shown in Figure 3, even though this is a selected response item, the type of reasoning this task elicits involves making connections between a problem context and mathematics, perhaps even visualizing or drawing a diagram modeling the situation described in the task. The mathematics involved in this task is not necessarily difficult, but it is a different type of reasoning that involves identifying and extending a cyclical pattern; essentially, this task requires the choice of a solution strategy that exemplifies Level II reasoning.

Mathematics Example 15: SEAL

A seal has to breathe even if it is asleep. Martin observed a seal for one hour. At the start of his observation the seal dived to the bottom of the sea and started to sleep. In 8 minutes it slowly floated to the surface and took a breath.

In 3 minutes it was back at the bottom of the sea again and the whole process started over in a very regular way.

After one hour the seal was:

A. at the bottom
B. on its way up
C. breathing
D. on its way down

Figure 3. Example of connections item from PISA (OECD, 2003, p. 51).

Another example of a task that exemplifies the connections level for the assessment pyramid is shown in Figure 4. The rock concert task requires mathematization of a problem context and the identification of key information that will support a valid estimation of the number of people attending the concert. The use of another selected response task here as an example is intentional. The format of a task does not necessarily determine the type of reasoning that is elicited. Indeed, if student work is limited to selecting one of the choices, A through E, this does not provide sufficient evidence to the teacher of the reasoning the student used.

For a rock concert a rectangular field of size 100 m by 50 m was reserved for the audience. The concert was completely sold out and the field was full with all the fans standing.

Which one of the following is likely to be the best estimate of the total number of people attending the concert?

A. 2 000
B. 5 000
C. 20 000
D. 50 000
E. 100 000

Figure 4. Example of connections item from PISA (OECD, 2003, p. 90).

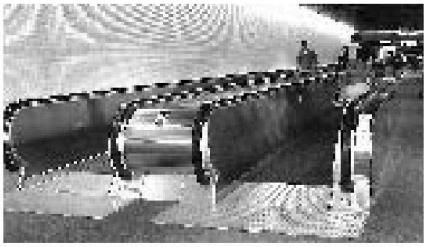
One could imagine adapting the rock concert task to make it a constructed response item without any choices shown. How might that change the difficulty of the task if a set of estimates is not provided? Another option would be to leave the choices for the estimated attendance, but add an additional prompt to explain or justify how the estimate was selected.

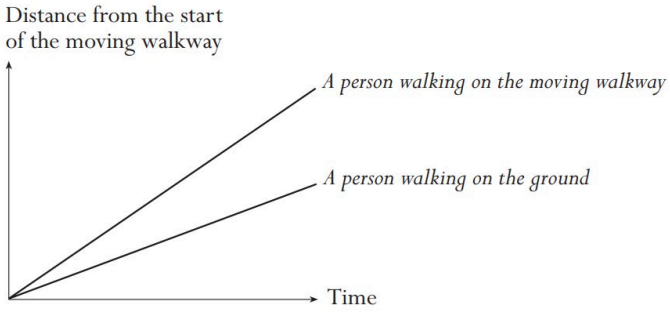
Analysis (Level III). For the analysis category, students are asked to mathematize unfamiliar situations; that is, to recognize and extract the mathematics embedded in the situation and use mathematics to solve the problem. Such tasks require analysis and/or the development of models and strategies. Mathematical arguments, generalization and proof reflect a deeper understanding of mathematical relationships and require some degree of synthesis and analysis to demonstrate this type of thinking. In this category, mathematics itself can be used as a context to promote further generalization of properties, relationships, patterns, and principles. Problems in the analysis category reveal students' abilities to plan solution strategies and implement them in less familiar problem settings that may contain more elements than those in the connections cluster. Having students create tasks can also involve this type of reasoning, as task design involves consideration of important aspects of a mathematics topic and the wording and representation of the task requires some degree of analysis about reasonable parameters, depending on the intended audience.

