

Secondary Mathematics Preservice Teachers' Assessment Perspectives and Practices: An Evolutionary Portrait¹

Matt Wallace

University of California, Davis

Tobin White

University of California, Davis

Received: 20 June, 2013/ Revised: 29, January 2014/ Accepted: 24 March, 2014

© Mathematics Education Research Group of Australasia, Inc.

Over the past two decades, mathematics education reformers have increasingly called for new approaches to assessment. This call urges mathematics teachers to merge their assessment with pedagogy, to employ assessments that engage students in the process of "doing mathematics," and to exploit assessment to support learning. This paper presents a research study of how six secondary mathematics preservice teachers learned to use such reform-based assessment practices while enrolled in one of three reform-minded teacher education programs. Data includes three interviews that concentrate on each preservice teacher's views on assessment. In addition, a series of classroom observations focus on the ways each preservice teacher assessed student learning. Analysis of the data revealed three distinct stages in the evolution of preservice teachers' assessment perspectives and practices – from traditional to more reform-based. Findings indicate that preservice teachers first focus on how to assess before considering other assessment functions such as what to assess and how to use assessment.

Keywords: classroom assessment • preservice teachers • teacher learning • assessment • mathematics reform

Introduction

Over the past two decades, mathematics education reformers have increasingly called for new approaches to assessment. Classroom assessment literature (e.g., Gardner, 2006), reform reports (e.g., Lund, 2008), and key reform documents (e.g., National Council of Teachers of Mathematics [NCTM], 1989; 1991; 1995; 2000) have urged mathematics teachers to transform their assessment practices from activities independent of instruction to exercises intertwined with classroom lessons, from practicing algorithms to engaging in problem solving, and from evaluating achievement to promoting student learning. In contrast to traditional practices, such reform-based assessment approaches are more appropriately aligned with current constructivist models of teaching and learning (Shepard, 2000), more relevant to present-day content standards that emphasise conceptual understanding (Brahier, 2001), and more consistent with the notion that *all* students can learn mathematics (Stenmark, 1991). Furthermore, when implemented effectively, reform-based assessments substantially improve student achievement (Black & Wiliam, 1998) and can greatly enrich student understanding (Shepard, 2001).

Despite such virtues and values, much evidence suggests that we have yet to see deep or widespread change in assessment within mathematics classrooms (e.g., Ohlsen, 2007; Senk, Beckmann, & Thompson, 1997; Watt, 2005). This is not to suggest that in-service mathematics teachers' are incapable of employing reform-based assessment practices (see, for example, Suurtamm, Koch, & Arden, 2010). Rather, as many reform reports and research studies maintain

¹ This research was supported in part by the Spencer Foundation.

(e.g., Brown, Cooney, & Jones, 1990; Popham, 2009), efforts to restructure assessment are inextricably linked to the preparation of new teachers.

Recently, there has been an increased effort to prepare preservice teachers (PSTs) in the ways of classroom assessment (e.g., Frykholm, 1999a; Shepard et al., 2005; Stiggins, 1999). Despite these efforts, research suggests that many PSTs still espouse and employ traditional assessment methods. For example, Volante & Fazio (2007) found that throughout their preparation program, primary/junior PSTs lacked confidence when assessing student learning and favoured using assessment for evaluative purposes. Concerning mathematics PSTs, Frykholm (1999b) found that many of the PSTs in a reform-oriented secondary mathematics teacher education program came to understand, value, and subscribe to central reform tenets. Fewer of those PSTs, however, believed such a model was feasible or included reform-based assessment practices in their daily instruction. Likewise, Pratt (2002) found that while many PSTs' assessments were aligned with their instructional plans, these plans did not promote problem solving, reasoning, or any other reform-based mathematical practices.

In order to better understand and perhaps improve on these reported *outcomes* of teacher preparation, our research focuses on the *processes* by which PSTs' assessment practices mature as they progress through teacher education. To this end, we conducted a study of the ways PSTs learn about, make sense of, and develop their own ideas about and approaches to assessment in a year-long reform-minded preparation environment. The central research question guiding this study was: How do the assessment practices of secondary mathematics PSTs evolve over the course of year-long teacher education programs? To answer this question, we draw upon extensive interviews, observations and other data collected over the course of a year in order to develop case studies of six secondary mathematics PSTs in three different reform-minded teacher education programs. We believe these nuanced accounts of PSTs' developing assessment practices represent important contributions to the knowledge base on preservice mathematics teacher learning.

Developing Assessment Perspectives and Practices

For this research, we use the general term 'assessment' to refer to *all* activities that provide evidence concerning what an individual student knows and is capable of doing in mathematics *before, during, and after* instruction (NCTM, 1995; Wiliam & Black, 1998). We use the term 'assessment perspectives' to refer to specific ways individual teachers describe their ideas about assessment. Finally, borrowing from Scribner & Cole's (1981) definition of *practice*, we use the term 'assessment practice' to refer to all recurrent and goal-directed events with the intended function of assessing a student's knowledge and capabilities in mathematics. Importantly, while perspectives and practices are presented as two separate aspects of assessment, the two can be tightly linked.

Dimensions of Mathematics Assessment Perspectives

There are a number of possible perspectives on the concept of assessment. Webb (1992), for example, noted that assessment could be considered a tool that provides evidence and feedback on what students know, a means of communicating what is valued and important to know, a source of information for policymakers, or a measurement of program-wide effectiveness. Frequently, however, assessment is presented as a dichotomy between summative and formative functions. Summative assessment, which is closely tied to traditional notions of assessing, concerns tasks designed to measure achievement at the end of a period of instruction, such as with tests. From this perspective, the purpose of assessment is to provide objective and valid information about student achievement. On the other hand, formative assessment, which is more

closely associated with current reform-based models of assessment, involves using assessment-elicited evidence to inform both teacher and student decisions for learning (Black & Wiliam, 1998; NCTM, 2007). From this perspective, assessment is viewed as a process that, through engaging in and feedback from the task, has a direct and positive influence on teaching and learning.

Figure 1 displays a spectrum of mathematics assessment practices. At one end are the practices that most typify traditional assessments and are rooted in behaviourist learning theory; at the other end are reform-based assessment practices which are based on constructivist and socio-cultural learning theories (Shepard, 2000). As the figure illustrates, both traditional and reform-based assessments differ along a number of dimensions. These dimensions draw on Harlen's (1998) description of formative and summative assessment practices, which identifies assessment in terms of its purposes, uses, timing, and feedback. In addition, given shifts in perspectives on mathematics teaching and learning (e.g., Cockcroft, 1982; NCTM, 1989), dimensions concerning assessment methods, tasks, and cognitive demands are also included. We elaborate on each of these dimensions below.

