Enhancing Teachers' Assessment of Mathematical Processes Through Test Analysis in University Courses*

Patricia D. Hunsader
University of South Florida, Sarasota-Manatee

Barbara Zorin
St. Petersburg College

Denisse R. Thompson
University of South Florida

Assessment is a critical component of the teaching and learning cycle. Yet, research suggests that teachers have often had insufficient preparation relative to the development and use of assessment. In this article, we share experiences and assignments we use with both preservice and in-service teachers within undergraduate and graduate university courses to enhance their focus on mathematics assessment, particularly assessment of processes and practices in classroom tests. We also share the results of teachers' analyses of classroom tests, their reactions to their analysis, and their reflections on the potential impact of the experiences on their future practice.

Key words: Mathematics assessment; Mathematical processes; Test analysis; Mathematics methods courses

The National Council of Teachers of Mathematics (NCTM) describes assessment of mathematical understanding as "the process of gathering evidence about a student's knowledge of, ability to use, and dispositions toward, mathematics and making inferences from that evidence for a variety of purposes" (1995, p. 3). Identifying learning goals and assessing whether instruction is helping students achieve those goals is a significant element of teaching (DeWitt & Freie, 2005; NCTM, 1989; Romagnano, 2001; Schoenfeld, 2002). Indeed, classroom assessment is a "bridge between teaching and learning, helping teachers collect evidence about student achievement in order to adjust instruction to better meet student learning needs" (Wiliam, 2007, p. 1054).

Although classroom assessment is key to maximizing student learning (Lukin, Bandalos, Eckhout, & Mickelson, 2004), research indicates that teachers are often not adequately prepared to make decisions about classroom assessment practices (Fan, Wang, & Wang, 2011; Koh, 2011; McMillan, Myran, & Workman, 2002; Quilter & Gallini, 2000; Stiggins, 2002), and they recognize these deficits in their preparation (Chelsey & Jordan, 2012; Mertler, 2009; Zhang & Burry-Stock, 2003; Zientek, 2007). Assessment courses are not required in all teacher education programs, and such courses are often not content specific (DeLuca & Klinger, 2010), providing little to no instruction in mathematics assessment.

Numerous educators recommend that mathematics assessment address more than just knowledge of content. It should also provide insights into students' higher-order thinking, reasoning, problem solving, communication, and conceptual understanding (Borko, Mayfield, Marion, Flexer, & Cumbo, 1997; Boston & Smith, 2009; NCTM, 1989, 1995, 2000). Because *how* students learn (i.e., the mathematical processes/practices through which they learn) is as important as *what* they learn (i.e., the specific mathematical content) (Kilpatrick, Swafford, & Findell, 2001; Ma, 1999; National Governors Association Center for Best Practices (NGA Center) and Council of Chief State School Officers (CCSSO), 2010; NCTM, 1989, 2000, 2006), assessments should focus not just on content mastery and procedural skills, but also on mathematical processes and practices that provide opportunities for students to demonstrate their mathematical thinking. As students provide insight into their thinking about important concepts and processes, teachers are positioned to modify and adjust classroom instruction to enhance learning.

Classroom assessment encompasses a range of forms, including informal observation, students' individual and group work, performance assessments, portfolios, journal writing, and tests. The National Research Council, in its report *Everybody Counts* (1989), highlights one form of classroom assessment that deserves attention: "We must ensure that tests measure what is of value, not just what is easy to test" (p. 70). Tests also need to "be sensitive enough to help teachers identify individual areas of difficulty in order to improve instruction" (NCTM, 1989,

* This manuscript was edited by Margaret Smith, editor of the *MTE* at the time the manuscript was initially submitted.

p. 192). Yet, Black and Wiliam note, "The tests used by teachers encourage rote and superficial learning even when teachers say they want to develop understanding; many teachers seem unaware of the inconsistency. … [questions and assessment methods] are not critically reviewed in relation to what they actually assess" (2010, pp. 83–84). Alkharusi, Aldhafri, Alnabhani, and Alkalbani's (2012) work indicates the potential of professional development to impact teachers' assessment knowledge and skills. It follows that enhancing teachers' ability to analyze their classroom tests objectively is an important area of focus within teacher preparation and a potentially fruitful area for professional development with in-service teachers.

This article focuses on an assessment experience we use within undergraduate and graduate university courses for both preservice and in-service teachers to enhance their ability to be critical consumers of classroom tests through analyzing and then modifying tests, as needed, to ensure tests provide opportunities for students to engage with important mathematical processes. In particular, we share the broad parameters of the experience used across different methods courses, and then focus in more detail on an expansion of the experience in a graduate course on current trends in mathematics education. We also share data collected from several instantiations of the experience to determine the extent to which the experience helped teachers reflect on the nature of tests accompanying their curriculum or used in their classes.

Rationale for Enhancing the Focus on Test Analysis Relative to Mathematical Processes

Smith and Stein (1998) note that to evaluate a task effectively, it is necessary to look beyond the task's surface features to analyze the kind of thinking required. Similarly, we believe it is important to look beyond the content being assessed by a test item to determine the extent to which it provides students with opportunities to engage with mathematical processes/practices teachers want to support. Test items may assess important content without focusing on mathematical processes but can often be easily modified to focus on processes while still assessing the intended content (e.g., Thompson, Beckmann, & Senk, 1997).

In our earlier research related to assessment, we built on the work of Senk, Beckmann, and Thompson (1997) and other researchers (e.g., Madaus, West, Harmon, Lomax, & Viator, 1992; Taylor, 1992) to develop a framework to analyze tests accompanying published curricula for the extent to which the test items provide potential opportunities for students to engage in important mathematical processes/practices. The language of our Mathematical Processes Assessment Coding (MPAC) framework is closely aligned to the language of the NCTM Process Standards (NCTM, 2000), providing content validity. The reliability of our framework was established for research purposes with coder agreement ranging from 90% to 98% across the framework criteria. (See Hunsader, Thompson, & Zorin, 2013, for information about the development of the framework.)

The results of our analysis of hundreds of chapter/unit tests accompanying published curricula from a range of publishers across elementary, middle, and high school grades indicate that such tests infrequently and inconsistently provide opportunities for students to engage in important mathematical processes and practices, such as reasoning, communication, connections, and representation (Hunsader et al., 2014; Hunsader, Thompson, & Zorin, 2013). These findings raised concerns that teachers primarily using the tests accompanying their curricula might not be preparing students for some of the new high-stakes assessments, such as our own Florida State Assessment, or the assessments associated with the Partnership for Assessment of Readiness for College and Careers (PARCC, 2013) or Smarter Balanced (2013). Furthermore, consistent with findings of researchers around the world (e.g., Blok, Otter, & Roeleveld, 2002; Delandshere & Jones, 1999; Garet & Mills, 1995; Madaus et al., 1992), we had documentation showing that many of the teachers in our state use the tests accompanying their curricula and are sometimes required to do so. Hence, we felt a need to support the development of teachers' assessment knowledge relative to mathematical processes/practices by helping them become critical consumers of the tests they use with their students and by raising their awareness of the potential lack of coherence between those tests and the high-stakes tests their students are required to take.

Assessment Experience in Our Courses

Context

The assessment experience described here has been refined over multiple semesters within undergraduate and graduate courses for preservice and in-service teachers at all levels, K–12.[1] The undergraduate preservice teachers were enrolled in the first of two elementary mathematics methods courses. The first course focuses on number

1 Throughout, we refer to our university students as *teachers*, regardless of whether they are preservice or in-service. We use the term *students* to refer to their K–12 students.

and operations, including work with whole numbers, fractions, and decimals, while the second course focuses on geometry, measurement, and basic statistics. At this point in their program, most teachers have participated in their first field experience but have had few opportunities to teach a lesson and have no experience in creating or modifying a test. The graduate teachers were enrolled in a master's-level methods course on the teaching of algebraic ideas in elementary or middle grades or in a trends course looking at current issues in mathematics education; all teachers in the methods course were in-service teachers, but those in the trends course were either preservice teachers in a Master of Arts in Teaching program or in-service teachers. Typical class sizes range from 20 to 30 for the undergraduate courses and from 15 to 20 for the graduate courses. The lead author was the instructor for the undergraduate courses, and the other authors were instructors for the graduate courses.

At our institution, both undergraduate and graduate programs require a measurement course, which focuses on general issues about test development, including a focus on assessing course objectives and designing tests. Therefore, in our work with teachers we have focused on investigating the extent to which the tests include opportunities for students to engage with mathematical processes and have not focused on other aspects of test development.

