



Two Teacher Educators' Approaches to Developing Preservice Elementary Teachers' Mathematics Assessment Literacy: Intentions, Outcomes, and New Learning

ABSTRACT

The purpose of this study was to examine and reflect on two teacher educators' approaches to developing preservice elementary teachers' mathematics assessment literacy. We explored the similarities and differences in preservice teachers' conceptions of good assessment practices and their critique of assessment items. We found that we, as course instructors, had different assumptions pertaining to the role of preservice teachers in the development of assessment and offered different assessment-related course activities. Despite these differences, there were more similarities than differences between the two groups of the preservice teachers with regard to their overall perceptions about good assessment practices and their critique of assessment items. However, we also observed differences in the criteria they used in critiquing assessment items. Discussions and implications are presented in accordance with these findings as a means to improve our own teaching and student learning.

KEYWORDS

preservice teacher education, assessment literacy, instructional activities and practices, faculty development

Preparing preservice teachers (PSTs) to be high quality teachers is a challenging and complex task. As new educational needs and expectations continuously arise, a major goal of teacher education is accommodating PSTs who are "learning to teach" through "a two-step process of knowledge acquisition and application or transfer" (Feiman-Nemser & Remillard, 1996, p.79). However, PSTs bring a variety of previous experiences, strengths, and limitations to teacher education programs. There is also a wide range of differences in the types of experiences PSTs are expected to engage in across different classes and programs. An underlying principle that exists in this variety is the notion of "equifinality" where the same final state may be reached by taking many different ways in open systems (von Bertalanffy, 1968). In particular when it comes to teaching methods, the principle of "equifinality" embraces different instructional methods that can be equally effective in achieving the same instructional goals. Avoiding a 'one-size-fits-all' approach, the notion

of equifinality allows teacher educators to leverage this variety of unique contexts and also requires teacher educators to continuously reflect upon their own practices and examine PSTs' outcomes rather than leaving the PSTs in the realm of ambiguity.

The purpose of this study is to collectively reflect upon our own teaching practices by investigating the similarities and differences in future teachers' conceptions of mathematics assessment and their critique of assessment items. The impetus for this joint project was based on the recognition of the inconsistencies and lack of communication among methods instructors and courses in the United States (Kastberg, Sanchez, Edenfield, Tyminski, & Stump, 2012), and the awareness of the importance of the development of scholarly inquiry in teaching (Chick, 2013). Although we were working at different teacher education programs, we shared many goals for effective elementary mathematics teachers. One important shared goal was preparing our PSTs to be teachers equipped with "assessment literacy." Yet, we noticed that different course activities and instructional approaches on the topic of classroom assessment were being used to achieve this goal. As a way to make our implicit assumptions more explicit and improve our teaching practice to better prepare PSTs, we attempted to investigate the similarities and differences in our preservice teachers' conceptions of mathematics assessment and their critique of assessment items.

FRAMING THE STUDY

In this study, we closely examined activities related to teacher assessment literacy because it is one of the areas emphasized the most in current educational trends. Research questions are also based on the challenges many mathematics teacher educators encounter. We intended to collaboratively examine course activities by "systematically assessing and evaluating the impact of our own teaching on students' learning" (Gilpin, 2011, p.1).

Teacher assessment literacy

Teaching, learning, and assessment are essential, interrelated cornerstones of education. Whenever there are shifts in educational principles or curriculum standards, all of these aspects are reexamined to reset goals. The upcoming implementation of the Common Core State Standards [CCSS] (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) in the United States is the most recent major curriculum reform. Accordingly, there is a call for assessments that align with the CCSS, and the development of these assessments is underway.

The continuous development of assessment theory and newly emerging perspectives calls for assessment-literate teachers. Assessment was traditionally viewed as a procedure at the end of instruction to measure the end results of student learning and provide accountability for the education system (Broadfoot, 2000). A new conception of assessment referred to as assessment *for* learning places emphasis on determining student progress throughout the instructional process (e.g., Black & Wiliam, 1998; Black, Harrison, Lee, Marshall, & Wiliam, 2004; Stiggins, 2002). Furthermore, the view of assessment *as* learning encourages teachers and students to use assessments as a meta-cognitive learning tool allowing students to monitor their own learning.

It is widely advocated that the one-size-fits-all assessment model is not feasible considering the complexity of the teaching and learning process. A balanced synergy among different views of assessment is needed to impact every stage of a student's learning process

(e.g., Earl, 2003; Stiggins, 2002). To create this balanced synergy, teachers are required to have a variety of knowledge and skills.