	Traditional ←—————→ Reform-based			
Purpose	To record achievement of individuals (evaluative)	To monitor progress or learning (summative)	To inform teaching (diagnostic)	To inform learning (formative)
Use	Report to student, parent, teachers, etc.	Feedback into teaching plans	For students and teacher	
Timing	Separate task	Intermission from classwork	Normal part of classwork	
Feedback	Criterion referenced	Student and criterion referenced	Student referenced	
Task	Closed	Open-middle	Open-ended	
Cognitive Demands	Procedures without connections	Procedures with connections	Doing mathematics	
Methods	Limited methods	Several methods	Variety of methods	

Figure 1. Dimensions of mathematics assessment practices.

Purpose. Just as with assessment perspectives, the purpose of assessment is frequently presented as a dichotomy between formative and summative functions. Associated with the reform-based approach, the purpose of formative assessment is to inform or advance learning through a process of collecting, interpreting, and using assessment-elicited evidence (Black & Wiliam, 1998; NCTM, 2007). Importantly, the student is an active participant throughout the formative assessment process (Harlen, 2006). By contrast, the purpose of summative assessment, which is closely tied to traditional notions of assessing, is to measure or obtain information about students' past or current achievement (O'Connor, 2002).

The Task Group on Assessment and Testing (1988) proposed two purposes of assessment in addition to the formative and summative functions mentioned above—diagnostic and evaluative. Diagnostic assessment, the group declares, "[is a] more focused use of formative assessment activities" (p. 25). That is to say, like formative assessment, the purpose of diagnostic assessments are to advance learning; however, diagnostic assessment involves identifying

student difficulties to inform instructional decisions, so students are passive recipients of the teacher's interpretation of their work. The purpose of evaluative assessment is to report on the performance of students, teachers, and schools for use in educational decisions such as admission, promotion, or certification. This information is typically based on summative assessments; evaluative activities, however, serve the purpose of classifying or ranking.

Use. Since the purpose of traditional mathematics assessment is evaluative and summative, these tasks are used to communicate or report on academic achievement (e.g., grades or report cards). An alternative to using assessment-based evidence to report on learning involves using student work to make decisions about what topics require further discussion and work. The central use of reform-based assessment, however, is to enable both teacher and student to identify next steps in learning and how to take these next steps. To that effect, teachers use assessment-elicited evidence of student learning to inform instruction, and students use assessment information to adjust the ways they are learning a particular concept (Wiliam, 2007).

Timing. Traditionally, mathematics students are assessed at the end of a period of instruction (Even, 2005). Here, assessment is treated as an official event that is separate from instruction or other learning activities. Reform-based tasks, on the other hand, are integrated with instruction such that "assessment...becomes a routine part of ongoing classroom activity rather than an interruption" (NCTM, 1995, p.13). Thus, reform-based assessment activities are opportunities for students to learn, as well as demonstrate what they know and are capable of doing. Between these two ends are assessments that do not necessarily occur during instruction nor at the end, but rather as an intermission. A teacher might pause in the middle of a lesson, for example, to monitor student progress before continuing her instruction.

Feedback. Feedback concerns information regarding student performance or understanding. Traditional assessment activities are judged against specific learning criteria, and the feedback, often in the form of marks or grades, reflects student performance towards meeting these criteria (Harlen, 2006; Shepard, 2001). With respect to reform-based assessments, teachers make qualitative judgments of individual student work, progress, or both. Feedback is in the form of comments and is directed towards giving students appropriately challenging guidance or support in relation to the learning goal, thus establishing conditions for on-going learning (Sadler, 1989). Less traditional yet not entirely reform-based feedback does not focus on the student but on correcting their mistakes (i.e., corrective feedback) and is not often generalisable to other tasks (Hattie & Timperley, 2007).

Tasks. Traditional assessments involve closed tasks that contain one correct answer with only one correct process for arriving at that answer (NCTM, 1995; Shepard, 2001). Moving away from the more traditional approach are open-middle tasks, which still require one correct answer but allow for students to employ one of several different solution strategies (Brahier, 2001). Reform-based assessments contain open-ended tasks that have many correct answers and many solution strategies (NCTM, 1999).

Cognitive demands. With traditional assessments, the emphasis is on recall, memorisation, and other low-inference tasks (Even, 2005; Shepard, 2001). During such assessments, students engage in basic procedural processes such as applying formulas or algorithms and are rarely required to extend their work to broader conceptual ideas or themes. Stein, Grover, and Henningsen (1996) have classified the cognitive demands of such tasks as "procedures without connections." Stein and colleagues classified procedurally-based tasks that also encourage students to make sense of their work within broader mathematical contexts as "procedures with connections." Genuine reform-based assessments, however, promote problem solving, analysing and interpreting, exploring patterns, making connections, conjecturing, reasoning and justifying, and

communicating mathematical ideas (NCTM, 1989, 1995). In essence, students are engaged in the process of "doing mathematics" just as mathematicians do (Stein et al., 1996).

Methods. Because the prevailing tool in traditional mathematics assessment is paper-and-pencil tasks consisting of short, close-ended questions (Even, 2005), students are assessed on their knowledge and skills by way of only a few methods (e.g., tests and quizzes). Reform-based assessments, on the other hand, focus on mathematical concepts, reflect the important role of problem solving, and include the full range of mathematical activity. Accordingly, reform-based assessment practices must include a combination of different methods to provide evidence of what a student knows and is capable of doing in mathematics (NCTM, 1989, 1995; Wiliam, 2007). So, in addition to paper-and-pencil tasks, reform-based assessments include presentations, portfolios, projects, observations, and so on.

Methods

Alex, Ben, Erika, Jake, Karen, and Mary (pseudonyms), are six secondary mathematics PSTs involved in this research. As shown in Table 1, the education and work experiences varied among these six PSTs. Importantly, Karen's work experience involved teaching English to non-native speakers, a job which first involved a month long teacher-training program.

Table 1 displays the course or courses that each PST taught during their field practicum along with the students' ages in each course.

Table 1
Background Information of PSTs

PST	Program	Undergraduate degree	Work Experience	Field Practicum Course(s) (Student Age)
Alex	A	Math	None	Algebra I (14) & Geometry (15-17)
Jake	A	Economics	≤ 5 years	Algebra I (14) & Geometry (15-17)
Karen	A	Math	≤ 5 years*	Pre-Algebra (15-17) & Algebra II (14-17)
Mary	A	Math	≤ 5 years	Algebra I (15-17) & Geometry (12-14)
Erika	B	Math	None	Algebra I (14-17) & Algebra II (14-17)
Ben	C	Engineering	> 15 years	Pre-Algebra (12)

Notes. Program B and C are shaded in grey to reflect a partnership between programs, and shared methods courses.

* Karen's work experience involved teaching.

These six PSTs were enrolled in one of three different reform-minded teacher education programs. Alex, Jake, Karen, and Mary all attended "Program A," a teacher education program at a large research-oriented university situated in a largely rural area in California; and as part of this program, all four were engaged in non-discipline specific courses focused on topics such as

educating diverse student bodies, as well as mathematics specific teaching seminars and methods courses. The year-long math teaching seminar and methods courses, which placed a particular emphasis on using technology to make mathematics accessible to all students, met weekly for two and four hours, respectively.