Describing Our Work with Teachers

Overview of the assessment experience. The assessment experience consists of several components completed over multiple class sessions. The general structure used in the undergraduate and graduate methods courses is outlined in Table 1, together with approximate time allocations for each component; the structure for the experience in the trends course is described later. A list of potential readings is found in Appendix A; a detailed discussion of the components follows the table.

Providing a background on standards and the language of assessment (Components I, II, and III). We have found it helpful for teachers to have some background knowledge of the Standards movement, and such background is discussed early in the semester and then referenced

Table 1
Components of the Assessment Experience in Methods Courses

Time frame	Component	Undergraduate Methods (preservice teachers)	Graduate Methods (in-service teachers)
0.5–2 hours	I. History of math standards	Overview of *Standards* from NCTM (1989) through Common Core State Standards (CCSSM, 2010) with emphasis on the NCTM Process Standards and the CCS Standards for Mathematical Practice	Overview of *Standards* as needed, with focus on process standards and standards for mathematical practice
0.5–1 hour	II. Readings on assessment	Assessment chapter from methods text (outside of class)	Research or practitioner journal articles about assessment (jigsaw in class)
0.25–0.5 hour	III. Introduction to language of assessment	Review assessment *of* and *for* learning, formative and summative roles of assessment, various forms of assessment (e.g., chapter tests, projects, journals, observations)	
0.75–1 hour	IV. Introduction to our framework and coding process	Review framework codes and coding of sample items, engage teachers in whole-group coding of items, discuss possible modifications of items	
0.5–1 hour	V. Small-group analysis	Small groups code and modify sample items and present their items, codes, and suggested modifications to the class for discussion and consensus	Small groups code sample items or 1–2 complete tests, develop possible modifications; whole-class discussion of small-group work
Varies	VI. Independent assignment	Analyze a collection of sample items and provide modifications as needed	Analyze complete tests and modify items as needed

Note. In the undergraduate methods course, Component I is discussed during the first week of class and Components III, IV, and V are discussed the 3rd week. For the graduate methods course, Component I is discussed during the first week as needed, with Components II, III, IV, and V covered during the 6th week of the 15-week semester.

during the assessment experience. To introduce teachers to the language of assessment, we provide a reading assignment that varies based on teachers' backgrounds. Preservice teachers read the portion of their methods text that introduces assessment. For in-service teachers, we add readings about classroom assessment appropriate to the teachers' grade level (see Appendix A).

To help set the stage for the analysis of tests, we discuss the Common Core State Standards for Mathematics (CCSSM; 2010), specifically, the Standards for Mathematical Practice, and the standards that served as the foundation of our framework, the NCTM Process Standards (NCTM, 2000) of problem solving, reasoning and proof, communication, connections, and representation. In particular, we discuss the power of mathematical processes and practices to elicit student thinking and reveal students' level of understanding of mathematical concepts, with the intention that such thinking provides teachers with insight into student learning that can help them modify instruction.

As part of our introduction, we also discuss that assessment is both *of* and *for* student learning, and that assessments can serve both formative and summative roles depending on how the information from the assessment is used (Joyner & Muri, 2011). We discuss various forms of assessment that are typically used in classrooms, and how each form serves a different purpose. For instance, formative assessment often includes journal writing, which provides opportunities for students to explain their thinking or describe how they are feeling about their learning; observations (including facial expressions), which give in-the-moment insight into students' understanding; and bellwork, which enables misconceptions to be caught early or prerequisite knowledge to be foregrounded. Summative assessments can include tests as well as projects that provide opportunities for students to engage in a problem over an extended period of time. Although mathematics teachers are unlikely to eliminate tests in the foreseeable future, nor do we suggest such a practice, we want teachers to realize there are many opportunities for assessment beyond paper-and-pencil tests, even though tests are the main focus of our assessment experience.

Providing a tool to analyze tests objectively (Component IV). Because we had previously developed the MPAC framework for analyzing mathematical processes on tests accompanying published curricula and used it to analyze hundreds of tests, we believed it would be a useful tool for teachers to analyze tests in an objective manner. Figure 1 illustrates our framework and the criteria used to analyze assessments for Reasoning and Proof, Communication, Connections, Role of Graphics, and Translation of Representational Forms. Figure 2 includes sample items we created for the purposes of this article to illustrate distinctions among the framework codes; the items span elementary to high school content and a range of topics. The codes for the sample items are provided in Table 2. For additional sample items and explanations of coding, see Hunsader et al. (2013, 2014).

Our criteria of *Reasoning and Proof* and *Communication* are related; it is possible for an item to elicit students' mathematical communication without engaging them in reasoning and proof, but all instances of reasoning and proof involve communication. Item 1 only directs students to record a numeric answer (*Communication* code N). Two different modifications of item 1 would have resulted in an affirmative code for *Communication:* if the item did not include a graphic but asked students to draw one (graphic communication), or if the item included the graphic but no equations and asked students to write the related multiplication and division equations illustrated by the graphic (symbolic communication). Items 3 and 4 are coded as involving *Reasoning and Proof* because both require students to provide a justification for their answer. Although item 2 includes a great deal of *Communication* (code Y) in the writing of equations, students are not asked to justify their responses.

We distinguish between these two criteria as follows: *Communication* involves students' sharing *what* they are thinking; *Reasoning and Proof* goes beyond that to elicit students' rationale for why they gave that response. (We originally developed the MPAC to analyze elementary tests, for which the term *justification* seemed softer than *proof*; as we expanded our work to middle and high school grades, we kept the term *justification*, recognizing that what counts as justification becomes more rigorous across the grades. We believe this description is in line with an essential principle related to all proof: "To specify clearly the assumptions made and to provide an appropriate argument supported by valid reasoning so as to draw necessary conclusions" [Hanna et al., 2009, p. 19].)

All four items are coded as involving *Connections*. Mathematical connections can take two major forms: real-world connections (items 2 and 4), and interconnections of mathematical concepts (items 1 and 3). Item 1 relates multiplication/division and item 3 relates ideas of parallel lines, corresponding angles, and similar triangles. In our work with test-item analysis, we have rarely found items set in a real-world context that also interconnect mathematical concepts. In those cases, we coded the item as *R* but acknowledge that other users may prefer to indicate both the *R* and *I* codes.

Students use Representation in various ways to express their mathematical thinking (Diezmann & McCosker,

Opportunity for Reasoning and Proof

N	The item does not direct students to provide justification for why they gave that response.
Y	The item directs students to provide a justification for why they gave that response.

Opportunity for Mathematical Communication

N	The item does not direct students to communicate what they are thinking through symbols (beyond a numeral answer), graphics/pictures, or words.
Y	The item directs students to communicate what they are thinking through symbols, graphics/pictures, or words.
V	The item only directs students to record a vocabulary term or interpret/create a representation of vocabulary.

Connections

N	The item is not set in a real-world context and does not interconnect mathematical concepts (e.g., multiplication and repeated addition, perimeter and area).
R	The item is set in a real-world context outside of mathematics.
I	The item is not set in a real-world context but interconnects mathematical concepts.

Representation: Role of Graphics

N	No graphic (graph, picture, or table) is given or needed.
S	A graphic is given but is not needed for the response, and the graphic does not explicitly illustrate the mathematics inherent in the problem (superfluous).
R	A graphic is given and is not needed, but the graphic explicitly illustrates the mathematics inherent in the problem.
I	A graphic is given and must be interpreted to answer the question.
M	Students are asked to make a graphic or add to an existing graphic.

Representation: Translation of Representational Forms (codes are bi-directional)

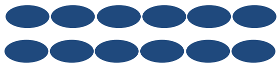
N	Students are not expected to record a translation between different representational forms of the problem.
SW	Students are expected to record a translation between verbal and symbolic representations.
GS	Students are expected to record a translation between symbolic and graphical (graphs, tables, or pictures) representations.
WG	Students are expected to record a translation between verbal and graphical representations.
TG	Students are expected to record a translation between two graphical representations.
A	Students are expected to record two or more translations among symbolic, verbal, and graphical representations of the problem.

Figure 1. MPAC framework for analyzing how test items engage students in mathematical processes/practices. Adapted from Hunsader, Thompson, & Zorin, 2013.

