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

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, 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, p. 1054).

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

Mathematics Teacher Educator • Vol. 4, No. 1, September 2015
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**

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

*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,
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?

\[ 2 \times \_\_\_ = 12 \]

\[ 12 \div 2 = \_\_\_ \]

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.

<table>
<thead>
<tr>
<th>Event</th>
<th>100 m</th>
<th>200 m</th>
<th>400 m</th>
<th>800 m</th>
<th>1500 m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jared</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Rayvon</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Chou</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>

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.

---

Table 2

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

<table>
<thead>
<tr>
<th>Item</th>
<th>Reasoning &amp; Proof</th>
<th>Communication</th>
<th>Connections</th>
<th>Representation: Graphics</th>
<th>Representation: Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>N</td>
<td>N</td>
<td>I</td>
<td>R</td>
<td>N</td>
</tr>
<tr>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>R</td>
<td>N</td>
<td>SW</td>
</tr>
<tr>
<td>3</td>
<td>Y</td>
<td>Y</td>
<td>I</td>
<td>M</td>
<td>WG</td>
</tr>
<tr>
<td>4</td>
<td>Y</td>
<td>Y</td>
<td>R</td>
<td>I</td>
<td>N</td>
</tr>
</tbody>
</table>

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.
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

<table>
<thead>
<tr>
<th>Original Response</th>
<th>Segment of Thought</th>
<th>Code</th>
<th>Theme</th>
</tr>
</thead>
<tbody>
<tr>
<td>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)</td>
<td>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...</td>
<td>Organized</td>
<td>B: Framework affords</td>
</tr>
<tr>
<td>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)</td>
<td>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.</td>
<td>Value of modifications</td>
<td>D: Aids in modifications and building of quality tests</td>
</tr>
</tbody>
</table>

---

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

“...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

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

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)
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.
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, Fellon, 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


Enhancing Teachers’ Assessment of Mathematical Processes


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:

Readings:
Chapter 5 of Van de Walle text

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

Texts:


Readings:
(completed in class using a jigsaw strategy)

For elementary teachers:


For middle-grades teachers:


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

**Texts (referenced during the assessment experience):**


**Readings (varied based on grade level of teachers):**

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


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


*Mathematics Teacher Educator* • Vol. 4, No. 1, September 2015

**Week devoted to Mathematical Practices:**


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

**Skim based on appropriate grade level:**


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


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

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

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

(Return to page 85)