The moving walkway task (see Figure 5) involves the context of "people movers" often found at an airport. A simple adaptation of this task that might be more familiar to some students could involve an escalator with a stairway next to it. This task requires not only making sense of a mathematical representation of the situation, but reflection about the relative slopes of two linear functions and how a third situation compares to the two that are shown in the graph. This task also motivates students to consider various ways of making sense of graphs that illustrate distance versus time. It is an unfamiliar mathematical task presented for a familiar situation. To add a reasonable line to the graph for the third situation, students need to reflect on the situation and the ways in which the graph illustrates what is happening.

On the right is a photograph of moving walkways.

The following Distance-Time graph shows a comparison between "walking on the moving walkway" and "walking on the ground next to the moving walkway."





Assuming that, in the above graph, the walking pace is about the same for both persons, add a line to the graph that would represent the distance versus time for a person who is standing still on the moving walkway.

Figure 5. Example of a reflection item from PISA (OECD, 2003, p. 91).

Figure 6 is an adaptation of a task that is included as an active learning activity in the first semester calculus course at the University of Colorado Boulder. It is included here to exemplify another aspect of Level III reasoning in the domain of functions and graphing. The first two items are from a sheet of 16 similar tasks in which students use the information provided to sketch a reasonable graph that fits the list of properties. Students need to make sense of the information and how it, collectively, can be translated into a reasonable graph. This aspect of the task exemplifies Level II reasoning.

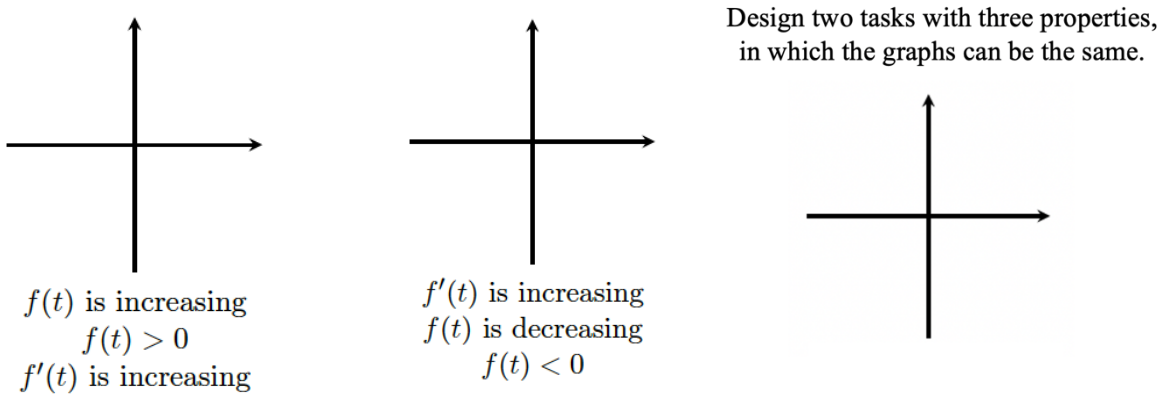


Figure 6. Sketching Snippets (BOALA, 2018).

The third part to the task asks students to design their own problems with a specific constraint: the two problems they design need to result in the same graph. The type of reasoning involved to create a task (or tasks) to fit specific parameters is not necessarily difficult, but it requires knowledge of how information about functions and derivatives influences a graph consistent with the information provided.

Future mathematics teachers should first grasp the meaning of the terms reproduction, connections, and analysis before they reference reasoning categories as Level 1, Level 2, and Level 3. The visual representation of a pyramid suggests that the distribution of tasks across different types of reasoning is not equal. As suggested by Dekker (2007), the ratio among Level I : Level II : Level III type tasks should be approximately 3 : 2 : 1. That is, at least half the tasks should elicit recall of knowledge and procedures, one third should elicit Level II type reasoning (connections), and one sixth should reflect Level III reasoning (analysis). As noted in Figure 1, "over time, assessment questions should fill the pyramid" (Shafer & Foster, 1997, p. 3). This does not mean that each assessment opportunity (e.g., quiz, test, assignment) should reflect this ratio; rather, the distribution of reasoning to be experienced occur over a course of study. It also is likely that some topics in mathematics are more conducive to a higher proportion of tasks that elicit reasoning beyond recall, but the goal should be to represent a range of reasoning for each topic. In a multi-year professional development project that focused on teacher application of the assessment pyramid in the design and use of assessments (Webb, 2011), we found that teachers did move to greater use of tasks that elicited reasoning beyond recall. In this study by Webb (2011), even though the adoption of the 3 : 2 : 1 ratio was not achieved, a significant increase in the percent of problems eliciting reasoning beyond recall was observed:

> Teacher use of tasks that elicit higher order thinking increased from a median of 11.8% in their baseline portfolios to 14.2% in Year 1 and 19.5% after Year 2. Overall, this represents a 65% increase from Y0 to Y2 in the median percentage of higher level tasks used by teachers in response to PD. (p. 8)

Applying the Framework to Support Planning for Formative Assessment

Reflecting on prior experiences with classroom assessment and engaging in critical analysis of status quo assessment practices is necessary to promote student-centered dispositions toward assessment. Knowledge and skills with respect to reasoning goals and how they can be addressed in classroom assessment can be developed through the discussion and analysis of mathematical tasks and activities, and how they reflect the different types of reasoning and difficulty illustrated in the assessment pyramid. The focus on tasks can appear to be an overt focus on resources and processes related to summative assessment. However, these emergent knowledge, skills, and dispositions should be extended to the questions and activities that included as part of instruction, in particular to inform formative assessment practices.

Formative assessment is an integral part of classroom assessment. While the purpose of summative assessment is for reporting and grading artifacts of what has been learned, the purpose of formative assessment is to gather information (through students' responses) to inform instruction. At the heart of formative assessment processes is the provision of feedback so students can gauge how their current understandings compare to instructional goals. Information (i.e., potential feedback) can be provided by instructional materials, classmates, math tutoring centers, instructors, online videos, etc. The information becomes feedback when it is actionable—i.e., when students use the information to make progress toward the instructional goals (Wiliam & Leahy, 2015; Hattie et al., 2016).

A framework for assessment can be used to expand opportunities for formative assessment when the instructor recognizes the potential reasonings that might be elicited from specific questions and tasks. The responses to Level I (recall) tasks are often limited to incorrect or correct answers, unless the task involves recall of a sequence of procedures that might demonstrate some partial recall of skills as in the "solve an equation" and "find the average" tasks shown in Figure 2. The actionable feedback that can be provided to such responses is often limited to error correction, reminders of valid strategies, revisiting the steps of an algorithm, etc. In contrast, Level II and III tasks typically involve a range of solution strategies—they are more open tasks. Students' responses to these types of problems can reveal relationships between different valid approaches, which promotes mathematical connections and relational reasoning. Such tasks also can invite justifications and mathematical argumentation, which can be probed and honed with additional questions, qualifications, counter-examples, and related representations. The type of feedback that is provided to this type of student reasoning can invite subsequent reflection and analysis. For Level II questions, the potential connections that are elicited can lead to feedback that develops those connections. That is, the feedback generated from these types of tasks also can lead to additional enhancement of the same type of reasoning.

Professional development that has focused on the application of the Dutch Assessment Pyramid (Webb et al., 2004; Webb, 2009) begins with the examination of summative assessments and the extent they collectively address the reasoning goals outlined in the framework. The critique and improvement of summative assessments then leads to re-articulating content goals *as well as the reasoning goals* for a unit or course so that assessments have a greater likelihood of eliciting evidence of student understanding. New tasks are selected, adapted, or designed so that summative assessments address more than Level I reasoning and are more closely aligned to the "reasoning beyond recall" goals. The value of this work is that when attention is given to analyzing and improving summative assessments, this aspect of goal-driven assessment design can influence instructional planning, including more informed selection of questions and tasks used in lessons (Her & Webb, 2004). This process of reflecting on, and planning for, possible student responses to instructional tasks supports more productive and focused formative assessment practices.