2011; Heritage & Niemi, 2007; Lowrie, Diezmann, & Logan, 2011). *Role of Graphics* captures how students interact with the graphics to provide a successful response (Berends & van Lieshout, 2009). When students must interpret (item 4) or create (item 3) a graphic, it is clear the item involves graphics. In contrast, some items include a graphic, but there is enough information provided in the stem of the item that a student could ignore the graphic and still respond correctly. In item 1, the counters are arranged in such a way that the mathematical relationship between 2 rows with 6 in each row and 12 counters broken into two sets is apparent. In this case, although the graphic is not needed to answer the question, it clearly represents the mathematics in the item (code *R*). If item 1 omitted the equations and asked students to create equations based on the graphic, it would require interpretation of the graphic and receive a code of *I*. If item 2 had included a picture of some birds and lizards, we would have coded it *S*.

In *Translation of Representational Forms*, we look for items that involve multiple representations by asking students to record a translation from one form of the item (symbols, words, or graphics) to a different one, a

1. Raul placed his counters in an array.



What number completes these related facts?

2 × ___ = 12

12 ÷ 2 = ___

2. Walton has 14 pet birds and lizards. Together, his pets have a total of 44 legs. How many of Walton's pets are birds? Solve using a system of equations. Record both equations, and the answer.

First equation: ____________

Second equation: ____________

Number of birds: ____________

3. Prove that the segment joining the midpoints of sides AB and AC in triangle ABC is parallel to side BC and half its length. Draw a diagram to support your work.

4. Jared, Rayvon, and Chou are the top athletes on their track and field team. The table below shows how they placed in 5 events for the last track meet.

Event	100 m	200 m	400 m	800 m	1500 m
Jared	2	1	3	4	3
Rayvon	3	3	2	3	2
Chou	1	4	4	1	4

For an upcoming meet involving sprints (100 m, 200 m, and 400 m) and medium distances (800 m and 1500 m), the school can only send two athletes. Based on these results, which two should the school send? Justify your decision.

Figure 2. Sample items to illustrate framework codes.

Table 2

Codes for Items in Figure 2 Based on the Framework in Figure 1

Item	Reasoning & Proof	Communication	Connections	Representation: Graphics	Representation: Translation
1	N	N	I	R	N
2	N	Y	R	N	SW
3	Y	Y	I	M	WG
4	Y	Y	R	I	N

critical skill in developing deep understanding of concepts (Bossé, Adu-Gyamfi, & Cheetham, 2011; Gagatsis & Shiakalli, 2004). In making this determination, it helps teachers to consider what would be involved in translating an item from English to Spanish. That translation would need to capture all the critical elements of the problem in the second language, far beyond merely providing an answer in the second language. Item 1 in its original form does not require a translation. However, if it omitted the equations and asked students to write equations to represent the graphic, or omitted the graphic and asked students to create a graphic to represent the equations, we would have recorded a code of *GS*, indicating a translation between graphic and symbolic representations. Item 2 is coded as *SW*; to complete the task, students must translate the word problem into symbolic equations. Item 3 involves translating the words of the problem into a graphic (code *WG*).

Because our assessment work began as a research project analyzing tests provided by textbook publishers without considering the written instruction in the text or the

classroom experiences of students, we did not include problem solving in the framework. A determination of whether an item might engage students in problem solving cannot be made without understanding students' background knowledge or the instruction they have experienced; however, it is clear that many procedural items (e.g., $2.38 \times 4.76 = ?$ for middle grades students) would likely provide few opportunities for problem solving regardless of classroom instruction. Although we have continued to exclude problem solving from our framework and our initial analysis of tests, we revisit this issue in the discussion.

Practice coding and modifying items (Component V). After introducing teachers to the MPAC framework and its relationship to the NCTM process standards, teachers work in a whole-class activity to create a shared understanding of the framework criteria and codes. Teachers code a selection of items or a whole test, discuss challenges and disagreements in coding, and discuss possible ways to modify items so that potential opportunities to engage with the mathematical processes/practices become evident. The author team creates the items, which are representative of items on published tests we have analyzed and involve a variety of framework codes. We discuss the risk of analyzing individual items in isolation and the goal of integrating a range of processes/practices across an entire test rather than attempting to embed all processes/practices within each item. After whole-class coding, small groups work to analyze and modify more items or tests. They then share their work with the whole class for additional discussion.

Independent assessment assignment (Component VI). As a follow-up to the class experience, teachers complete an independent assignment related to analyzing tests or test items. Undergraduate preservice teachers are given a set of 5–8 items created by the author team to model a range of framework codes across grade levels with content that varies based on the content of the course. For each item, teachers record a code for each framework criterion and a rationale for their selection and then suggest a modification to the item as needed. At the end of the assignment, they reflect on their use of the framework and how this work might impact how they formulate chapter tests in their future classrooms.

Graduate in-service teachers bring a different perspective to the experience because they are typically working with students whose abilities and challenges they understand. Such teachers have taught content units and then assessed their students' mastery of that content. Regardless of whether they have used their own tests or those provided either by a publisher or their district, these teachers have graded such tests and observed what they reveal about their students' understanding. To increase the relevance of the assignment for these teachers, we have them analyze two to five whole tests of their choosing, whether self-authored or provided by their textbook publisher or school district. We challenge teachers to reflect on their personal beliefs about the appropriate level of mathematical processes/practices across an entire test. We do not suggest a target level; that decision must rest with a classroom teacher who knows her/his students and what other assessment measures are used. However, we require teachers to modify at least two items strategically so that the processes/practices become more evident.

Evaluation of the independent assignment. The emphasis placed on the independent assignment for a course grade is typically about 10% in the undergraduate and graduate methods courses based on the amount of class time dedicated to assessment and the overall expectations for the assignment. Later in this article, we share an adaptation/extension of the assignment for a course in which more time is allocated to assessment.

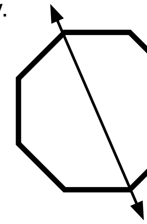
Our grading criteria are not focused on whether teachers apply our framework correctly or incorrectly but rather on their rationale for their analysis and modification of items and their reflection on their own assessment practices. Such a grading decision is consistent with our intent to enhance teachers' ability to be critical consumers of their tests, rather than suggest that there is a perfect test or a predetermined ratio of processes/practices that should be present in tests.

Figure 3 shows two items that preservice teachers analyzed, along with their codes and rationales. We evaluated the code and rationale holistically, and the examples are ones that would have received partial credit; either a code different from what we assigned is given but with a reasonable rationale, or a code consistent with ours is given but the rationale is lacking in some respect.

Teacher responses A and C include a code that is different from how we coded the item (V and R, respectively), but the rationales demonstrate good thinking about how students might engage with the items. For item 1, Teacher A understood that students were not expected to explain their thinking but missed that the item asks students to interpret a representation of vocabulary. For item 2, Teacher C makes an incorrect assumption that students are expected to complete the graphic but chooses the correct code based on that assumption.

Teachers B and D coded the item as we expected, but their explanations do not fully communicate a rationale for the code. Teacher B's rationale incorrectly implies that the student has to write a response including vocabulary.

Item 1: Tell if the line is a line of symmetry.



Framework Criterion: Communication

Teacher A: Code N. No communication is represented in this question. Students do not have to show their thought processes, nor does it engage students in the use of mathematical language.

Teacher B: Code V. The test item directs students to communicate their mathematical thinking solely through vocabulary.

Item 2: Samantha has 40 dolls. She puts them in 5 equal groups. How many dolls are in each group?

40 dolls in all				
?	?	?	?	?

Framework Criterion: Role of Graphics

Teacher C: Code M. A partial graphic is given and the item directs students to add to the existing graphic. Students fill in the blanks where there are question marks.

Teacher D: Code R. A graph is given, and it would be helpful for a student to use it. There is no direction on how to use it, or that they need to use it to show work.

Figure 3. Sample test items analyzed by preservice teachers with their codes and rationales.

Teacher D acknowledges that the graphic may support students but does not communicate awareness that the graphic completely illustrates the mathematics in the problem.

Teacher Reflections on the Assessment Experience and the Framework

We continue to be interested in teachers' perspectives on the assessment experience embedded within our methods courses so we can improve the overall experience, maximize its potential to impact teachers and their students, and ensure that the time spent is perceived as worthwhile by the teachers. Thus, after implementing the assessment experience in several course sections, we began a formal research study to collect data about teachers' perspectives on the use of the framework and their perceptions about how the analysis of tests/test items might influence their future classroom assessment practices. We report data only from those sections in which teachers were completing the assessment experience for the first time, provided informed consent for the use of their work, and for which we have all their work electronically.