There are many different ways to define assessment literacy. Generally, knowledge and skills are characterized in terms of two overarching purposes: (a) identify, select, or create assessments optimally designed for various purposes, such as accountability, instructional program evaluation, student growth monitoring and/or promotion, and diagnosis of specific student needs; (b) analyze, evaluate, and use the quantitative and qualitative evidence generated by external summative and interim assessments, classroom summative assessments, and instructionally embedded formative assessment practices to make appropriate decisions to improve programs and specific instructional approaches to advance student learning (Kahl, Hofman, & Bryant, 2013, p 8). We utilized the first component, identifying and selecting assessment, to understand PSTs' assessment literacy.

Challenges for mathematics methods course for PSTs

There has been a plethora of research about the preparation of elementary mathematics teachers. Some researchers are concerned about the knowledge, beliefs, and attitudes PSTs bring to teacher education programs (e.g., Ball, 1990). Many mathematics teacher educators would concur that it is not feasible to incorporate all necessary knowledge and skills into the short time period of a methods course. Teacher educators strive to effectively design methods courses that consist of a specific selection of content, tasks, and resources. Ensor (2000) calls this collection a *privileged repertoire* in that "it involves a particular selection and combination of mathematics for the production of pedagogic tasks, a particular selection of pedagogic resources to facilitate this and arrangement of these tasks into sequences as lessons" (p. 235). How a privileged repertoire is constructed depends on the views of good teaching by particular instructors and/or of teacher education programs. This may explain the wide range of enacted activities that exist in methods courses. Many research studies imply that these inconsistent variations may have resulted from the different priorities of individual teacher educators (Kastberg et al., 2012; Kosnik & Beck, 2009; Taylor & Ronau, 2006). Courses often operate in isolation with unexamined assumptions about the effectiveness of course activities. In this regard, several challenging questions can be asked. First, what are individual instructors' rationales for their privileged repertoires? Second, how do we know the impact of instructors' privileged repertoires on PSTs' development? Third, to what extent do instructors' privileged repertoires influence PSTs' own privileged repertoires?

Issues under consideration

This project set the space for us to investigate and share our own practices and reflections in relation with others. We do not attempt to ascertain which approach is better. Rather, we see this as an opportunity to learn from each other. Our study was framed around the following questions:

1. What are the similarities and differences in learning objectives, expected outcomes and the enacted instructional activities and assignments regarding assessment literacy between two instructors' courses?
2. Are there any differences in conceptions of assessment between students of two instructors as students finish their methods courses?

3. Are there any differences in evaluating the quality of assessment items between students of two instructors as students finish their methods course?

By answering these research questions, we intended to provide collective reflection on our PSTs' learning outcomes and needs. This study collectively examined our course design, student outcomes, and our own learning from the results as a means to improve our own teaching and student learning.

METHOD

Background of students and contexts

Data from this study came from two elementary mathematics methods course instructors at two sites in the United States and the PSTs who were enrolled in each of the courses. Instructor 1 had 34 students and Instructor 2 had 37 students. For most PSTs, it was the last semester before their full-time student teaching. Although PSTs in both classes were about to embark on their professional journey as novice elementary teachers, their backgrounds were different. Instructor 1's students were undergraduate students who majored in elementary education. Instructor 2's students were graduate students who already held bachelor's degrees in fields related to education.

Data sources

Three main data sources were analyzed to answer the research questions: (1) annotated course syllabi; (2) a belief survey on good assessment practices; and (3) critique of assessment items. All of them were created and administered towards the end of the fall semester of 2013.

Annotated course syllabi

Our original course syllabi showed that we each proposed learning objectives and rationales for course activities around the topic of assessment literacy. We created annotated syllabi to track thinking about our course design. The annotated syllabi included course description, institutional and program context, theoretical rationale for the design of activities and assignments, and critical reflection.


Survey on good assessment practices


To identify PSTs' conceptions of good mathematics assessment, we asked an open-ended question: *What constitutes good mathematics assessment?* There were no specific formats or choices for PSTs' written responses in order not to limit their initial thoughts. We utilized this survey as a data source to see any similarities or differences in our PSTs' espoused beliefs.

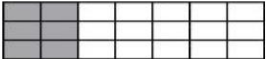
Critique of assessment items

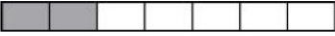
For this task, we used five items as shown in Figure 1. These were sample items from two consortia that have been developing assessments aligned with the Common Core State Standards (i.e., the Smarter Balanced Assessment System [SBAS] and the Partnership for Assessment of Readiness for College and Careers [PARCC]). We asked our PSTs to rank the quality of these assessment items from most preferred to least preferred. In addition, we asked for the basis of their evaluation for the most/least preferred item and the aspects they wanted to modify to improve the least preferred item. From this task,

Item 1. Which model below best represents the fraction $\frac{2}{5}$?

A. 

B. 

C. 

D. 


Item 2. Patricia needs to read for 120 minutes each week.

- She read for 26 minutes on Monday.
- She read for 39 minutes on Tuesday.
- She read for 38 minutes on Thursday.