Erika and Ben attended Programs B and C, respectively. Program B was part of a large liberal arts college in a major metropolitan area in California. In the same metropolitan area was Program C, a small alternative teacher education program designed for working professionals. Both programs B and C required courses similar in nature to Program A; however, the mathematics specific courses in these programs emphasised problem-based teaching and learning, and the year-long methods course met weekly for three hours. Program C was affiliated with program B, and students in both programs often took courses together, as was the case with Ben and Erika and their methods courses.

None of the programs offered specific assessment courses. Instead, the mathematics teaching methods courses in each program included class meetings, discussions, and assignments concerning assessment. These opportunities are described in more detail in the Findings section below.

While enrolled in their respective programs, all six PSTs apprenticed with one or two experienced mathematics teachers as part of a field practicum. During the field practicum, each PST first took on a nonparticipant-observer role, and then progressively assumed greater responsibility until taking over all classroom duties.

Data Sources

A primary source of data for this research involved three 45 to 100 minute semi-structured interviews with each PST. These three interviews took place at equal intervals during each PST's respective program. The first interview (1st month into the program) involved general background information questions, and questions designed to elicit data on each PST's perspectives on assessment. The second interview was conducted near the middle (month 5 or 6) of each program. The third interview was conducted with each PST during the ninth or final month of his or her program. The intent of the second and third interviews was to track any changes in PSTs' perspectives over time. Thus, the interview questions varied across PSTs, as questions were asked in the light of previous responses.

During each interview, PSTs were also asked to sketch out an assessment plan for one unit of a hypothetical algebra class, which included a description of all assessment tasks for the unit, the functions such tasks would serve, and the types of information such tasks would elicit from students. Next, each PST was asked to choose an algebra topic (not necessarily related to the unit), outline a lesson, and elaborate on his or her assessment activities and the purpose of each assessment activity.

Another primary source of data concerned classroom observations with each PST during the field practicum. Each PST was observed teaching two discrete units of instruction, one occurring in the beginning of the field practicum and one towards the end. Several class meetings were observed over the course of each unit, totalling anywhere from 270 to 320 instructional minutes depending on the unit topic(s), school schedule, and so forth. Field notes were recorded during each observation. Included in the field notes were all perceivable moments and opportunities for assessment, including teacher-led instruction, classroom discussions, class and student activities, and teacher-student interactions. Along with the field notes, all available classroom documents were collected.

Immediately following one observation in the beginning, middle, and end of each unit, 15- to 45-minute post-observation interviews were conducted with each PST. During these interviews, PSTs were asked to describe the ways that they assessed their students, discuss what

they were thinking and doing with regard to each specific assessment, and reflect on the effectiveness and information gleaned from each practice. Additionally, PSTs were asked to explain why other moments witnessed in class were thought not to be assessment.

For insight into the assessment concepts and practices that each PST was exposed to, non-participant observations were conducted during roughly a quarter of all teaching methods course meetings. These observations occurred during class sessions intended for assessment-related concepts as identified through course syllabi or from notification by the methods instructor. Field notes were recorded during each observation to document the overall structure, discussions, and activities of the class, with a particular eye towards assessment-related matters.

To supplement the interview and classroom observation data, a number of assessment-related methods course assignments were collected, including lesson plans, journal reflections, term papers, and, most notably, the *Performance Assessment for California Teachers* (PACT). Briefly, the PACT, which is designed to determine whether a credential candidate has met a minimum performance standard considered necessary for teaching, includes an assessment component that requires PSTs to collect and analyse evidence of student learning.

Data Analysis

The analysis began with separate coding of the interview and observation data for each individual PST. Data were separated into one of three categories based on when the data were collected—beginning (0–3 months), middle (4–6 months), and end (7–9 months). Within each time interval, the data were then further subdivided into assessment perspective and practice categories. Importantly, these broad categories were not mutually exclusive, as some data reflected both a perspective and illustrated a practice.

Within each of these perspective and practice categories, the data were then carefully coded. The coding categories for assessment perspectives included descriptions of assessment, as well as ideas or examples of assessment methods, purposes, and content. With assessment practices, PSTs' assessment tasks or methods (e.g., test, homework, etc.) were individually coded using the specific dimensions of assessment (previously shown in Figure 1). In some cases, a particular assessment received two codes for the same category. A quiz that contained both open-ended and closed questions, for example, would receive both corresponding codes.

The results of this coding process were then examined to look for changes or similarities in each PST's assessment perspectives and practices over time. Afterwards, the content and general themes of the coded data were summarised into profiles of each PST's assessment perspectives and practices at the beginning, middle and end of their credential program year. The findings from this analysis are presented in the following sections.

Findings

The findings of this study revealed three different stages in the evolution of PSTs' assessment perspectives and practices—test-oriented, task-oriented, and tool-oriented. These three different stages are described below. Importantly, the timelines along which PSTs progressed were not uniform, but nevertheless followed a predictable pattern.

Test-oriented perspectives and practices. The test-oriented assessment perspective primarily concerned the notion that assessment was simply another name for a "test" or a "grade." The leftmost assessment perspective column of Table 2 contains excerpted quotes from each PST to help illustrate this common test-oriented view of assessment. For example, when asked to share her views on assessment, Erika replied: "I think that assessment means evaluation, which I would

say is grading." In addition, to ensure that each individual student's work reflected their own understanding, assessments were to be an independent task in which students must write out their entire solution strategy. This also ensured that assessment grades were both accurate and dependable.

Table 2 shows samples of PST responses. To help illustrate the timelines along which each preservice teacher progressed through these stages, the three-month time intervals used in our analysis are placed in parenthesis under each excerpted quote in Table 2.