Reflections from undergraduate preservice teachers. Preservice teachers responded to the following questions:

- To what extent did the framework impact your ability to analyze whether a test item gave students an opportunity to engage in the mathematical processes you evaluated?
- How might these assessment experiences impact your formulation of chapter tests in mathematics when you are a teacher?

Preservice teachers' responses were analyzed according to grounded theory and emergent design (Glaser & Strauss, 1999; Patton, 2002). Responses to the two questions from 23 preservice teachers in the elementary methods course were coded and discussed by the authors using a consensus model (Charalambous, Delaney, Hsu, & Mesa, 2010; Ding & Li, 2010; Roseman, Stern, & Koppal, 2010) to identify segments that seemed to indicate a coherent thought. We identified and coded 72 thought segments among the responses and then collapsed related codes into six themes, labeled as indicated

below, along with the percentage of teachers who had at least one thought segment that fell into that theme.

A: Comments about framework criteria and what teachers need to use the framework effectively (83% of teachers)

B: What the framework affords the user (57% of teachers)

C: Quality of framework criteria (57% of teachers)

D: Value of analysis for modification of items and building of quality tests (13% of teachers)

E: Comments on the overall assignment (17% of teachers)

F: Other (9% of teachers)

Table 3 illustrates one teacher's response to the two questions, the thought segments we coded, and the theme in which that code was eventually placed.

We discuss teachers' reflections for themes A–D, leaving discussion of theme E to a later section of this article.

Theme A: *Comments about framework criteria and what teachers need to use the framework effectively.* Eighty-three percent of the teachers made comments about some aspect of the framework criteria and supports that would help in using it. Some reflections detailed specific

Table 3

A Preservice Teacher's Response to the Two Questions With Codes and Themes Identified

Original Response	Segment of Thought	Code	Theme
I found the way that the MPAC framework was set up was helpful. It was organized well so that I could assess the problems easily. Under each term, like Reasoning and Proof, it said Justification, which gave me a quick clue to make the assessment accessible. I have never done any sort of assessments on student problems… I really liked having to try to correct the problems (trying to make them better). I felt that it helped me make more sense of what was wrong with the problem in the first place. (T215 response to question 1)	I found the way that the MPAC framework was set up was helpful. It was organized well so that I could assess the problems easily.	Organized	B: Framework affords
	Under each term, like Reasoning and Proof, it said Justification, which gave me a quick clue to make the assessment accessible. I have never done any sort of assessments on student problems…	Clear criteria	C: Quality of criteria
	I really liked having to try to correct the problems (trying to make them better). I felt that it helped me make more sense of what was wrong with the problem in the first place.	Value of modifications	D: Aids in modifications and building of quality tests
The only thing I found challenging about the MPAC-5 framework was that I was second-guessing myself on certain codes. Some items seemed to be able to be one code or the other (it seemed to be both!). At first I had a hard time figuring out if a problem had translation or asked the student to translate and then when I read more about translation I was more comfortable coding. (T215 response to question 2)	The only thing I found challenging about the MPAC-5 framework was that I was second-guessing myself on certain codes. Some items seemed to be able to be one code or the other (it seemed to be both!). At first I had a hard time figuring out if a problem had translation or asked the student to translate and then when I read more about translation I was more comfortable coding.	Struggle with criteria	A: Comments about criteria and what teachers need to use the framework effectively

criteria teachers found challenging (e.g., distinguishing between Reasoning and Proof and Communication); others provided suggestions that might support future work with the framework or acknowledged that analyzing assessment items was an entirely new experience. Typical responses include:

> **"I found that it may be helpful to provide a sample question to each of the coding frameworks [criteria] to help . . . differentiate between them." (T221)**

> **"I also found it difficult to complete this assignment because I have never thought about nor had [to] practice assess [sic] test questions." (T226)**

Theme B: *What the framework affords the user.* Fifty-seven percent of the teachers made comments that we interpreted to indicate that they were thinking beyond the framework criteria to what the analysis reveals about the items and their ability to reveal student thinking, as suggested in the following representative comment:

> **"The [framework] made it very obvious that all of the questions required little to no thinking . . . If there was thinking required, students did not need to show it. . . . It is easier to see a mistake when a student's work is shown." (T211)**

Theme C: *Quality of criteria.* Fifty-seven percent of the teachers made comments that focused on how the structure and wording of the framework criteria helped in their analysis, while others noted a desire to internalize the framework criteria to guide their future assessment practices. The following comment suggests how one teacher considered turning the criteria into questions to ask herself when developing tests:

> **"The codes served as the questions that we (should) innately ask ourselves when sorting through these types of questions. They served as question prompts for those of us that are still digesting the information and trying to make sense of it." (T205)**

Theme D: *Analysis aids modification of items and building of quality tests.* Thirteen percent of teachers made comments that suggested a growing awareness of the complexity of crafting quality assessments, as reflected in the following comment:

> **"I never realized that so much went into writing a test question." (T226)**

Reflections from graduate in-service teachers. The in-service teachers in the graduate methods course reflected on the following three questions, with the first and third questions identical to those asked of the preservice teachers.

- To what extent did the framework impact your ability to analyze whether a test item gave students an opportunity to engage in the mathematical processes you evaluated?
- After using the framework to analyze tests, what did you notice and what conclusions did you draw about the tests you analyzed?
- How might these assessment experiences impact your formulation of chapter tests in the future?

Responses from the 19 graduate teachers were coded into 83 thought segments that were collapsed into the following themes using the same process we used for the preservice teachers' responses.

G: Concern about test quality (100% of teachers)

H: Impact of assessment experiences on future assessment practices (84% of teachers)

I: Framework facilitates analysis of tests (47% of teachers)

J: Other (11% of teachers)

Again, we briefly discuss responses within these themes.

Theme G: *Concern about test quality.* Although these teachers have experience assessing their own students, for some this was their first experience analyzing the tests they used. All teachers made some comment that fell within this theme, with most indicating some level of disappointment with their tests after the analysis. The following comment suggests this teacher realized that her tests did not provide opportunities to elicit student thinking:

> **"I noticed that the items that I analyzed did not aptly allow the students to reason or justify their thinking. . . . I feel like that does the student a real disservice because I can't truly interpret the data and drive my instruction if I don't know the reasoning or provide my students with the opportunity to justify their thinking." (T606)**

Theme H: *Impact of assessment experience on future assessment practices.* The focus on test analysis was designed to heighten teachers' awareness of the importance of including processes and practices within test items and to impact their future assessment practice. Reflections

from 84% of the teachers appear to indicate they recognize the potential negative impact on students' learning when they limit assessment to the tests they currently use. In contrast to the work cited by Black and Wiliam (2010), these typical responses indicate an awareness of the need to alter future practices and a willingness to engage in the challenging work required to effect change:

> **"In the future, prior to administering a math test with students, an analysis of the test questions will be conducted using the . . . framework . . . to ensure that the mathematics concept and processes are being evaluated adequately." (T617)**

> **"The coding and reviewing of assessments has given me a lot to think about. Do I even bother looking at publisher based assessment or do I develop my own from scratch every time? The answer is easy, develop my tests from now on . . ." (T612)**

Theme I: *Framework facilitates analysis of tests.* This theme pertained to 47% of the teachers' responses, suggesting that they believed the framework facilitated their analysis of the extent to which their tests engaged students in important mathematical processes/practices. The following comment is typical of the responses:

> **"Using the . . . framework helped me to evaluate whether or not test items aligned to the depth that is expected with Common Core." (T604)**

Teachers' reactions to the overall assessment experience. One question for us has been whether the time and effort required to build teachers' assessment knowledge relative to mathematical processes has the potential to impact their students' learning and whether it is worth the investment of time in the courses. Responses from both preservice and in-service teachers seem to suggest that the experience has raised their awareness of the need to look at items objectively to ensure they provide opportunities for students to engage with the processes. However, responses from the in-service teachers suggested deeper thought, likely because they were reflecting on current practices and not just on future practices. Responses such as the following suggest to us that the experience is likely worth the time investment, although checking teachers' actual classroom test practices was beyond the scope of the course or our related research endeavors.

