Beyond the Criteria:
Evidence of Teacher Learning in a Performance Assessment

By Anissa R. Stewart, Jennifer N. Scalzo, Nicole Merino, & Katherine Nilsen

Research on university-based teacher preparation has been routinely scrutinized. Current criticisms by policy makers and scholars are focused on the need for empirically based evidence on if and how teacher preparation matters (e.g., Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009; Darling-Hammond, 2008; National Research Council [NRC], 2010). One of the strongest sources of evidence comes from the report *Studying Teacher Education: The Report of the AERA Panel on Research and Teacher Education* (Cochran-Smith & Zeichner, 2005), which clearly illustrates that teacher education programs influence preservice teachers' thinking about teaching and learning, self-awareness, and beliefs and attitudes (see also Clift & Brady, 2005; Hollins & Guzman, 2005). Further evidence exists on the positive effects of assessments in university teacher education programs (e.g., Bunch, Aguirre, & Téllez, 2009; Darling-Hammond & Snyder, 2000; Nagle, 2009; Snyder, Lippincott, & Bower, 1998), but lack of consistent evidence threatens the

Anissa R. Stewart is program director with University Extension, University of California, Santa Barbara; Jennifer N. Scalzo is a lecturer in the Teacher Education Program of the Gevirtz Graduate School of Education, University of California, Santa Barbara; Nicole Merino is director of PACT/TPA with the Stanford Center for Assessment, Learning, and Equity, Stanford University, Stanford, California; and Katherine Nilsen is program manager with University Extension, University of California, Santa Barbara. anissa.stewart@ucsb.edu, jscalzo@education.ucsb.edu, nmerino@stanford.edu, & knilsen@ucsb.edu

sustained belief that teacher education programs enhance teacher effectiveness (Grossman, 2008).

Although teacher preparation programs are required to show evidence of preservice candidates' teaching ability, most assessments have been in the form of subject matter tests (Cochran-Smith, 2006); many states use one or more of the Educational Testing Service's (ETS) Praxis series tests. Nevertheless, research indicates a weak correlation between these tests of content knowledge and teacher effectiveness (Darling-Hammond, 2006; K. J. Mitchell, Robinson, Plake, & Knowles, 2001; NRC, 2001). As Darling-Hammond (2010) argued, "current measures for evaluating teachers are not often linked to their capacity to teach" (p. 2).

Even U.S. Secretary of Education Arne Duncan (2009) underscored the need for better assessments of the pedagogical skills of new teachers when he identified the efforts of the American Association of Colleges for Teacher Education (AACTE) and its 800 colleges and universities to improve student learning through developing a national assessment of teacher candidate readiness, a performance-based assessment modeled after the Performance Assessment for California Teachers (PACT). The PACT, a teaching performance assessment, is designed to measure effective teaching through assessing five domains (with rubrics covering Assessment, Reflection, Academic Language, Planning, and Instruction). The current performance assessment, endorsed by the AACTE and the Teacher Performance Assessment Consortium (TPAC), is known as the edTPA and comprises 33 states and the District of Columbia.[1]

Purpose of the Study

The purpose of this study was to examine what teacher candidates made visible about their practices and understandings of the teaching and learning process in constructing their performance assessments. This study was designed to examine the kinds of teaching practices teacher candidates utilized in the classroom, specifically examining how candidates who scored highest on certain PACT rubrics (in the domains of Assessment, Reflection, and Academic Language) planned instructional supports, assessed, and reflected in ways significantly different than those who scored lowest on PACT rubrics. For this study, we examined 12 performance assessments completed by preservice teachers from a Central Coast California Teacher Education Program. Although various types of assessments are required during this program, PACT offers the most comprehensive evidence of how teacher candidates engage in the practice of teaching and learning after having participated in variety of teacher preparation courses and while completing their fieldwork.

As of July 1, 2008, all candidates admitted to a credential program in California are required to pass a teacher performance assessment (TPA). PACT is an approved TPA, along with edTPA and two others. PACT is subject specific and is "designed to measure and promote candidates' abilities to integrate their knowledge of content, students, and instructional context in making instructional decisions and

reflecting on practice" (Pecheone & Chung, 2007, p. 5). To complete one of the PACT Teacher Events, candidates must submit teacher artifacts and commentaries centered on the five dimensions of teaching: planning, instruction, assessment, reflection, and academic language. Artifacts for the first four dimensions include lesson plans, video clips of teaching, student work samples, and daily reflections on instruction. Academic language is examined through these artifacts and candidates' commentary responses and evaluates "how their lessons and instruction help to build students' acquisition and development of academic language" (Pecheone & Chung, 2007, p. 10), including the vocabulary, symbols, and language demand central to the learning segment.

In this study, two phases of analysis were conducted. The first phase was a discourse analysis that focused on how teacher candidates who scored highest on the performance assessment described their teaching and students' learning in ways that were clearly different than the ways those candidates who scored lowest on the assessment learned. In this phase, we constructed telling cases, a means by which the teacher candidates' discursive choices become descriptions of formerly invisible social conditions (see J. C. Mitchell, 1984). These telling cases support our grounded inferences of how different candidates engaged in and reflected on their teaching and learning practices. In the second phase, we focused more on assessments of candidates who scored highest on the PACT to highlight differences in practices related to academic language development across disciplines. Through these phases of analysis, we addressed the following research questions:

1. What kinds of teaching practices did teacher candidates who scored highest on the Assessment, Reflection, and Academic Language rubrics use? How were these practices or strategies different from the practices or strategies those candidates who scored lowest on the same rubrics used?