How many more minutes does Patricia need to read this week? () minutes

Item 3. Represent each fraction to the correct location on the number line.

$\frac{1}{2}$, $\frac{3}{4}$, $\frac{6}{2}$



Item 4. Circle all the equations that are true.

- $8 \times 9 = 81$
- $54 \div 9 = 24 \div 6$
- $7 \times 5 = 25$
- $8 \times 3 = 4 \times 6$
- $49 \div 7 = 56 \div 8$

Item 5. Write a fraction that is equivalent to $\frac{3}{4}$.

$\frac{3}{4} = \frac{[]}{[]}$

Figure 1. Items used for the selection task

each of our PSTs' overall tendencies in ranking the items and their criteria used in their ranking were compared.

DATA ANALYSIS

Annotated syllabi were compared and contrasted. We particularly focused on the course objectives, class topics and activities, and course assignments that were designed for the development of assessment literacy. Instructors' annotations indicated their intentions and justifications for course activities and assignments. Some of these were ex-

PLICIT as shared with students via the syllabus. Others were implicit and were meant to provide a model that could be adapted for use in future classrooms. The intentions and justifications were used to interpret similarities and differences from each group of PSTs' perceptions of assessment upon completion of the course.

For two data sources, the belief survey and the critique task, we used an inductive content analysis approach (Grbich, 2007). We read all of the responses and created codes based on the raw data from the following processes: (a) initial reading of each response, (b) finding emerging categories and subcategories in each aspect, (c) coding the categories, and (d) checking inter-rater reliability. There was 100% agreement on the coding of 89% of the examples. To answer our research questions, we presented our data using descriptive statistics (e.g., frequencies for each category). Our final reflections were based on the identified similarities and differences from this data analysis.

FINDINGS

Despite our differing assumptions on the role of preservice teachers in the development of assessment items and different course activities, we observed some similar tendencies between the two courses with respect to PSTs' overall perceptions of good assessment practice and their critique of assessment items. However, we observed gaps that existed between their conception of good assessment practice and their evaluation of actual assessment items. More details on how two instructors' personal views are reflected in their course activities and how it has influenced their preservice teachers' assessment literacy are discussed in the following section.

Assessment practices exemplified in the course

To answer our first question (i.e., similarities/differences in learning objectives, expected outcomes and the enacted instructional activities and assignments regarding assessment literacy between two instructors' courses), we compared and contrasted our original course syllabi and annotated syllabi.

Similarities

One similarity between the two courses was that assessment was perceived as an important part of teaching practice. The following assessment-related course goals and objectives were noted in Instructor 1's syllabus:

- Create, modify, and assess appropriate curricula to meet cognitive, affective and psychomotor learning objectives
- Demonstrate effective instructional practices
- Assess the progress of students who are learning mathematics and be able to remediate for students who are having difficulties

Instructor 2 also included the following goals and objectives:

- Developing your pedagogy: Pedagogy is a word that encompasses many aspects of teaching, including the work a teacher does "behind the scenes" to plan for instruction, as well as the teaching and assessing that take place in the classroom itself.
- Developing knowledge of curriculum, planning, and assessment

Through the review of the annotated course syllabi and extended communications, we came to an agreement that assessment literacy could cover virtually all aspects of as-

assessment practices. The first challenge this led us to was how to effectively design course tasks that would have a lasting impact on PSTs in the short time given in one methods course. The second challenge was how to effectively orchestrate the topic of assessment literacy along with other major priority topics commonly expected to be addressed in a mathematics methods course. Although there was a consensus on the importance of addressing assessment literacy in a mathematics methods course, there were differences in terms of the aspects that were emphasized and those that were deemed not as important as others.

Differences

Coping with the aforementioned challenges, we found that there were some differences in the course activities that aimed to address the topic of assessment literacy. Table 1 shows all course activities, assignments, and practices intended to develop teacher assessment literacy in each instructor's class.

One notable difference was the elements of assessment literacy each instructor intended to highlight through the course assignments and activities. Instructor 1 approached assessment literacy from the premise that it is important for PSTs to experience the item development process as this is a critical ability of a professional teacher. If teachers were excluded from this process, it would contribute to deskilling of teachers due to the separation of conception from execution (Apple, 1986, 995). Thus, regardless of two readily available major assessments that align with the CCSS (i.e., SBAS and PARCC that will be fully implemented in the 2014-15 academic year), Instructor 1 believed that it was important for PSTs to experience the process of assessment item development and implementation in the teacher preparation program. In contrast, Instructor 2 held a supposition that it is more important for educators to critically select, revise, and use them in their current educational context, rather than to create assessment items from scratch. This supposition was particularly supported by the availability of the aforementioned two major assessments. As a result, Instructor 2 focused more on the PSTs pedagogical skills related to the effective use of existing resources. Therefore, Instructor 1 devoted more time discussing assessment-related topics throughout the semester whereas Instructor 2 focused on developing the ability to determine fundamental ideas embedded in the problems and possible student misconceptions and strategies.