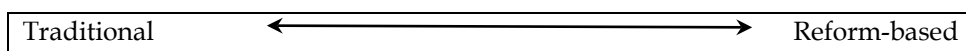
Table 2
Preservice Teachers' Assessment Perspectives

Assessment Perspectives			
	Test-oriented	Task-oriented	Tool-oriented
Alex	"From the student point of view, [assessment] is a way to get a good grade." From the teacher point of view "it's a way to evaluate student learning." (0-3 months)	"[Almost] anything you do is an assessment piece." Assessment can be a "formative" task, which is "weighted less," or a "summative" task, which is "weighted more." (4-9 months)	N/A
Ben	"...when I thought of assessment I would think 'give them a test' ..." (0-3 months)	"I've come to see [assessment] ... for both testing and for monitoring student progress." (0-3 months)	"Static assessment seeks only a snapshot of a student's knowledge..." With "dynamic assessment" the student is "actually learning [while] being assessed." (4-9 months)
Erika	"Assessment means evaluation, which I would say is grading." (0-3 months)	"[Assessments are] anything that lets me know what my kids know. They can be formal or informal. The difference is that formal is points and informal is not points." (4-9 months)	N/A
Jake	"[Assessment] literally meant a test or quiz" [and it is] very procedural ... just 'How do you solve this problem?'" (0 - 3 months)	"The word 'assessment' means ... someone is giving you information ... and you are checking [it]. [Assessment] doesn't have to be formal, it can just be questions to kids." (0-6 months)	"Formative assessment is ... [when] you let [students] know where they went right or wrong so they keep doing the right thing or change what they did wrong. Summative assessment is where you are not giving

			students any feedback ... [and] it counts towards their grade." (7 - 9 months)
Karen	N/A	"Formal for me means you give a grade and informal means you don't." (0-3 months)	"Formative assessment is anything that tells me what to do next, or tells a student what they don't know and what they need to do next. Tests are summative." (4-9 months)
Mary	"Assessment is a value that I give to student's understanding [of procedurally based questions]." (0-3 months)	Assessment can be a formal test-like task, an informal non test-like task, a formative and ungraded task, or a summative task used to determine the success of the unit ... " (4-6 months)	"With formative assessment I am checking what [students] know as we are doing it and I am modifying my lesson ... [With] summative assessment you don't modify anything." (7-9 months)

Although assessment was described as a test, PSTs' test-oriented assessment practices were not limited to tests. In fact, each PST employed two or three assessment tasks outside of tests during this stage. Alex, for example, assessed using post-lesson worksheet and "exit pass" tasks, Erika assessed using homework and end-of-unit student presentations, and Mary assessed using pre-lesson warm-up tasks, post-lesson individual student whiteboard tasks and exit passes. As these examples illustrate, PSTs' assessments consisted of a limited number of individual tasks that occurred separately from, and predominantly at the end of, instruction. With each of these tasks, students were prompted to solve closed problems by applying an algorithm. By and large, these questions focused on the most immediately discussed algorithm with little connection to any broader concepts; however, each PST did employ one form of assessment that promoted connections between procedures and their underlying concepts. One of Erika's student presentations, for example, prompted students to solve an equation "in as many ways as possible" to promote the idea that equations can be solved in a variety of ways. The purpose of every assessment task was to measure how well students understood the most immediately discussed topics. After each assessment was completed, PSTs judged student responses based on accuracy and used the results for grades, to decide what topics required further discussion, or both. These test-oriented assessment practices are shaded lightly in grey in Figure 2.

Alex, Ben, Erika, Jake, and Mary – five of the six PSTs – entered teacher education with these test-oriented perspectives and practices. These perspectives and practices largely reflected each PST's K-16 school experiences. As Mary noted during her initial interview, "All that I was given [in school] was just tests and projects, so that's all I really know about assessment."



Purpose	To record achievement of individuals (evaluative)	To monitor progress or learning (summative)	To inform teaching (diagnostic)	To inform learning (formative)
Use	Report to student, parent, teachers, etc.	Feedback into teaching plans	For students and teacher	
Timing	Separate task	Intermission from classwork	Normal part of classwork	
Feedback	Criterion referenced	Student and criterion referenced	Student referenced	
Task	Closed	Open-middle	Open-ended	
Cognitive Demands	Procedures without connections	Procedures with connections	Doing mathematics	
Methods	Limited methods	Several methods	Variety of methods	




Key:  Test-, task-, and tool-oriented practices  Task- and tool-oriented practices  Tool-oriented practices

Figure 2. Changes in assessment practices.

Over the course of the first few months of each teacher education program, all six PSTs were introduced to a variety of novel assessments, such as student presentations and clickers (interactive response devices). Often, these introductions involved the methods instructor modelling a contemporary assessment method or task for the class while asking PSTs to participate by taking on a student role. Afterwards, each methods instructor invited PSTs to talk through the assessment as teachers. In addition, PSTs were assigned to read practitioner-oriented articles that addressed topics such as questioning strategies, monitoring student progress during instruction, and grading. Following each assignment, each instructor asked the class to reflect on the readings, share their thoughts, and discuss the literature in relation to their practices or other teachers' practices that they witnessed.

Around the same time as these introductions to novel assessments in the methods classes, PST's test-oriented assessment perspectives and practices began evolving into task-oriented ones. This change was not simultaneous, however. Ben and Jake both entered this stage at some point during the beginning (0-3 months) of their respective programs, while Alex, Erika, and Mary did not do so until the middle (4-6 months). Karen entered her teacher education program with task-oriented perspectives and practices, which could potentially be explained by her prior teacher training.

Task-oriented perspectives and practices. The middle assessment perspective column of Table 2 contains excerpted quotes to illustrate each PST's task-oriented assessment perspective. As the table helps illustrate, the task-oriented assessment perspective is defined in part by a broader view of assessment. For example, during the second interview, Erika noted, "I thought assessment were tests, but it's *anything* that lets me know what my kids know," (emphasis added). As this quote highlights, at this stage each PST came to view assessment as more than a test or a grade. Along with this broader view came a distinction between tasks that were graded and tasks that were not necessarily graded but used to assess students nonetheless. Karen, for example, distinguished between "formal" and "informal" assessments. "Formal means you give a grade and

informal means you don't," she explained. These comments typify PSTs' focus at this stage on the kinds of tasks that were being employed. Erika, Ben, and Jake also noted that assessment involved a more and less precise way to be implemented. As Erika put it, "We had all this research to read which proved to us that [reform-based] assessments work...once we were all won over it was like, 'OK, we need to learn how to do this the *right* way because we don't know'," (emphasis added).

The cells shaded moderately in grey in Figure 2, along with the cells shaded lightly in grey, reflect PSTs' task-oriented practices. As the figure illustrates, during this stage PSTs' practices consisted of individual and closed tasks, which focused on procedures with and without any connections to other mathematical ideas, and were used for both grading and to decide what topics needed further review. While these elements of their practice mirrored those of the test-oriented practices, PSTs were now employing more assessment methods. In fact, PSTs employed an average of five different assessment methods, including verbal questions, observations, and poster presentations. Moreover, all PSTs were now implementing tasks prior to their lessons, as well as during lessons as an intermission from instruction. Jake, Karen, and Mary, for example, periodically gave their students problems to solve as a checkpoint to determine students' readiness for the next part of the lesson.

Feedback also became more of a priority during this stage of development. All PSTs corrected students' inaccuracies through verbal or written feedback to help clear up any misunderstandings.

Perhaps the most significant difference between the test- and task-oriented practices, however, concerned the purpose of assessment. That is, in addition to designing assessment tasks to measure class progress, each PST also created diagnostic assessment tasks to inform their instruction. For instance, Karen and Mary used daily in-class activities to decide if students needed to review a topic before proceeding with instruction.