> **"I think that this process has benefited me greatly in that as I analyzed each question I found myself referring back to the NCTM Process Standards for clarification and I realized very quickly that I was hindering my little problem solvers. I am on the math committee for 1st grade and we plan the unit but I create the assessments and although I use a CCSS database to draw from I elect not to use essay questions and opt for fill in the blank and multiple choice questions not realizing that I have eliminated the opportunity for the students to reason." (T606)**

> **"As I create my own tests or rework the tests from the textbook company, my new motto will be 'Go below the surface, make them dig deeper.'" (T615)**

> **"These assessment experiences confirm why I make a majority of my test and quizzes. My students are given an opportunity to show a representation of different methods and reasoning to explain work and find the answers!" (T618)**

Adapting and Extending the Assessment Experience

In contrast to the undergraduate/graduate methods courses in which we use the previously described assessment experience, the graduate Current Trends course allows for more time to be spent on assessment, integrating the previously described experience with further investigations of national and international assessments (e.g., National Assessment of Educational Progress [NAEP], Trends in International Mathematics and Science Study [TIMSS], and Programme for International Student Assessment [PISA]). Table 4 highlights the topics for the first 8 weeks of the course, illustrating how assessment is related to issues of curriculum reform; readings for the week related to assessment are included in Appendix A.

Teachers' Test Analysis and Modifications

To gauge the effectiveness of the framework to guide teachers' analysis and modification of their own tests, we reviewed the responses to the assignment in Figure 4 from 14 teachers who provided consent. Again, our focus was not on right/wrong coding but on their overall analysis and reflection about their findings. Teachers' work on the assignment generally reflected appropriate coding according to the framework and item modifications that were targeted to enhance the focus on processes the teachers believed were underrepresented on each test. Per step 2 of the guidelines, teachers summarized their test coding results. Although they analyzed tests of their own choosing, 13 of the 14 teachers generally found results similar to findings from our previous research using the framework (Hunsader et al., 2013, 2014). That is, tests accompanying published curricula often provide limited opportunities for students to engage with mathematical processes, with

Table 4
Topics Related to Curriculum Reform and Assessment in the Current Trends Course

Week	Topic
1	Current issues in mathematics education
2	Historical perspectives on curriculum up to the era of NCTM Standards
3	Recent national recommendations and reports, from the 1989 NCTM Standards to the 2000 Principles and Standards
4	Common Core State Standards and relation to Florida standards
5	***Classroom assessments with a focus on the NCTM mathematical process standards, introduction to MPAC framework, and group and whole-class coding of sample tests***
6	***Focus on the mathematical practices***
7	***National and international assessments, including NAEP, TIMSS, and PISA***
8	***PARCC and Smarter Balanced assessments and relation to Florida State Assessments***

Note. Bold and italics are used to indicate weeks with a particular focus on assessment.
Because 4 weeks of the course are spent on assessment issues, the independent assignment accounts for about 30% of the overall course grade. Figure 4 provides the assignment guidelines along with the scoring rubric.

reasoning and proof typically the least visible process. The visibility of the other processes vary depending on the content focus of the test. The following excerpts are typical of teachers' responses:

> **"I was able to get insight on what aspects of student understanding that can or can not be learned from the tests I analyzed. . . . Reasoning and Proof was completely absent from 3 out of 5 test [sic], and added up to less than 10% of the two tests that it was in. This is a hard gap to ignore. Reasoning and proof are arguably the heart of mathematics, but it is gone in assessment practices; likely because of time constraints, yet multiple questions without mathematical processes can't be the better option. Also intriguing is the absence of mathematical communication. Assessment should include what students are thinking . . ." (T710)**

> **"The tests analyzed showed that the mathematical processes of reasoning and proof are not integrated into assessment. The grade 6 tests should use these processes because students at this level should begin to prepare for the use of formal proofs in their mathematics. The tests provided little opportunity for mathematical communication which does not align with the mathematical standards. Connections to the real world were used but mostly with superficial examples that does [sic] not truly show the importance of mathematics in a real world context. Overall there were a lot of graphics used but not many needed interpretation to solve the problem. Many of the graphics used were just tables that contained the information of the problem. This was just a way to represent the information but it was not necessary to interpret the graphic. Furthermore, translation of representational forms was rarely seen in these assessments." (T709)**

Step 3 of the assignment challenges teachers to use their analysis to suggest modifications for 10 test items, then describe the focus of the modifications and how the modifications would change their results from step 2. Teachers had considerable latitude in determining the focus of their modifications. As the following excerpt indicates, some chose to modify items across a range of processes:

> **"I tried to include all the processes in different test items that I modified. Introducing more of these processes in the assessments would allow students to present or illustrate what they are thinking and why they thought their responses were correct, in lieu of one final answer. Showing the process of solution communicates to the teacher where the student may have difficulty with a particular topic or where the student may have made a small error that would affect their final answer. Including more real-world connections in the test items would make the items more relatable to students, while allowing them to think analytically: what is the item asking of me, what would I do if I were in this situation?" (T703)**

Analyze 5 tests that cover at least three different mathematics content areas. You may use the tests that accompany your curriculum materials, tests you construct yourself, or district wide assessments.

1. Focus on the extent to which the tests address the mathematical processes of reasoning and proof, communication, connections, and representations (role of graphics and translation among representations). Use the Framework shared in class. Share your coding in some type of template. Include a copy of the original tests with your work.
2. Summarize your results. You will likely want to use the test as the unit of analysis. Prepare a short narrative paper that describes the extent to which the processes are evident in these assessments and give insight into students' understanding. Support your results with references to your detailed results from #1. It would be appropriate to include a brief table with some averages across tests. Keep this summary to 3–5 pages.
3. Suggest possible modifications of items to enable you to assess mathematical processes that seem to be assessed minimally on the tests you analyzed. Describe the focus of your modifications and how use of the modified items would change the results in #2. Provide modifications for at least 10 items, with modifications designed to get at different processes. Identify any processes still not adequately addressed and how you will attend to them.
4. Find at least 10 items from NAEP, TIMSS, or PISA that address the content topics on your tests and have the potential to engage students with mathematical processes. Compare and contrast these items with the original items on your tests and/or your modified items.
5. Briefly (2 pages max) describe how your investigation provides insights into alignment or disconnects with your classroom assessment and the national assessments related to Common Core (e.g., PARCC, Smarter Balanced).

Grading Rubric

A All items are done at an exceptional level. Items are coded according to the framework and shared in an understandable manner. The summary from #2 is clear, well written, and coherent, with appropriate evidence to support claims. Reasonable modifications are made that enable more of the processes to become evident. Selections from NAEP, TIMSS, or PISA are appropriate, with appropriate comparisons and contrasts. The discussion related to national assessments is appropriate with excellent insights.

B All items are done at an above average level. Items are coded according to the framework and shared in an understandable manner. The summary from #2 is clear, well written, and coherent, with appropriate evidence to support claims. Reasonable modifications are made that enable more of the processes to become evident. Selections from NAEP, TIMSS, or PISA are good, with appropriate comparisons and contrasts. The discussion related to national assessments is appropriate with good insights.

C All items are done at an average level. Items are coded according to the framework and shared in an understandable manner. The summary from #2 is acceptable, with some minor grammar errors or incoherent statements, with evidence to support most claims. Reasonable modifications are made that enable more of the processes to become evident. Selections from NAEP, TIMSS, or PISA are good, with appropriate comparisons and contrasts. The discussion related to national assessments is appropriate with average insights.

Figure 4. Guidelines for extended assessment assignment in graduate Current Trends course.

As the following excerpts indicate, some teachers focused on modifications related to a particular process they perceived as neglected:

> **"The process that seemed to be the most neglected in items for the five tests analyzed is Reasoning and Proof. Since this process was lacking the most compared to the other process, the modifications of the assessment items will attempt to correct the minimal evidence of assessment of Reasoning and Proof. A possible way to modify the items in order to address the disparity of this process is to require students to explain their response." (T707)**

> **"Most of the unit assessments are lacking in the area of student reasoning and proof, and communication. These tasks are critical as they allow for a greater insight into student understanding." (T714)**

In step 5 of the assignment, teachers consider the results of their investigation and discuss the relationship between their classroom tests and the expectations of national assessments related to the Common Core, even though Florida's assessments are not associated with PARCC or Smarter Balanced. All teachers indicated the assessment experience enhanced their awareness of the need to be more intentional about their classroom assessment

decisions; they recognized the need to ensure that processes reflected on high-stakes assessments, either state or national, are integrated into their classroom assessments. The following excerpts are representative of teachers' comments:

> **"According to the PARCC website, the items on the PARCC assessment are classified into three separate tasks [sic] types. Type 1 assesses understanding and application, type 2 requires students to justify their work and show reasoning, and type 3 has items written in real-world context (PARCC). These task types are aligned with the items that were being investigated in this assignment and should therefore be adhered to when constructing assessments of my own." (T703)**

> **"This investigation provided me with awareness that reliance on the assessments that accompany textbooks may not provide adequate insight into a student's ability to display the mathematical processes outlined by NCTM Process Standards, and CCSS Standards for Mathematical Process. . . . With the new national assessments related to Common Core (e.g., PARCC, Smarter Balanced, . . .), students will be required to show competency in the process standards." (T707)**

In fact, one teacher developed her own test by modifying items from her curriculum materials to address more of the processes. She noted:

> **"By doing so I created a test that was more diverse and had a variety of mathematical processes. The students overall did better compared to previous chapter tests and I, as a teacher was able to assess them more accurately." (T702)**

In summary, teachers' work on the assignment suggests they found limited evidence of opportunities for students to engage with the processes and practices in the tests they analyzed. Using the results of their analysis, teachers demonstrated their ability to make targeted modifications in an effort to better balance students' engagement in the processes within a single classroom test. Finally, their responses to step 5 suggest an awareness of potential disconnects between the inclusion of processes and practices on the tests they analyzed and what is expected of students in new state/national assessments. We believe our goal, to enhance teachers' ability to be critical consumers of assessments, was realized in this case.