Perception on good mathematics assessment

To answer our second question (i.e., differences in conceptions of assessment), we examined and organized PSTs' written responses to the open-ended question of *what constitutes good mathematics assessment*. PSTs responses were coded into categories that emerged from our analysis. Most PSTs addressed multiple facets of good assessment. Thus, their statements were coded in multiple categories. Table 2 shows the categories and frequencies of responses from each instructor's class.

Similarities

The two groups' overall perceptions on good assessment practice were similar in each major category. Both groups perceived that assessment *for learning*, commonly referred to as formative and diagnostic assessments, was the most important purpose of assessment. This demonstrated that they view learning as a process of contextualized knowledge construction, as well as viewing assessment practice as a meaningful way to

Table 1. *Enacted course activities and assignments for assessment literacy*

	INSTRUCTOR 1	INSTRUCTOR 2
Major assessment literacy-related course activities/assignments	<Common> <ul style="list-style-type: none"> • Review of CCSS for instructional planning and assessment • Lesson planning and implementation/demonstration (assessment is one component) 	
	<ul style="list-style-type: none"> • Pre and post classroom assessment conception surveys • Critique of sample assessment items used in statewide standardized assessment • Math assessment packet: (a) create student assessment interview items that are appropriate for student teaching grade level, (b) conduct mock interviews with peers in the university class, (c) conduct one-on-one assessment interviews with K-8 students, (d) analyze assessment data and devise potential intervention plans 	<ul style="list-style-type: none"> • Create mathematical problems that they could use for teaching and assessing student understanding aligned with the CCSS. • Create an effective teaching strategy: (a) select lesson plans that include the selected teaching strategy, (b) create or find an assessment to measure students' understanding and the effectiveness of intervention (c) implementing lesson activities, (d) analyze students' understanding
Other assessment practices implicitly exemplified in the course	<Common> <ul style="list-style-type: none"> • Self-assessment opportunities for various course activities/assignments • Peer-assessment opportunities for group activities/projects 	
	<ul style="list-style-type: none"> • Partner exam opportunities as a choice for two performance-oriented exams • Ongoing feedback on math assessment packet materials including self-, peer-, and instructor feedback. 	

promote the development of individual understanding (Taylor, 1994). Most of them believed that the assessment items that required higher-level cognitive demand were better than the items that asked for simple procedures or routine computations. These PSTs also valued the assessment items related to real-life contexts and contained varying difficulty levels and formats.

Differences

The only subcategory that showed a distinct difference in the cognitive demands category was *representation*. Statements in this subcategory highlighted the importance of students' ability to visually represent given information using various modes of representation and teachers' ability of incorporating a variety of representations in the assessment as a means to accurately access students' understanding.

Critique task

To answer our third question (i.e., differences in evaluating the quality of assessment items), we administered and analyzed a critique task using the five items listed in Figure 1. Table 3 shows the general tendency of the most and least preferred selections.

Table 2. *PSTs' perspectives on good assessment*

Category	Sub-category	Frequency Instructor 1 (N=34)	Frequency Instructor 2 (N=37)
Purpose: <i>Why</i> assess?	· Assessment of learning	6 (17.6 %)	3 (8.1 %)
	· Assessment for learning	9 (26.5 %)	9 (24.3 %)
	· Assessment as learning	5 (14.7 %)	3 (8.1 %)
Cognitive demands: <i>What</i> to assess?	· Procedure/computation	1 (3 %)	0 (0 %)
	· Understanding/critical thinking	11 (32.4 %)	8 (21.6 %)
	· Representation	6 (17.6 %)	0 (0 %)
	· Explanation/reasoning	5 (14.7 %)	6 (16.2 %)
	· Application to real-word contexts	5 (14.7 %)	6 (16.2 %)
Mathematical complexity	· Easy and difficult problems	7 (20.1 %)	7 (18.9 %)
Formats: <i>How</i> to assess?	· Traditional/formal	0 (0 %)	0 (0 %)
	· Alternative/informal (e.g., performance-based assessment)	6 (17.6 %)	5 (13.5 %)
Other issues considered	· Consideration of students' diverse background/ learning styles	2 (5.9 %)	1 (2.7 %)
	· Consideration of student engagement/motivation	2 (5.9 %)	2 (5.4 %)

Note. The majority of PSTs addressed multiple categories. These responses were coded in multiple categories as long as the categories were present in their written responses.