Though not included in our framework, a common issue raised with PSTs' task-oriented practices was a lack of adequate information about student learning. Erika, Alex, Mary, and Jake all spoke of being surprised by their students' performance on various assessments. This suggests that they did not have a solid understanding of what their students knew or were capable of doing leading up to these particular assessments. Furthermore, Erika and Alex described themselves as having only a "broad view" of student understanding at the end of a unit's worth of instruction.

Alex and Erika finished teacher education with these task-oriented perspectives and practices. The assessment perspectives and practices of the remaining four PSTs, however, evolved from task- to tool-oriented. Ben and Karen's perspectives and practices evolved into this stage by the middle (4-6 months) of their respective programs, while Jake and Mary did so by the end (7-9 months).

During this latter half of teacher education, each methods instructor presented PSTs with a succession of increasingly sophisticated frameworks for thinking about assessment; first by describing assessment as a means of knowing what students know, then by adding that assessment was also a way of informing subsequent lessons, and finally by discussing assessment in terms of its "formative" and "summative" functions. In essence, both instructors described formative assessment as tasks used to inform teachers and students of progress towards an instructional goal, and to make decisions about how to proceed towards reaching that goal. Summative assessment, on the other hand, was described by both instructors as "evaluative" and often characterised as a test or other end-of-unit exam.

At the same time that these frameworks were being discussed, PSTs also engaged in a variety of practical activities. All six PSTs were required to develop lesson plans, which included descriptions of how students were to be assessed and how that information would be used. Both methods courses also required PSTs to write a "Total Assessment Plan." For this exercise, PSTs were asked to explain the types of assessments used in class, the purpose of each assessment, the

information gleaned from each task, and how assessment tasks were used individually and collectively. In addition to tasks such as these, PSTs examined, analysed, and reflected on samples of their students' work as part of their PACT.

Tool-oriented assessment perspectives and practices. The rightmost column of Table 2 reflects PSTs' tool-oriented assessment perspectives. The defining feature of this tool-oriented perspective was a distinction between assessments used for grading purposes, and assessments used to support student or teacher learning. The assessment tasks that were used for grading purposes, which were most commonly referred to as "summative assessments," resembled the graded tasks discussed above (e.g., Karen's "formal assessments"). The assessments that were used to support learning, often called "formative assessments," were viewed as tasks that either provided direction for improvement or facilitated student learning. Karen, for example, noted that, "[Formative] assessment is anything that tells me what to do next, or tells a student what they don't know and what they need to do next," while Ben commented that this is a "form of assessment where [students] are actually learning as [they] are being assessed." In other words, these PSTs now viewed assessment as a tool that served a particular purpose.

When comparing the task-oriented and tool-oriented practices, many characteristics are similar. For example, PSTs employed the same assessment methods, and these methods continued to focus on procedures with and without connections to other concepts. Though there were many similarities, in this stage the "formative assessment" tasks were designed or used to "scaffold" ideas so that students were able to make connections or draw conclusions, to "differentiate" work so that tasks were tailored to individual student's needs, or in some other way to enable students to construct a deeper understanding of the concepts discussed in class. The cells shaded in grey with some texture in Figure 2 represent these practices that are unique to the tool-oriented stage.

A few practices were not shared among all four PSTs in the tool-oriented group. While most assessment tasks contained closed questions, Ben's tests included a number of open-ended problems that were graded based on a rubric that took into account student growth in understanding. Though the assessment tasks focused on procedures, Mary facilitated class discussions in which students were encouraged to develop and challenge each other's theorems, postulates, or other ideas. More, these assessments took place as part of the lesson. No other PST employed such a technique, thus they were not reflected in the collective tool-oriented practices. It is also important to note that, with tool-oriented practices, no PST discussed issues with the adequacy of information on student learning, nor did any of these four PSTs mention being surprised by any student performances.

In the next section, to help illustrate the progression through test-, task, and tool-oriented stages, we present a close examination of Mary's assessment perspectives along with one of her assessment methods during the beginning, middle, and end of teacher education.

Mary: From "How?" to "What?"

In the following illustrative account, we present Mary's assessment perspectives and "whiteboard checks" practice during the beginning, middle, and end of her teacher education program, which also extends across all three stages. Because Mary taught two courses during her field practicum, the first 2 whiteboard check methods described below were employed in an Algebra I course, and the final whiteboard check in a Geometry class. Mary's whiteboard checks always involved 30-centimeter by 45-centimeter dry erase whiteboards, which were given to every student in the class. Students were able to place the whiteboards on their desk to work, and then hold them over their heads to display to the teacher or to other students in the classroom. Depending on the activity, students might have written one problem and solution on their whiteboard at a time, a

number of answers without any other work, or something in between. Students' work was never graded. Furthermore, Mary never documented or otherwise kept track of students' answers or solution strategies. Mary did, however, examine and assess students' work as they displayed their answers.

Mary's test-oriented perspective and whiteboard check. Initially, Mary described assessment as a "value that [she gave] to student understanding," which was based on their "ability to process skills and computation." For Mary, assessment tasks must provide the teacher with information about individual student understanding. Homework and other out-of-class projects were not assessment, for example, because students could receive help in completing the assignment. "How do I know my students don't go home and have parents [say], 'Oh, I know how to do that, I'll help you.' That's not really assessment because I don't know whose knowledge it is," Mary explained.

At this point, Mary's whiteboard checks immediately followed her instruction. For these activities, Mary gave her students a number of problems to solve. Each problem prompted students to apply the mathematical procedure, algorithm or skill recently illustrated and discussed during the lesson. "The problems will have varying degrees of difficulty," Mary said, "But [they are] all about practicing a skill." As students worked, Mary walked around to judge the accuracy of their work and "...tell students if they got it right or not." According to Mary, the purpose of these particular tasks was "to get a read on exactly what each student knows and where they are going wrong." She continues:

So, from every whiteboard check I hope to have an idea of what holes people have. So, is it like a skill hole related to what we are doing, or is it a hole from adding or subtracting integers, or a hole from fractions?

With this information Mary was able to "monitor [her] students' learning during class to tell when the students have enough to move on and when we need some extra time to review."

Analysis. Mary's description of assessment as a "value" derived from a student's "skill and computation" work is indicative of the test-oriented assessment perspective. As noted earlier, despite such descriptions not all assessment tasks were graded. Indeed, Mary did not grade students' whiteboard check work. Still, Mary intended this assessment to monitor or "get a read on" her students' level of understanding. These whiteboard checks were separate tasks that occurred after her instruction. With regard to content, the questions focused on the most immediately discussed "skill" or, in other words, procedures without connections. As Mary observed student work and offered feedback, her comments were focused on accuracy. Thus, her feedback was criterion referenced. Once the assessment was complete, Mary used the assessment information to monitor student progress against her instructional plans.