Summary

This paper offers an approach to achieve Guiding Principle 5, in particular, knowledge, skills, and dispositions related to classroom assessment. Since classroom assessment is often under-addressed in pre-service mathematics education, and classroom assessments are more typically designed around familiar tasks that appear in instructional materials rather than designed to promote specific types of reasoning, it is important for instructors to have a framework that can be used to justify the choices that are made. Even in cases where instructors have limited influence on the design of summative assessments, the approaches outlined here can be applied in instruction for more formative purposes. When mathematics instructors can justify their choice of questions and tasks that they use in lessons, and the potential student reasoning that might be elicited, they are better prepared to adapt instruction in ways that reflect students' interests and ideas.

References

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–74.

Boulder Omaha Active Learning Alliance (July, 2018). *Sketching snippets*. Retrieved from http://math.colorado.edu/activecalc1/index.html

Chappuis, J., Stiggins, R. J., Chappuis, S., & Arter, J. A. (2012). *Classroom assessment for student learning: Doing it right-Using it well* (2nd Ed.). Upper Saddle River, NJ: Pearson Education, Inc.

Dekker, T. (2007). A model for constructing higher level assessments. *Mathematics Teacher, 101*(1), 56–61.

Hattie, J., Fisher, D., Frey, N., Gojak, L. M., Moore, S. D., & Mellman, W. (2016). *Visible learning for mathematics, Grades K–12: What works best to optimize student learning*. Thousand Oaks, CA: Corwin Press.

Her, T., & Webb, D. C. (2004). Retracing a Path to Assessing for Understanding. In T. A. Romberg (Ed.), *Standards-Based Mathematics Assessment in Middle School: Rethinking Classroom Practice* (pp. 200–220). New York, NY: Teachers College Press.

Mathematics Teacher Education Partnership (2014). *Guiding principles for secondary mathematics teacher preparation*. Washington, DC: Association of Public and Land-grant Universities. Retrieved from mtep.info/guidingprinciples

Organization for Economic Co-operation and Development (2003). *The PISA 2003 assessment framework-mathematics, reading, science and problem-solving knowledge and skills*. Paris: Author. Retrieved from https://www.oecd.org/education/school/programmeforinternationalstudentassessmentpisa/33694881.pdf

Shafer, M. C., & Foster, S. (Fall 1997). The changing face of assessment. *Principled Practice in Mathematics and Science Education*, *1*(2), 1–8.

Webb, D. C. (2009). Designing professional development for assessment. *Educational Designer, 1*(2), 1–26. Retrieved from: http://www.educationaldesigner.org/ed/volume1/issue2/article6

Webb, D. C. (April 2011). *Integrating assessment principles and disciplinary content in professional development: Supporting teacher change in classroom assessment*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Webb, D.C. (2017). The Iceberg Model: Rethinking mathematics instruction from a student perspective. In L. West & M. Boston (Eds.), *Annual Perspectives in Mathematics Education*: *Reflective and Collaborative Processes to Improve Mathematics Teaching*. Reston, VA: National Council of Teachers of Mathematics.

Webb, D. C., & Peck, F. A. (2019). From tinkering to practice—The role of teachers in the application of realistic mathematics education principles in the United States. In M. van den Heuvel-Panhuizen (Ed.),

International Reflections on the Netherlands Didactics of Mathematics (pp. 21–40). New York, NY: Springer Nature.

Webb, D. C., Romberg, T. A., Dekker, T., de Lange, J., & Abels, M. (2004). Classroom Assessment as a Basis for Teacher Change. In T. A. Romberg (Ed.), *Standards-Based Mathematics Assessment in Middle School: Rethinking Classroom Practice* (pp. 223–235). New York: Teachers College Press.

Webb, D. C., Stade, E., & Grover, R. (2014). Rousing students' minds in post secondary mathematics: The undergraduate learning assistant model. *Journal of Mathematics Education at Teachers College, 5*(2), 39–47.

Wiliam, D. (2007). Keeping learning on track: Classroom assessment and the regulation of learning. In F. K. Lester Jr. (Ed.), *Second handbook of mathematics teaching and learning* (pp. 1053–1098). Greenwich, CT: Information Age Publishing.

Wiliam, D., & Leahy, S. (2015). *Embedding formative assessment: Practical techniques for K–12 classrooms*. West Palm Beach, FL: Learning Sciences International.