Table 3. *Distribution of PSTs' preference in evaluating the sample assessment items*

	Most Preferred		Least Preferred	
	Instructor 1 (N=34)	Instructor 2 (N=37)	Instructor 1 (N=34)	Instructor 2 (N=37)
Item 1 (Visual representation of 2/5)	5 (14.7 %)	6 (16.2 %)	13 (38.2 %)	11 (18.9 %)
Item 2 (Word problems with addition)	8 (23.5 %)	9 (24.3 %)	7 (20.6 %)	5 (13.5 %)
Item 3 (Number line fraction problem)	11 (32.4 %)	12 (35.3 %)	4 (11.8 %)	4 (10.8 %)
Item 4 (Operation equation problem)	7 (20.6 %)	7 (18.9 %)	6 (17.6 %)	6 (16.2 %)
Item 5 (Equivalent fraction to 3/4)	3 (8.8 %)	2 (5.4 %)	4 (11.8 %)	9 (24.3 %)

Each item was considered as most preferred and least preferred, suggesting the different set of evaluation criteria each PST may hold. Item 3 was selected as most preferred in both instructors' classes, and item 1 as least preferred. A varying preference toward item 5 is noticeable. To illustrate similarities and differences in PSTs' evaluation criteria, Table 4 shows the categories of rationales used in PSTs' selection of the most and least preferred items as indicated in their written responses. We observed that our PSTs tend to use the same criteria in selecting these items. For example, three PSTs in Instructor 1's class addressed *clarity* when they described their justifications for choosing the most preferred item (e.g., it is a good item because of high-level of clarity) and 13 PSTs use the same reason to address weaknesses (e.g., it is a weak item because of low-level of clarity). In addition, Table 4 illustrates that our PSTs considered various aspects of each assessment item including clarity, cognitive demand, mathematical complexity, format of assessment, and other student-related issues such as motivation and fairness.

Similarities

A large portion of the PSTs in both courses paid attention to the cognitive aspects assessment items required of students when selecting a most preferred item. For example, around 53% of PSTs in Instructor 1's class and 38% PSTs in Instructor 2's class highlighted the importance of measuring deep understanding of mathematics in the selection of most preferred item. In particular, the cognitive demand required for students is the most popular reason for their selection of most preferred. Yet when asked to select a least preferred item, PSTs in both classes pointed out other reasons including clarity, personal mathematical and pedagogical preference, mathematical complexity, format, and other issues. Interestingly, when PSTs' tendency in the selection of a least preferred item is compared to that of a most preferred item, PSTs in both classes seemed to show a favor toward certain assessment formats and the affective aspects of assessment in the selection of a least preferred item than in the selection of a most preferred one.

Differences.

We also noticed many differences in PSTs' rationales for the selection of the most preferred and least preferred items from the two instructors' classes. First, in comparison to PSTs in Instructor 2's class, more of the PSTs in Instructor 1's class paid more attention to clarity in wordings, directions, and representations used for both most preferred and least preferred assessment items. For example, PSTs in Instructor 1's class evaluated that Item 1 was least preferred because the unequal parts used in choice (A) shown below were not clear enough to represent $2/5$, in particular in comparison to choice (B). Six out of 13 PSTs who evaluated Item 1 as the least preferred addressed this concern of not making the inequality of parts more evident. Whereas, this was not an issue at all in Instructor 2's class since all of them clearly discerned the unequal parts (see Table 4).

In contrast, a relatively higher number of PSTs in Instructor 2's class evaluated the items based on their personal preference in mathematical difficulty or modes of representation. Another difference was that Instructor 2's PSTs used more non-mathematics-specific features in evaluating items such as the motivational level of the assessment item.

DISCUSSIONS AND IMPLICATIONS

The opportunity to examine our own course design and our PSTs' conceptions of assessment literacy allowed us to reflect upon what we hoped to achieve through our methods courses. We were also able to reflect upon the types of knowledge and skills our PSTs felt they possessed as they were about to exit our teacher education programs. We admit that what we did in our individual courses may truly be insufficient to cover everything our PSTs need to know. Nevertheless, valuable ideas from this SoTL project will give course instructors new concepts to consider.

Instructors similar, but different emphasis and expectations

The opportunity to create and share annotated course syllabi provided us a venue to clarify our thoughts behind the course design, which was more beneficial than informal exchanges about teaching approaches. Although our courses have evolved over the years of teaching, we realized that we had little to document the thinking process behind the course design. The similarities in our annotated course syllabi showed that both of us valued experiential and authentic learning as well as active meaning-making through