Mary's task-oriented perspectives and whiteboard check. By the middle of her preparation program, Mary described assessment as an exercise in which "students are telling [her] what they know." Mary distinguished between "informal" and "formal" assessment. "Informal I do everyday like...whiteboards. [They're] just things that aren't like tests," Mary said. "Formal assessment," on the other hand, "like tests and quizzes...[are] things for grades." In addition to informal and formal, Mary distinguished between "formative assessments," which were tasks that "happen often, they happen during the lesson, and [she didn't] grade them," and "summative assessments," which were "tests" designed to "determine the success of a unit ...".

At this time, Mary designed her whiteboard checks to "dictate what [she did] in class." Following a brief lesson, which typically involved one or two example problems, Mary gave her students a "skill drill" problem. These problems mirrored the examples in the lesson, and entailed one correct answer. After students completed the problem, they raised their boards over their

heads to display their work to Mary. Mary quickly assessed their answer and "[told] all students if they got it right or exactly what step they got wrong." For example, during a whiteboard check in which students were asked to simplify the rational expression $\frac{3x(x-3)}{(x-3)}$, Mary said "Good" to a

number of students as they held up the correct answer. And when a student held up a whiteboard with the answer " $3x/1$ ", Mary replied: "Simplify it more."

After all students showed their work to Mary, she made a decision about how to proceed. If a majority of the students got the problem wrong, then Mary would review the problem with the class and then give students a similar type of problem to solve. However, "If by looking at the white boards the majority of the class (at least 75%) got it [right], then [Mary would not] go over it." Instead, Mary would give students an incrementally more difficult problem. She explained:

I start by looking for if they know how to do problem 'A'? Then, do they know how to do problem 'B'? Then, can they do 'A' and 'B' in one problem? If they can do 'A' and 'B' in one problem, now can they do 'C'?

After students successfully completed problems A, B, and C, Mary believed that students were "prepared to see the new material." Accordingly, she would continue with the next part of the lesson.

Analysis. At this point, Mary's description of assessment was much broader than her initial description that equated assessment to a value. Though Mary distinguished between both formal and informal assessments, as well as formative and summative assessments, both distinctions were concentrated primarily on grades. That is, formal and summative assessments were tasks that were graded; informal and formative assessments were tasks that either went ungraded or weighed less on students' grades than formal or summative assessments. These views highlight Mary's progression from a test- to a task-oriented perspective.

Along with the change in Mary's views, there were also a number of changes to her whiteboard check. Instead of conducting these checks simply to monitor student progress, Mary now designed her whiteboard checks to "dictate" what she taught and when she taught it. Because Mary conducted whiteboard checks at various points throughout her lecture, they acted more as an intermission than a separate task. As students displayed their work, Mary's feedback judged accuracy (e.g., "good") as well as offered corrective suggestions (e.g., "simplify it more"). Thus, the feedback at this stage was both student and criterion referenced.

While some of Mary's whiteboard check practices changed, there were also a few that did not. The questions that students were asked to solve were still closed, they focused on the most immediately discussed skill or procedure with little connection to any broader concepts within the unit, and Mary continued to use the information from these tasks to make decisions about what topics to review.

Mary's tool-oriented perspectives and whiteboard check. Mary continued to distinguish between informal and formal assessments, as well as formative and summative assessments, at the end of her teacher education program. For Mary, the informal-formal distinction continued to concern the form of assessment (i.e., tests and non-tests); however, the difference between formative and summative assessment was now about how tasks were being used. "With formative assessment I am checking what they know as we are doing it and I am modifying my lesson depending on what they do and don't know... [With] summative assessment you don't modify anything," Mary explained. For Mary, modifying lessons involved trying to "build off prior knowledge," as well as "differentiating," and "scaffolding." Building off prior knowledge involved using assessment information to decide when to move forward with instruction. Differentiating involved assigning different tasks to different students based on need, which allowed Mary to "spend extra time with slower students and give them the support needed...without making the material repetitive for

students who have already mastered the objective." Scaffolding involved "break[ing] one problem into really simple questions," in order to make the task more accessible to students all the while better identifying student misconceptions to make more specific instructional decisions about what and when to review or offer individual student support.

Mary organised her whiteboard checks during this time "to move seamlessly from [the warm-up] into the lesson." As students sat down at their desk each day they immediately went for their whiteboards to work on a problem that was presented on the front board. Once finished, students held their whiteboards over their head. Mary then examined each student's work and provided feedback. For example, after asking the question "How many degrees are in a triangle?" Mary said "good" to a number of students as they held up their boards with the answer "180°." And when students held up "180" as an answer, Mary gave comments such as "Show units [of measurement]."

During this time, Mary also "differentiated" the task by asking students who correctly answered the problem to help those who did not. Like before, when a majority of students did not answer the problem correctly, Mary would review it with the class. Unlike before, for this review, Mary often asked a student to share their solution and then challenged the rest of the class to either prove or disprove the answer. Once a majority of students answered the question correctly, Mary gave the class another problem to answer. Instead of increasing in difficulty, as with the previous whiteboard checks, Mary designed these problems to "build up to stuff...[and] scaffold the conversation." That is, Mary created and structured the whiteboard check questions to help foster connections for students between their current understandings and new concepts that were part of the day's lesson. As part of the whiteboard check in which Mary asked for the sum of the interior angles in a triangle, for example, Mary projected three polygons (see Figure 3) on the board. Each polygon had lines leaving one vertex and connecting to all the remaining vertices, which divided the polygons into smaller triangles. Mary then asked, "How many triangles are in each of the polygons?" and "What is the sum of the interior angles of each polygon?" Through a discussion of these questions, Mary hoped that students would generate some ideas about how to answer the question: "What is the formula for the sum of the interior angles of a n-gon?"

Warm-Up

(do not erase until you have completed #5)

1. How many degrees are in a triangle?
2. How many sides do the following shape have?



3. How many triangles are in each of the polygons in #2?
4. What is the sum of the interior angles of each polygon in #2?
5. What is the formula for the sum of the interior angles of a n-gon?

Figure 3. Mary's whiteboard check with scaffolded questions.

Analysis. Mary's new distinction between formative assessments, which were used to modify instruction, and summative assessments, which were not used to modify anything, are most consistent with the tool-oriented assessment perspective.

Concerning her whiteboard checks, Mary no longer designed these tasks to obtain information about student understanding, or to inform her instruction. Although Mary's whiteboard checks continued to do just that, now the purpose of these tasks was to support students' understanding of mathematical topics. Because these checks started at the outset of most class days and were incorporated into the lesson, this assessment was a normal part of instruction. While problems still largely focused on procedures, with Mary's scaffolding of questions the task also helped bridge connections to broader mathematical concepts. Moreover, when students answered problems incorrectly, Mary facilitated class discussions in which students were encouraged to develop and challenge each other's ideas and solution strategies. Much like with her earlier versions of whiteboard checks, Mary used the information from these tasks to decide what topics to review. However, these reviews involved class discussions where students were called upon to examine and analyse each other's work. Furthermore, as a way of differentiating the task, Mary incorporated peer-to-peer instruction during whiteboard checks when certain students correctly solved the problem while others were still working. In other words, students were just as actively involved as Mary in using assessment information.