Discussion and Limitations

Our intention for the assessment experience was to provide teachers with an objective means to analyze classroom tests in relation to opportunities for students to engage with the mathematical processes, which we believe are at the heart of mathematics and of the Common Core, and to help them become critical consumers of the tests they often use as part of their assessment practice. Teachers' reflections have corroborated our belief that teachers with all levels of experience need an awareness check, particularly with respect to the mathematical processes and practices embedded in tests. The teachers in our courses perceived the benefit of this experience for their future practice, whether we allocated a relatively small proportion of class time (1–2 weeks) or a longer period of time (up to 4 weeks) to assessment issues. Based on our prior work in analyzing tests across publishers and grade levels (Hunsader et al., 2013, 2014), we were not surprised by how infrequently and inconsistently the tests teachers analyzed engaged students in mathematical processes/practices. However, many teachers were shocked to realize that the tests they have been using miss many opportunities to elicit the kind of mathematical thinking they value and want to assess.

We continue to refine and adjust the assessment experience to ensure that it achieves our stated purpose. For instance, we have recently realized that we should make some changes that we believe would enhance the experience for preservice undergraduate teachers. We initially provided a single set of items for these teachers to both analyze and modify. However, we found that they struggled to modify items we had developed that already incorporated multiple processes/practices, contributing to a misconception that every item should engage students in every process. So, in the future, we plan to provide separate sets of items for analysis and for modification. Items provided for analysis will include a range of opportunities to engage students in mathematical processes/practices; in contrast, items provided for modification will reflect important content but be written with limited opportunities to engage in the processes/practices.

We have also learned that when having teachers strategically modify test items, it is helpful to focus their attention on the strengths of each item. Building on the strengths of existing test items is more efficient and likely more effective than creating new items from scratch or attempting to improve weak ones. We have also found that teachers benefit from actually answering both the original test items and their modified versions. Answering an original item helps teachers focus on what the item is actually

asking of students rather than what a first glance implies, particularly for items that have multiple correct answers. Answering modified versions helps teachers consider whether the complexity and length of response required by their modifications are realistic within the context of a unit test and/or for the intended grade level. For guidance on the process of modifying existing items, see Hunsader, Thompson, and Zorin (2014).

One limitation of our framework is that it does not include a criterion for problem solving. As indicated by NCTM (1989), true problem solving engages students in tasks different from typical exercises found in a textbook or as part of classroom instruction. Because preservice teachers typically have limited knowledge of the types of problems to which students have been exposed, it can be difficult for them to determine whether a task is simply an exercise or involves more non-routine solution approaches. However, in-service teachers are in a better position to determine whether a task is similar to ones their students have previously seen or solved. Hence, in our future work with in-service teachers we plan to add a criterion for problem solving to differentiate simple exercises from non-routine problems that cannot be solved by applying a procedure, formula, or algorithm known to the student.

Direction for the Future

Now, more than ever, teachers need to be equipped to analyze classroom tests for how items engage students in mathematical processes and practices (CCSSO, 2010; NCTM, 2000). McGehee & Griffith (2001) analyzed assessment trends across many U. S. states and found that high-stakes tests act as agents of change in curriculum design and classroom assessments. In this specific case, high-stakes tests have the potential to make a positive difference in student learning by focusing classroom instruction and assessment on important mathematical processes and practices. The two major consortia developing high-stakes assessments for the CCSSM, namely PARCC and Smarter Balanced, have woven mathematical processes and practices into their task-development guidelines. Students need opportunities to engage regularly in the kind of thinking espoused in the NCTM Process Standards and the CCSS Standards for Mathematical Practice during both instruction and assessment. Although teachers have no control over the content or processes/practices included in high-stakes tests, they can control the nature of the tests they use in their classrooms. We believe the assessment experiences outlined in this article will support teachers' ability to analyze and modify their classroom tests. Tests that include important processes and practices support the kinds of thinking that should prepare students for success in the classroom and in their future personal and professional lives.

Whether using our framework or a different set of criteria as the basis of test analysis, we believe there is a need to increase the level of attention in teacher preparation and professional development to support teachers' ability to be critical consumers of their classroom tests. Providing an objective tool to evaluate classroom tests and the support necessary to analyze and modify test items is a means to build teachers' assessment skills. Classroom teachers, mathematics teacher educators, math coaches, and those who design and lead professional development experiences for teachers could use the structure of our framework and edit, add, or substitute criteria to suit their learning and assessment goals and reflect the mathematical processes and practices they value. Such knowledge could be integrated with other issues in test development (e.g., content alignment, learning objectives, the writing of multiple-choice items) that are often a part of more general measurement and assessment courses/experiences.

Our framework, with its focus on NCTM's mathematical processes, was developed prior to the release of the Common Core Standards and its focus on mathematical practices. We do not believe our framework is incompatible with the Standards for Mathematical Practice. A comparative analysis of these practices and our framework, using the work of Koestler, Felton, Bieda, and Otten (2013) as a guide, reveals that elements of each of the practices align with one or more of our framework criteria (see **Appendix B**). We believe that our criteria allow for the analysis of mathematical processes that most educators would agree are important for students' conceptual development. We believe the complex nature of the practices and the fact that they are much less discrete than the NCTM Process Standards make it more difficult to analyze written tests for opportunities to engage with the practices without some knowledge of the textbook's instruction, classroom instruction, or students' responses. Analyzing based on the processes still provides insight into potential to engage with the mathematical practices.

Teachers' reflections indicate a desire to integrate what they have learned from the experience into their own assessment practices. Future work might seek to gather longitudinal data to determine the extent to which these experiences effect real and long-term change in teachers' assessment practices.

Some of the in-service teachers plan to share the framework with their colleagues and suggest working with school-based leaders to enact assessment experiences in their schools similar to what they experienced in our courses. Although our work has been confined to our

university courses, we believe the in-class activities and independent assignments we designed could be adapted for use in professional development workshops for teachers of mathematics across grade levels, as suggested by these teachers' plans. We look forward to continuing the assessment conversation with mathematics teacher educators, practitioners, coaches, and researchers.

References

Alkharusi, H., Aldhafri, S., Alnabhani, H., & Alkalbani, M. (2012). Educational assessment attitudes, competence, knowledge, and practices: An exploratory study of Muscat teachers in the Sultanate of Oman. *Journal of Education and Learning*, *1*(2), 217–232.

Berends, I. E., & van Lieshout, E. C. (2009). The effect of illustrations in arithmetic problem-solving: Effects of increased cognitive load. *Learning and Instruction*, *19*(4), 345–353.

Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *92*(1), 81–90.

Blok, H., Otter, M. E., & Roeleveld, J. (2002). Coping with conflicting demands: Student assessment in Dutch primary schools. *Studies in Educational Evaluation*, *28*, 177–188.

Borko, H., Mayfield, V., Marion, S., Flexer, R., & Cumbo, K. (1997). Teachers' developing ideas and practices about mathematics performance assessment: Successes, stumbling blocks, and implications for professional development. *Teaching and Teacher Education*, *13*(3), 259–278.