Table 4. Distribution of rationales for most and least preferred item

Category	Description	Most Preferred		Least Preferred	
		Ins. 1 (N=34)	Ins. 2 (N=37)	Ins. 1 (N=34)	Ins. 2 (N=37)
<i>Clarity</i>	Clarity in wording/directions/representations used	3 (8.8%)	0 (0%)	13(38.3%)	1 (2.7%)
<i>Personal Preference</i>	PSTs' personal preference in math topics/representation	2 (5.9%)	8 (21.6%)	1 (2.9%)	10 (27.0%)
<i>Cognitive Demand</i>	Factual	0 (0%)	0 (0%)	4 (11.8%)	10 (27.0%)
	Understanding	18(52.9%)	14(37.8%)	7(18.9%)	13(35.1%)
	Explanation	6(17.6%)	0 (0%)	3(8.8%)	10(27.0%)
	Application	2 (5.9%)	3 (8.1%)	1 (2.9%)	1 (2.7%)
	Multiple Concepts	1 (2.9%)	8 (21.6%)	1 (2.9%)	1(2.7%)
	Contexts to analyze	4 (11.8%)	3 (8.1%)	1 (2.9%)	1(2.7%)
	Effectiveness of Representation	5 (14.7%)	13 (35.1%)	1 (2.9%)	1(2.7%)
<i>Math Complexity</i>	Question is easy/difficult to students	1 (2.9%)	1 (2.7%)	4 (11.8%)	5 (13.5%)
<i>Format</i>	Multiple choice/Multiple answers/Open-ended	1 (2.9%)	5 (13.5%)	3 (8.8%)	10 (27.0%)
<i>Other Issues</i>	Motivation, Fairness	0 (0%)	3 (8.1%)	1 (2.9%)	5 (13.5%)

Note. The majority of PSTs addressed multiple categories. These responses were coded in multiple categories as long as the categories were present in their written responses.

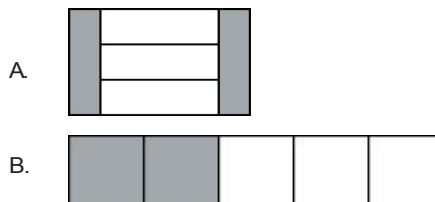


Figure 2. Graphic representations used in Item 1

processes of critical reflection. We tried to help our PSTs engage in the actual assessment process of designing, scoring, and interpreting assessment for various purposes with an emphasis on formative assessment processes (Shepard, Hammerness, Darling-Hammond, & Rust, 2005). However, different expectations existed when it came to assessment designing activities. Instructor 1 strictly asked for developing assessment items because she believed that it was one of the essential activities PSTs should experience. By engaging in this development process, PSTs would be prepared for evaluating the flood of assessment items and methods that will be available in their future career. Since teachers have indicated that they are more concerned with the day-to-day issues related to the application of assessment processes (Rogers, 1991), Instructor 1 felt that more fundamental assessment issues could be addressed freely during the teacher education program while PSTs actually engaged in the assessment item development process. In contrast, Instructor 2 allowed PSTs to adapt already existing assessment items developed by others. She believed that PSTs could benefit more from pre-designed assessment items with autonomy

for adaptation and that they would be able to create their own classroom assessment as they become mature teachers in the future. In short, there was a difference in the two instructors' expectations regarding what should be emphasized during the teacher education program and how the PSTs' experience in the program would impact their future.

Another difference was the amount of implicit modeling of assessment practices. Both instructors utilized *assessment for learning* approaches with PSTs by inviting them to participate in formative assessment processes (e.g., self-, peer-evaluations or ongoing instructor feedback on long-term projects). Instructor 1 added more of these components hoping that the actual modeling would provide more positive and authentic experience of assessment (e.g., James & Pedder, 2006; Wiliam, 2011). Although we observed a clear difference in each instructor's emphasis and expectation, it was challenging to interpret what this distinction meant.

Resulting differences in PSTs' conceptions and selection of good assessment items

PSTs responses to the question, "*What constitutes good mathematics assessment?*" demonstrated their conceptions of good assessment practice. Although it was an open-ended question, it was interesting to see that the emerging themes and frequencies were very similar regardless of the instructors' different emphasis and expectations in course activities/assignments. The only distinctive category between two classes was *representation*. Instructor 1 explained that her PSTs had extensive discussion on modes of representation while they developed their own assessment items and participated in mock assessment interviews with their peers. Both instructors' were satisfied with their PSTs' responses, which were well aligned with the current educational trends and expectations of teachers. However, this extreme similarity in PSTs' conceptions of good mathematics assessment in both classes raised some questions. We were not sure whether the design of course activities and assignments positively impacted PSTs' conceptions of good mathematics assessment or if their conceptions were developed regardless of their learning experiences in our courses. Could we obtain similar results if we asked the same questions at the beginning of the semester? Were these PSTs using "espoused theories" that represent what they said they would do in a certain situation (Argyris & Schon, 1974)? These were questions we could not answer with the data from PSTs' responses to the question about good assessment. PSTs' responses in the critique task partially revealed their "theory-in-use" that represents what they actually do in evaluating assessment items (Argyris & Schon, 1974).