For all that changed with Mary's whiteboard check assessment practice, there were still a few things that remained the same. Mary's whiteboard check questions often called for an algorithm or procedure, albeit to develop students' understanding of concepts, and all questions were closed in that they typically involved a numeric answer and all questions entailed one correct solution. In addition, Mary's feedback still focused on accuracy and offered corrective suggestions (i.e., student and criterion referenced).

These three accounts of Mary's whiteboard checks offer a concrete example of how PSTs' assessment perspectives and practices evolved. Initially, when Mary described assessment as a "value" assigned to students' procedural tasks, her whiteboard checks served as an opportunity for students to practice the procedure discussed during lecture, and as an opportunity for her to monitor learning and make decisions about topics to review (i.e., test-oriented perspectives and practices). When Mary's assessment perspective developed into a division between (formal and summative) graded and (informal and formative) ungraded tasks, her whiteboard check worked more as a checkpoint in which she was able to make decisions about whether to review or proceed with instruction as circumstances dictated during lecture (i.e., task-oriented assessment perspectives and practices). Once Mary developed the view that assessment tasks could be used to modify lessons, her whiteboard checks involved questions that were crafted in such a way as to help students draw conclusions or develop understandings on their own. Furthermore, Mary used the information from whiteboard checks to "differentiate" the task for students, and facilitate whole-class discussions (i.e., tool-oriented assessment perspectives and practices).

What we see here is a notable shift in how Mary designed and used her whiteboard checks, from the more traditional practice of monitoring what students learned from instruction and addressing their misconceptions, to a more reform-based practice of designing the task to facilitate learning, and using insights gleaned from the task to support individual student's and whole class needs. What we don't see here, on the other hand, is much growth or change from the closed and procedural tasks that Mary used at the outset of her teacher education program.

Discussion and Implications

We identified three stages that occurred in PSTs' assessment perspective and practice evolution. In the first, test-oriented stage, PSTs described assessment as a test or a grade, and their

assessments involved many traditional practices including procedural and closed tasks that were designed and used for summative purposes. With the second task-oriented stage, PSTs described assessment as anything that elicited information about student learning, and they distinguished between graded and ungraded assessment tasks. During this stage, PSTs added more tasks to their practice, designed certain tasks to inform their teaching, and began to implement some tasks during class time (as opposed to at the end of it). In the third tool-oriented stage, PSTs distinguished between assessment for grading and learning purposes, and they employed more reform-based practices, namely, designing and using assessment to facilitate student learning.

In light of previous research highlighting PSTs' struggles to adopt key reform-based assessment practices, we find this evolution encouraging. However, this evolution reveals more than PSTs' progression from traditional to more reform-based assessment perspectives and practices. We interpret this progression as evidence that, when learning about assessment, PSTs initially focus on *how to assess*.

This initial focus was apparent in PSTs' increased number of assessment methods from the test-oriented to the task-oriented stage. Though PSTs typically added a few assessments from the task- to the tool-oriented stage, the number of tasks added, if any, was usually small. Additionally, in expressing task-oriented views, Erika, Ben, and Jake indicated an interest in assessing "the right way." Among other things, this suggests a focus on assessment implementation, or how to assess. Finally, a common issue with PSTs' task-oriented assessment practices involved a shortage of information about individual student knowledge. This can be attributed to a number of things, not the least of which is a focus on implementing the task (as opposed to accessing student understanding).

Only after learning how to assess do PSTs begin to consider other assessment functions. This is perhaps best illustrated by a comment made by Mary at the end of her teacher education program: "My assessment is not 'how can I assess?' anymore, it is '*Why am I assessing?*' and '*What is most important to assess?*'" (emphasis added). In addition, PSTs began to consider "*How do we use assessment?*" to evaluate and support learning. This was apparent in each PST's assessment perspectives, as the tool-oriented stage represents the first time that PSTs distinguished between assessments for grading purposes and those for supporting student and teacher learning. Furthermore, it was apparent in each PST's practice, as their classroom assessments took on new and reform-based functions. Such was the case with Mary's whiteboard checks, for example, where by the end of her program she scaffolded her questions to help build on students' prior knowledge, and differentiated the tasks to attend to students' learning needs.

From this evolutionary portrait, there are a number of important implications for teacher education. First, five of the six PSTs entered teacher education with test-oriented or traditional notions of assessment. This not only seems consistent with the research we have reviewed concerning PSTs' assessment practices, it also reaffirms research findings about PSTs' assessment knowledge upon entering teacher education (e.g., Maclellan, 2004). Thus, while this research involved a relatively small group of participants, it would seem safe to predict that PSTs, at least in the not too distant future, will continue to enter teacher education programs with traditional and therefore limited notions of assessment. Accordingly, mathematics teacher educators must continue to help develop, and look for new approaches to enrich and broaden, PSTs' notions of mathematics classroom assessment.

Although literature identifying effective assessment education in teacher education programs is sparse (DeLuca & Klinger, 2010), there are a number of documents that outline various ways to promote reform-based assessment among PSTs. For example, Otero (2006) suggests teacher educators ground the concept of formative assessment in relevant learning theory, and Shepard et al. (2005) suggests PSTs engage in tasks such as analysing student work and designing assessment plans. While these may indeed be ingredients in the recipe for

successful assessment education, based on our findings we argue that the first step involves introducing PSTs to novel assessment methods or forms. Through introductory activities, such as engaging in an assessment task as an assessee and then discussing it as an assessor, PSTs' perspectives and practices become broader and less traditional. Other introductory activities have led to similar results (e.g., A. Herrington, J. Herrington, Sparrow, & Oliver, 1998). Our findings suggest that as PSTs continue to move through teacher education and better understand how to assess, they must then be given opportunities to learn about reform-based assessment purposes, content, functions, and the like. Structuring teacher education courses and coursework in such a way is essential to meet PSTs' assessment learning needs.

Still, based on this research, PSTs will continue to consider assessment an individual task that involves closed and procedurally-based problems. It is imperative that teacher education programs address these issues to help PSTs (and their students) realise the full potential of reform-based assessment. Thus, mathematics teacher educators must carefully consider and explore new ways to foster PSTs' use of more open-ended assessment tasks that encourage students to engage in the process of "doing mathematics."

Conclusion

We began by noting that research on PSTs' assessment practices, which suggests that PSTs struggle to employ the ideas and methods emphasised in teacher education, is largely outcome focused. If it is worth reforming the assessment practices of PSTs, then it is worth thoroughly understanding their emerging assessment perspectives and practices as they prepare for their teaching careers. The intent of this paper was to shed light on PSTs' processes of learning to assess while traveling through reform-minded teacher education programs. It is our hope that, just as with previous work on teacher learning, this research provides valuable insight that can be used to help ensure that sound contemporary assessment practice is learned, appropriately applied, and sustained throughout a teacher's professional career.