Bossé, M. J., Adu-Gyamfi, K., & Cheetham, M. (June 15, 2011). Translations among representations: Teacher beliefs and practices. *International Journal for Mathematics Teaching and Learning*, 1–23. http://www.cimt.plymouth.ac.uk/journal/default.htm

Boston, M. D., & Smith, M. S. (2009). Transforming secondary mathematics teaching: Increasing the cognitive demands of instructional tasks used in teachers' classrooms. *Journal for Research in Mathematics Education*, *40*(2), 119–156.

Charalambous, C. Y., Delaney, S., Hsu, H.-Y., & Mesa, V. (2010). A comparative analysis of the addition and subtraction of fractions in textbooks from three countries. *Mathematical Thinking and Learning*, *12*, 117–151.

Chelsey, G. M., & Jordan, J. (2012). What's missing from teacher prep. *Educational Leadership*, *69*(8), 41–45.

Delandshere, G., & Jones, J. (1999). Elementary teachers' beliefs about assessment in mathematics: A case of assessment paralysis. *Journal of Curriculum and Supervision*, *14*, 216–240.

DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, *17*(4), 419–438.

DeWitt, S. W., & Freie, C. (2005). Focusing methods on goals: Problems and possibilities. *The Social Studies*, *6*, 241–245.

Diezmann, C. M., & McCosker, N. T. (2011). Reading students' representations. *Teaching Children Mathematics*, *18*(3), 162–169.

Ding, M., & Li, X. (2010). A comparative analysis of the distributive property in U.S. and Chinese elementary mathematics textbooks. *Cognition and Instruction*, *28*, 146–180.

Fan, Y., Wang, T., & Wang, K. (2011). A web-based model for developing assessment literacy of secondary in-service teachers. *Computers & Education*, *57*, 1727–1740.

Gagatsis, A., & Shiakalli, M. (2004). Ability to translate from one representation of the concept of function to another and mathematical problem solving. *Educational Psychology: An International Journal of Experimental Educational Psychology*, *24*(5), 645–657.

Garet, M. S., & Mills, V. L. (1995). Changes in teaching practices: The effects of the Curriculum and Evaluation Standards. *Mathematics Teacher*, *88*, 380–389.

Glaser, B. G., & Strauss, A. L. (1999). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine Publishing.

Hanna, G., de Villiers, M., Arzarello, M., Dreyfus, T., Durand-Guerrier, V., Jahnke, H. N., . . . Yevdokimov, O. (2009). Proof and proving in mathematics education. Discussion document. In F.-L. Lin, F.-J. Hsieh, G. Hanna, & M. de Villiers (Eds.), *Proceedings of the ICMI 19 Study 19 Conference: Proof and proving in mathematics education* (Vol. 1, pp. 16–30). Taipei, Taiwan: National Taiwan Normal University.

Heritage, M., & Niemi, D. (2007). Toward a framework for using student mathematical representations as formative assessment. *Educational Assessment*, *11*(3–4), 265–282.

Hunsader, P. D., Thompson, D. R., & Zorin, B. (May 14, 2013). Engaging elementary students with mathematical processes during assessment: What opportunities exist in tests accompanying published curricula? *International Journal for Mathematics Teaching and Learning*, 1–25. http://www.cimt.plymouth.ac.uk/journal/

Hunsader, P. D., Thompson, D. R., & Zorin, B. (2014). Mathematical practices: Small changes in assessments = big benefits. In K. Karp & A. Roth McDuffie (Eds.), *Annual perspectives in mathematics education: Using research to improve instruction* (pp. 205–214). Reston, VA: National Council of Teachers of Mathematics.

Hunsader, P. D., Thompson, D. R., Zorin, B., Mohn, A. L., Zakrzewski, J., Karadeniz, I., Fisher, E. C., & MacDonald, G. (2014). Assessments accompanying published textbooks: The extent to which mathematical processes are evident. *ZDM: The International Journal on Mathematics Education*, *46*(5), 797–813. doi:10.1007/s11858-014-0570-6

Joyner, J. M., & Muri, M. (2011). *INFORMative assessment: Formative assessment to improve math achievement, grades K–6*. Sausalito, CA: Math Solutions.

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. Washington, D.C.: National Academies Press.

Koestler, C., Felton, M. D., Bieda, K. N., & Otten, S. (2013). *Connecting the NCTM Process Standards & the CCSSM Practices*. Reston, VA: National Council of Teachers of Mathematics.

Koh, K. H. (2011). Improving teachers' assessment literacy through professional development. *Teaching Education*, *22*, 255–276.

Lowrie, T., Diezmann, C. M., & Logan, T. (September 21, 2011). Understanding graphicacy: Students' making sense of graphics in mathematics assessment tasks. *International Journal for Mathematics Teaching and Learning*, 1–21. http://www.cimt.plymouth.ac.uk/journal/default.htm

Lukin, L. E., Bandalos, D. L., Eckhout, T. J., & Mickelson, K. (2004). Facilitating the development of assessment literacy. *Educational Measurement: Issues and Practice*, *23*, 26–32.

Ma, L. (1999). *Knowing and teaching elementary mathematics*. Mahwah, NJ: Lawrence Erlbaum.

Madaus, G. F., West, M. M., Harmon, M. C., Lomax. R. G., & Viator, K. A. (1992). *The influence of testing on teaching math and science in grades 4–12, executive summary*. Chestnut Hill: Boston College, Center for the Study of Testing, Evaluation, and Educational Policy.

McGehee, J. J., & Griffith, L. K. (2001). Large-scale assessments combined with curriculum alignment: Agents of change. *Theory into Practice*, *40*(2), 137–144.

McMillan, J. H., Myran, S., & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *The Journal of Educational Research*, *95*(4), 203–213.

Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools*, *12*, 101–113.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation 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. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics: A quest for coherence*. Reston, VA: Author.

National Governors Association Center for Best Practices (NGA Center) and Council of Chief State School Officers (CCSSO). *Common Core State Standards for Mathematics*. Washington, D.C.: NGA Center and CCSSO, 2010. http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf.

National Research Council. (1989). *Everybody counts: A report to the nation on the future of mathematics education*. Washington, D.C.: National Academies Press.

Partnership for Assessment of Readiness for College and Careers. (2013). "Item Development." http://www.parcconline.org/assessment-development (accessed July 24, 2013).

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.

Quilter, S. M., & Gallini, J. K. (2000). Teachers' assessment literacy and attitudes. *The Teacher Educator, 36*, 115–131.

Romagnano, L. (2001). The myth of objectivity in mathematics assessment. *Mathematics Teacher*, *94*, 31–37.

Roseman, J. E., Stern, L., & Koppal, M. (2010). A method for analyzing the coherence of high school biology textbooks. *Journal of Research in Science Teaching*, *47*, 47–70.

Schoenfeld, A. H. (2002). Making mathematics work for all children: Issues of standards, testing, and equity. *Educational Researcher*, *31*(1), 13–25.

Senk, S. L., Beckmann, C. E., & Thompson, D. R. (1997). Assessment and grading in high school mathematics. *Journal for Research in Mathematics Education*, *28*, 187–215.

Smarter Balanced Assessment Consortium. (2013). "Smarter Balanced Assessments." http://www.smarterbalanced.org/smarter-balanced-assessments/ (accessed July 24, 2013).

Smith, M. S., & Stein, M. K. (1998). Selecting and creating mathematical tasks: From research to practice. *Mathematics Teaching in the Middle School*, *3*, 344–350.

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment FOR learning. *The Phi Delta Kappan*, *83*(10), 758–765.

Taylor, L. T. (1992). Survey of algebra and geometry teachers' assessment practices, and an investigation of publisher-provided testing materials. *Dissertation Abstracts International*, *52*(11). (University Microfilms No. AAC9211232)

Thompson, D. R., Beckmann, C. E., & Senk, S. L. (1997). Improving classroom tests as a means of improving assessment. *Mathematics Teacher*, *90*, 58–64.

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. 1053–1098). Charlotte, NC: Information Age Publishing.

Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education*, *16*, 323–342.

Zientek, L. R. (2007). Preparing high-quality teachers: Views from the classroom. *American Educational Research Journal*, *44*(4), 959–1001.

Authors

Patricia D. Hunsader, College of Education, University of South Florida Sarasota-Manatee, 8350 N. Tamiami Trail, Sarasota, FL 34243; hunsader@sar.usf.edu

Barbara Zorin, Department of Mathematics, 6650 5th Avenue N., St. Petersburg College, St. Petersburg, FL 33733; Drbzorin@gmail.com

Denisse R. Thompson, College of Education, University of South Florida, 4202 E. Fowler Ave., Tampa, FL 33620; denisse@usf.edu

Appendix A: Course Readings Related to Assessment

Undergraduate Methods Course (Preservice Elementary Teachers)

Text:

Van de Walle, J. A., Karp, K. S., & Bay-Williams, J. M. (2010). *Elementary and middle school mathematics: Teaching developmentally* (7th ed.). Boston, MA: Allyn & Bacon.