In the critique task, PSTs' choices of good or poor assessment items were similar. However, there existed a wide range of justifications. It was notable that we could find some justifications that did not appear when PSTs talked about what constitutes good mathematics assessment. For example, we observed that some PSTs, particularly from Instructor 2's class, relied on their own personal preference, confidence, and familiarity with the given topics (e.g., fractions). We also noticed that many PSTs perceived the cognitive demand of the same assessment item differently as evidenced in the different set of evaluation criteria. This implied that our PSTs' knowledge about the kinds and level of thinking required of students in specific problems was still in the developing stage regardless of the seemingly coherent view on good assessment practices aligned with the current reform ideas. Furthermore, we observed that there were some discrepancies be-

tween the PSTs' conceptions of good assessments and their actual evaluation of assessment items. For example, while none of the PSTs mentioned a traditional format (e.g., multiple choice items as a component of good assessment practices), some PSTs considered it in the selection of assessment items (see Table 4). In addition, fewer PSTs seemed to consider student affect-related criteria such as students' learning style or engagement and motivation in the selection of assessment items. This result showed the gap between our PSTs' "espoused theory" and "theory-in-use."

What we learned from the observed similarities and differences

Instructor 1 paid attention to the criteria used in evaluating Item 1 in her class. Many PSTs in Instructor 1's class mentioned that Item 1 was a poor assessment item since the drawing did not clearly note the different sizes of parts, but Instructor 2's class emphasized the cognitive complexity and the possibility of measuring deeper understanding of fraction concepts resulting from such drawings. While agreeing with their evaluation, this difference left Instructor 1 with some questions. Why did only her PSTs pay attention to the minor difference shown in the visual representation? How important is this? Instructor 1 recalled that a large portion of peer discussion during the mock assessment interview session was related to effectively incorporating various representations into the development of mathematics assessment items. This result informed Instructor 1 of the need to systemically examine the level of discussion on mathematical representation that occurred during the peer group collaboration. By reviewing that data, it would be much clearer whether they focused on representation solely on the surface, or if they reached a much deeper level. Also, Instructor 1 noted that many assessment practices she implicitly intended to demonstrate were not visible in her PSTs' responses.

Instructor 2 noted that many of her PSTs used their personal preference as one of the criteria in evaluating the assessment item. While emphasizing adaptability of assessment items, Instructor 2 highlighted the importance of teachers' knowledge and skills and their autonomy in the creation of students' learning opportunities in instruction and assessment throughout the semester. Thus, PSTs' mathematical and pedagogical preference as one of the criteria in evaluating the assessment item seems to mirror the instructor's intention. This result seems to suggest Instructor 2 placed a careful emphasis on teachers' autonomy in the creation of students' learning opportunities in instruction and assessment. Instructor 2 observed that some PSTs with a limited understanding of mathematics tend to rely on their personal preference. For example, one PST choose item 1 as least preferred because he/she had a harder time with fractions.

Through this study, we had an opportunity to see what similarities and differences each of our PSTs exhibited with respect to assessment literacy and what those results meant to us. As we complete our first round of the study, it is time to look back on each of our 'privileged repertoires.' One major question remaining is about the implicit teaching practice we utilized. We tackled the topic of assessment literacy in various modes. Some were explicit discussion, activities, and assignments. Others were implicit demonstrations throughout the course. While striving to model assessment strategies in various ways, not simply by telling but using active engagement as the form of our content delivery, our intentions did not clearly resonate in PSTs' conceptions or evaluations.

This study opens an opportunity for further research. First, in the present study we focused only on PSTs' conceptions upon completion of the course. In the future, we

want to tie in before, during, and after stages so that we can get more valid data on the effectiveness of our teaching strategies. Second, we can further examine how explicit the discourse about good assessment practice needs to be to have a positive impact on PSTs' conceptions and practices. We may need some scaffolding ourselves to help our PSTs' understanding of how to promote their assessment skills which will in turn impact their students' learning.

Implications for other disciplines

In this joint project, we reflected upon our own teaching practices as teachers of future teachers in a specific discipline, mathematics education. Nonetheless, we believe that the findings of this study will also resonate with those who are involved in learning and teaching other disciplines because virtually all professions require people to possess the assessment-related knowledge and skills in order to competently perform their responsibilities (Popham, 2009).

Prior research has documented inconsistencies and a lack of communication among instructors and courses not only in mathematics education but also in other areas (e.g., Gregory, Ellis, & Orenstein, 2011; Kastberg, et al., 2012), and collaborative work has not been the norm at many schools and universities. Many individual instructors need to improve their own assessment literacy through identifying, selecting, or creating assessments designed for various purposes, as well as analyzing, evaluating, and using data to make instructional decisions (Kahl, Hofman, & Bryant, 2013). This level of assessment literacy is a key indicator of instructors' quality. We suggest that, as we experienced in this project, instructors in other disciplines should experience the collaborative opportunity where they review and compare their rationales of course assignments and assessments.