Acknowledgements

We would like to thank each of the preservice teachers for their willingness to participate in this research.

References

- Berliner, D. (1994). Expertise: The wonder of exemplary performances. In J. Mangieri & C. Block (Eds.), *Creating powerful thinking in teachers and students: Diverse perspectives* (pp. 161-186). Fort Worth, TX: Harcourt Brace College Publishers.
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139-148.
- Brahier, D. J. (2001). *Assessment in middle and high school mathematics: A teacher's guide*. Larchmont, NY: Eye on Education.
- Brown, C., & Borko, H. (1992). Becoming a mathematics teacher. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 209-239). New York: Macmillan.
- Brown, C., Cooney, T., & Jones, D. (1990). Mathematics teacher education. In W. R. Houston (Ed.), *Handbook of research on teacher education* (pp. 639-656). New York: Macmillan.
- Cockcroft, W. (1982). *Mathematics counts: Report of the committee of inquiry into the teaching of mathematics in schools*. London: HMSO.
- DeLuca, C., & Klinger, D. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17, 419-438.

- Even, R. (2005). Using assessment to inform instructional decisions: How hard can it be? *Mathematics Education Research Journal*, 17(3), 45–61.
- Frykholm, J. (1999a). Assessment in mathematics teacher education: Introducing preservice teachers to assessment reform. *The Teacher Educator*, 34, 244–258.
- Frykholm, J. (1999b). The impact of reform: Challenges for mathematics teacher preparation. *Journal of Mathematics Teacher Education*, 2, 79–105.
- Fuller, F. (1969). Concerns of teachers: A developmental conceptualization. *American Educational Research Journal*, 6, 207–226.
- Gardner, J. (Ed.). (2006). *Assessment and learning*. London: Sage.
- Hammerness, K., Darling-Hammond, L., & Bransford, J., with Berliner, D., Cochran-Smith, M., McDonald, M., & Zeichner, K. (2005). How teachers learn and develop. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world* (pp. 275–326). San Francisco: Jossey-Bass.
- Harlen, W. (1998). Classroom assessment: A dimension of purposes and procedures. In K. Carr (Ed.), *SAME papers* (pp. 75–87). Hamilton, New Zealand: Centre for Science, Mathematics and Technology Educational Research, University of Waikato.
- Harlen, W. (2006). On the relationship between assessment for formative and summative purposes. In J. Gardner (Ed.), *Assessment and learning* (pp. 103–117). London: Sage Publications.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Herrington, A., Herrington, J., Sparrow, L., & Oliver, R. (1998). Learning to teach and assess mathematics using multimedia: A teacher development project. *Journal of Mathematics Teacher Education*, 1, 89–112.
- Lund, A. (2008). Assessment made visible: Individual and collective practices. *Mind, Culture, and Activity*, 15, 32–51.
- Maclellan, E. (2004). Initial knowledge states about assessment: novice teachers' conceptualizations. *Teaching and Teacher Education*, 20, 523–535.
- National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.
- National Council of Teachers of Mathematics (1991). *Professional standards for school mathematics*. Reston, VA: Author.
- National Council of Teachers of Mathematics (1995). *Assessment standards for school mathematics*. Reston, VA: Author.
- National Council of Teachers of Mathematics (1999). *Mathematical assessment: A practical handbook for grades 9 – 12*. Reston, VA: Author.
- National Council of Teachers of Mathematics (2000). *Principles and standards for school mathematics*. Reston, VA: Author.
- National Council of Teachers of Mathematics (2007). *What is formative assessment?* Reston, VA: Author.
- O'Connor, K. (2002). *How to grade for learning*. Thousand Oaks, CA: Sage.
- Ohlsen, M. T. (2007). Classroom assessment practices of secondary school members of NCTM. *American Secondary Education*, 36(1), 4–14.
- Otero, V. K. (2006, June). Moving beyond the "get it or don't" conception of formative assessment. *Journal of Teacher Education*, 57, 247–255.
- Performance Assessment for California Teachers [PACT]*. Retrieved July 7, 2014 from <http://www.pacttpa.org/>
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, 48(1), 4–11.
- Pratt, E. O. (2002). Aligning mathematics teacher work sample content with selected NCTM standards: Implications for preservice teacher education. *Journal of Personnel Evaluation in Education*, 16, 175–190.
- Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119–144.
- Scribner, S., & Cole, M. (1981). *The psychology of literacy*. Cambridge, MA: Harvard University Press.
- Senk, S. L., Beckmann, C. E., & Thompson, D. R. (1997). Assessment and grading in high school mathematics classrooms. *Journal for Research in Mathematics Education*, 28, 187–215.
- Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4–14.
- Shepard, L. (2001). The role of classroom assessment in teaching and learning. In V. Richardson (Ed.), *Handbook of research on teaching* (pp. 1066–1099). Washington, DC: American Educational Research Association.

- Shepard, L., Hammerness, K., Darling-Hammond, L., Rust, F., Snowden, J., Gordon, E., Gutierrez, C., & Pacheco, A. (2005). Assessment. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world* (pp. 275–326). San Francisco: Jossey-Bass.
- Stein, M., Grover, B., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Education Research Journal*, 33, 455–488.
- Stenmark, J. K. (1991). *Mathematics assessment: Myths, models, good questions, and practical suggestions*. Reston, VA: National Council of Teachers of Mathematics.
- Stiggins, R. (1999). Evaluating classroom assessment training in teacher education. *Educational Measurement: Issues and Practice*, 18(1), 23–27.
- Stiggins, R. (2008). *Assessment manifesto: A call for the development of balanced assessment systems*. Portland, OR: Assessment Training Institute.
- Suurtamm, C., Koch, M., & Arden, A. (2010). Teachers' assessment practices in mathematics: Classrooms in the context of reform. *Assessment in Education: Principles, Policy & Practice*, 17, 399–417.
- Task Group on Assessment and Testing (1988). *National curriculum: Task group on assessment and testing. A report*. London: Department of Education and Science and the Welsh Office.
- Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education*, 30, 749–770.
- Watt, H. (2005). Attitudes to the use of alternative assessment methods in mathematics: A study with secondary mathematics teachers in Sydney, Australia. *Educational Studies in Mathematics*, 58, 21–44.
- Wiliam, D. (2007). Keeping learning on track: Classroom assessment and the regulation of learning. In F. K. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 1053–98). Charlotte, N.C.: Information Age Publishing and National Council of Teachers of Mathematics.
- Webb, N. (1992). Assessment of students' knowledge of mathematics: Steps toward a theory. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 661–683). New York: Macmillan.
-

Authors

Matt Wallace, School of Education, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA

Email: mtwallace@ucdavis.edu

Tobin White, School of Education, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA

Email: tfwhite@ucdavis.edu