Readings:

Chapter 5 of Van de Walle text

Graduate Methods Course (In-Service Elementary or Middle Grades Teachers)

Texts:

Cuevas, G., J., & Yeatts, K. (2001). *Navigating through algebra in grades 3–5.* Reston, VA: National Council of Teachers of Mathematics. (elementary teachers)

Friel, S., Rachlin, S., & Doyle, D. (2005). *Navigating through algebra in grades 6–8.* Reston, VA: National Council of Teachers of Mathematics. (middle grades teachers)

Stump, S., Roebuck, K., & Bishop, J. (2009). *Algebra for elementary and middle school teachers: An inquiry approach.* (2nd ed.) New York: Pearson Learning Solutions. (all teachers)

Readings:
(completed in class using a jigsaw strategy)

For elementary teachers:

Barlow, A. T. (2010) Building word problems: What does it take. *Teaching Children Mathematics*, *17*(3), 140–148.

Breyfogle, M. L., & Williams, L. E. (Dec. 2008/Jan. 2009). Designing and implementing worthwhile tasks. *Teaching Children Mathematics*, *15*(5), 276–280.

Clarke, D. M., & Clark, B. A. (2003). Encouraging perseverance in elementary mathematics: A tale of two problems. *Teaching Children Mathematics*, *10*(4), 204–209.

Lee, J.-E. (2007). Context in math learning: Problems and possibilities. *Teaching Children Mathematics*, *14*(1), 40–44.

Wilson, L. D. (2004). On tests, small changes make a big difference. *Teaching Children Mathematics*, *11*(3), 134–137.

Zorin, B., Hunsader, P. D., & Thompson, D. R. (2013). Assessments: numbers, context, graphics and assumptions. *Teaching Children Mathematics*, *19*(8), 480–488.

For middle-grades teachers:

Chappell, M. F., & Thompson, D. R. (1999). Modifying our questions to assess students' thinking. *Mathematics Teaching in the Middle School*, *4*(7), 470–474.

Lambert, R., & Stylianou, D. A. (2013). Posing cognitively demanding tasks to all students. *Mathematics Teaching in the Middle School*, *18*(8), 500–506.

Smith, M. S., & Stein, M. K. (1998). Selecting and creating mathematical tasks: From research to practice. *Mathematics Teaching in the Middle School*, *3*(5), 344–350.

Stylianou, D. A., Kenney, P. A., Silver, E. A., & Cengiz A. (2000). Gaining insight into students' thinking through assessment tasks. *Mathematics Teaching in the Middle School*, *6*(2), 136–144.

Graduate Current Trends Course (Preservice and In-Service Elementary, Middle-Grades, or High School Teachers)

Texts (referenced during the assessment experience):

Hirsch, C. R., Lappan, G. T., & Reys, B. J. (2012). *Curriculum issues in an era of Common Core State Standards for Mathematics.* Reston, VA: National Council of Teachers of Mathematics.

Koestler, C., Felton, M. D., Bieda, K. N., & Otten, S. (2013). *Connecting the NCTM process standards & the CCSSM practices.* Reston, VA: National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston VA: Author.

Readings (varied based on grade level of teachers):

***Week devoted to* Classroom assessments, with a focus on the process standards:**

Bleiler, S. K., & Thompson, D. R. (2012/2013). Multi-dimensional assessment of the Common Core. *Teaching Children Mathematics*, *19*(5), 292–300. (all teachers)

Hirsch, C. R., Lappan, G. T., & Reys, B. J. (2012). Introduction to cultivating mathematical practices and habits of mind. In C. R. Hirsch, G. T. Lappan, & B. J. Reys (Eds.), *Curriculum issues in an era of Common Core State Standards for Mathematics* (pp. 90–92). Reston, VA: National Council of Teachers of Mathematics. (all teachers)

Hunsader, P. D., Thompson, D. R., Zorin, B., Mohn, A. L., Zakrzewski, J., Karadeniz, I., Fisher, E. C., & MacDonald, G. (2014). Assessments accompanying published textbooks: The extent to which mathematical processes are evident. *ZDM: The International Journal on Mathematics Education*, *46*(5), 797–813. (all teachers)

Zorin, B., Hunsader, P. D., & Thompson, D. R. (2013). Assessments: Numbers, context, graphics, and assumptions. *Teaching Children Mathematics*, 19(8), 480–488. (all teachers)

Process standards from the NCTM* Principles and Standards *(all teachers):

Goldenberg, E. P., Mark, J., & Cuoco, A. (2012). *An algebraic-habits-of-mind perspective on elementary school mathematics.* In C. R. Hirsch, G. T. Lappan, & B. J. Reys (Eds.), *Curriculum issues in an era of Common Core State Standards for Mathematics* (pp. 93–104). Reston, VA: National Council of Teachers of Mathematics. (elementary teachers)

Mark, J., Cuoco, A., Goldenberg, E. P., & Sword, S. (2012). Developing mathematical habits of mind. In C. R. Hirsch, G. T. Lappan, & B. J. Reys (Eds.), *Curriculum issues in an era of Common Core State Standards for Mathematics* (pp. 93–104). Reston, VA: National Council of Teachers of Mathematics. (middle-grades teachers)

Process Standards from National Council of Teachers of Mathematics (2000). *Principles and Standards for School Mathematics*. Reston, VA: Author. (all teachers)

Cuoco, A. Goldenberg, E. P., & Mark, J. (2012). Organizing a curriculum around mathematical habits of mind. In C. R. Hirsch, G. T. Lappan, & B. J. Reys (Eds.), *Curriculum issues in an era of Common Core State Standards for Mathematics* (pp. 111–120). Reston, VA: National Council of Teachers of Mathematics. (high school teachers)

***Week devoted to* Mathematical Practices:**

Koestler, C., Felton, M. D., Bieda, K. N., & Otten, S. (2013). *Connecting the NCTM Process Standards & the CCSSM Practices.* Reston, VA: National Council of Teachers of Mathematics.

***Week devoted to* National and International Assessments, NAEP, TIMSS, PISA:**

Skim based on appropriate grade level:

Neidorf, T. S., Binkley, M., Gattis, K., & Nahara, D. (2006). *Comparing mathematics content in the National Assessment of Educational Progress (NAEP), Trends in International Mathematics and Science Study (TIMSS), and Program for International Student Assessment (PISA) 2003 Assessments: Technical report.* Washington, D. C.: National Center for Education Statistics.

OECD. (2011). *Strong performers and successful reformers in education: Lessons from PISA for the United States.* OECD Publishing. http://dx.doi.org/10.1787/9789264096660-en

Programme for International Student Assessment. (2009). *Take the test: Sample questions from OECD's PISA assessments.* OECD.

TIMSS & PIRLS International Study Center. (2007). *TIMSS 2003 mathematics items: Released set fourth grade.* International Association for the Evaluation of Educational Achievement.

TIMSS & PIRLS International Study Center. (n.d.). *TIMSS 2003 mathematics items: Released set eighth grade.* International Association for the Evaluation of Educational Achievement.

***Week focused on* Explorations related to PARCC, Smarter Balanced, and connections to Florida tests and classroom assessments:**

Doorey, N. A. (Dec. 2012/Jan. 2013). Coming soon: A new generation of assessments. *Educational Leadership, 70*(4), 28–34.

(**Return to page 73**)

Appendix B: Comparison of CCSSM's Standards for Mathematical Practice and the Criteria in the MPAC Framework

	Standards for Mathematical Practice	Alignment to MPAC Framework Criteria
1	Make sense of problems and persevere in solving them	Connections, Representation: Graphics, Representation: Translation
2	Reason abstractly and quantitatively	Communication, Connections, Representation: Translation
3	Construct viable arguments and critique the reasoning of others	Reasoning and Proof, Communication
4	Model with mathematics	Communication, Connections, Representation: Translation, Representation: Graphics
5	Use appropriate tools strategically	Representation: Graphics
6	Attend to precision	Communication, Reasoning and Proof
7	Look for and make sense of structure	Connections, Representation: Graphics
8	Look for and express regularity in repeated reasoning	Reasoning and Proof, Connections

(Return to page 85)