It would also be worthwhile to invite students to share their conceptions of assessment and their critique of assessment items or tasks in various disciplines. This investigation may elicit the gap that exists between students' "espoused theory" and "theory-in-use" as we reported here and provide implications for instructors. Similar analytical frameworks, such as the ones used for this study in Table 2 and Table 4, can be developed as a basis for analysis. This assessment practice will help students become engaged in their learning as self-regulated and reflective learners. We hope that our study generates many different ways in which the design and results could be incorporated into others' classrooms.

Ji-Eun Lee is Associate Professor in the Department of Teacher Development and Educational Studies at Oakland University (USA).

Ji-Won Son is Assistant Professor in the Department of Learning and Instruction at University at Buffalo, SUNY (USA).

REFERENCES

- Apple, M. W. (1986). *Teachers and texts: A political economy of class and gender relations in education*. New York: Routledge & Kegan Paul.
- Apple, M. W. (1995). *Education and power*. New York: Routledge.
- Argyris, C. and Schön, D. (1974) *Theory in practice: Increasing professional effectiveness*, San Francisco: Jossey-Bass.

- Ball, D. L. (1990). The mathematical understandings that PSTs bring to teacher education. *The Elementary School Journal*, 90(4), 449-465.
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139-148.
- Black, P., Harrison C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, 86, 9-21.
- Broadfoot, P. (2000). Assessment and intuition. In T. Atkinson & G. Claxton (Eds.), *The intuitive practitioner: On the value of not always knowing what one is doing* (pp. 199-219). Buckingham: Open University Press.
- Chick, N. (2013). Difference, privilege, and power in the scholarship of teaching and learning: The value of humanities SoTL. In K. McKinney (ed.), *The scholarship of teaching and learning: In and across the disciplines* (pp. 15-33). Bloomington: Indiana University Press.
- Earl, L. (2003). *Assessment as learning*. Thousand Oaks, CA: Corwin.
- Ensor, P. (2000). Recognizing and realizing "best practice" in initial mathematics teacher education and classroom teaching. In J. Bana, & A. Champman (Eds.), *Mathematics education beyond 2000. Proceedings of the 23rd annual conference of the mathematics education research group of Australasia* (pp.235-242). Fremantle, Western Australia: MERGA.
- Feiman-Nemser, S. & Remillard, J. (1996). Perspectives on learning to teach. In F. Murray (ed.), *The Teacher Educator's Handbook* (pp.63-91). San Francisco: Jossey-Bass.
- Gilpin, L. S. (2011). Scholarship of Teaching and Learning Trades. *International Journal for the Scholarship of Teaching and Learning*, 5(2). Article 4. Retrieved from: <http://digitalcommons.georgiasouthern.edu/ij-sotl/vol5/iss2/4>
- Grbich, C. (2007). *Qualitative data analysis: An introduction*. Thousand Oaks, CA: Sage.
- Gregory, E., Ellis, J. P., & Orenstein, A. N. (2011). A proposal for a common minimal topic set in introductory biology courses for majors. *The American Biology Teacher*, 72(1), 16-21.
- James, M., & Pedder, D. (2006). Professional learning as a condition for assessment for learning. In J. Gardner (Ed.), *Assessment and learning* (pp. 27-44). London, UK: Sage.
- Kahl, S., Hofman, P., & Bryant, S. (2013). *Assessment literacy standards and performance measures for teacher candidates and practicing teachers*. Dover, NH: Measured Progress.
- Kastberg, S., Sanchez, W. B., Ednfield, K. W., Tyminski, A., & Stump, C. (2012). What is the content of Methods? Building an understanding of frameworks for Mathematics Methods Courses. *Proceedings for the Thirty-fourth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Kalamazoo, Michigan.
- Kosnik, C., & Beck, C. (2009). *Priorities in teacher education: The 7 key elements of preservice preparation*. New York: Routledge.
- National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for Mathematics*. Washington, DC: Author.
- Popham, W. J. (2009, June 4). *Is Assessment Literacy the "Magic Bullet"?* [Web log comment]. Retrieved from <http://www.hepg.org/blog/19>.
- Rogers, T. (1991). Educational assessment in Canada: Evolution or extinction? *The Alberta Journal of Educational Research* 37, 179-192.

- Shepard, L., Hammerness, K., Darling-Hammond, L., & Rust, F. (2005). Assessment. In L. Darling-Jammond & J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 275-326). San Francisco, CA: Jossey-Bass.
- Stiggins, R. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758-765.
- Taylor, C. (1994). Assessment for measurement or standards: The peril and promise of large scale assessment reform. *American Educational Research Journal*, 13(2), 231–262.
- Taylor, P. M., & Ronau, R. (2006, Fall). Syllabus study: A structured look at mathematics methods courses. *AMTE Connections*, 12-15.
- Von Bertalanffy, L. (1968). *General systems theory*. New York: Braziller.
- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3-14.