2. Are these differences evident across teaching practices for different subject areas?

In answering these research questions, we sought to highlight what distinguished a strong performance assessment from a weaker one, based on scores for the Assessment, Reflection, and Academic Language rubrics. We focused on scores for these rubrics because our candidates consistently receive higher scores in the Planning and Instruction rubrics. We conducted this analysis specifically to inform the design of our own courses here at the university and to better support our teacher candidates going through the performance assessment process. An additional aim of this study is to inform a larger audience of teacher educators utilizing performance assessments to measure teacher candidate learning. As teacher educators and trained PACT scorers at our universities, we drew in this research and subsequent analysis from a number of experiences in working with teacher candidates, including performance assessment and master's project coordinators

and instructors for courses on curriculum design and instruction, English language development and specially designed academic instruction in English, educational psychology, literacy courses, and science methods.

The Performance Assessment for California Teachers

The PACT is a standardized performance assessment that includes Embedded Signature Assignments that vary among institutions.[2] PACT was developed to assess a teacher candidate's ability to plan lessons that provide opportunities for students to learn, design, and analyze assessments; to reflect on what occurred during and as a result of the instruction; and to propose next steps for the students' learning processes.

In the assessment, teacher candidates are required to consider the classroom context in which they are teaching and to plan lessons that are appropriate for their group of students. They also are prompted to provide specific support for English language learners, students with Individualized Education Programs (IEPs) or 504 Plans, or students who may struggle with content. These various documents, descriptions, and explanations are organized into a Teaching Event (TE). The TE comprises five tasks: Task 1 includes the Context for Learning Form and Context Commentary; Task 2 includes lesson plans, instructional materials, and Planning Commentary; Task 3 includes the video of classroom teaching and Instruction Commentary; Task 4 includes the assessment rubric, three student work samples, and the Assessment Commentary; and Task 5 includes daily reflections and the Reflection Commentary. Evidence of attention to academic language development is embedded across each of the tasks.

Trained and calibrated scorers evaluate the candidates' performances; these scorers are mostly faculty and supervisors within their own teacher education programs. Scorers evaluate the PACT TE using 12 four-level rubrics divided by task (Table 1). To pass the TE, a teacher candidate must achieve at least a Level 2 on 10 of the 12 rubrics and not receive two scores of Level 1 within the same task.[3] When a candidate passes PACT, he or she is deemed ready to take over his or her own classroom. Although our data derive solely from the PACT TE, we argue that the findings discussed in this study have implications for any university-based teacher education program that uses or plans to use performance assessments to evaluate teacher candidates' knowledge about teaching and learning.

A Conceptual Framework for Studying Teacher Learning

We understand teacher learning to be a continual process of socially constructed and reconstructed teaching and learning experiences. Two bodies of research inform our view of teacher learning and subsequent study of performance assessments: (a) teacher capacity, or what teachers should know and demonstrate as effective

Table 1
Task Name, Rubric Numbers, and Guiding Questions for the PACT Rubrics

Task	*Rubric*	*Guiding Questions*
Planning	1	How do the plans support student learning of strategies for understanding, interpreting, and responding to complex text? (TPEs 1, 4, 9)
	2	How do the plans make the curriculum accessible to the students in the class? (TPEs 1, 4, 5, 6, 7, 8, 9)
	3	What opportunities do students have to demonstrate their understanding of the standards and learning objectives? (TPEs 1, 5, 11)
Instruction	4	How does the candidate actively engage students in their own understanding of how to understand, interpret, or respond to a complex text? (TPEs 1, 5, 11)
	5	How does the candidate monitor student learning during instruction and respond to student questions, comments, and needs? (TPEs 2, 5)
Assessment	6	How does the candidate demonstrate an understanding of student performance with respect to standards/ objectives? (TPEs 1, 3)
	7	How does the candidate use the analysis of student learning to propose next steps in instruction? (TPEs 3, 4)
	8	What is the quality of feedback to students? (TPEs 3, 4)
Reflection	9	How does the candidate monitor student learning and make appropriate adjustments in instruction during the learning segment? (TPEs 2, 10, 12, 13)
	10	How does the candidate use research, theory, and reflections on teaching and learning to guide practice? (TPEs 10, 11, 12, 13)
Academic Language	11	How does the candidate describe the language demands of the learning tasks and assessments in relation to students at different levels of English language proficiency? (TPEs 1, 4, 7, 8)
	12	How do the candidate's planning, instruction, and assessment support academic language development? (TPEs 1, 4, 7, 8)

teaching, including support for students' academic language development, and (b) performance assessments and what they reveal about candidates' teaching practices and understanding of their practices.

Teacher Capacity

Grant (2008) defined *teacher capacity* as "a teacher's knowledge, skills and dispositions" (p. 127). McDiarmid and Clevenger-Bright (2008) discussed the evolving and expanding conceptions of teacher capacity, drawing attention to the role of teachers' subject matter knowledge and responsibilities for providing access to all students. These conceptions of teacher capacity have advanced from a skill-focused view or "old formula of knowledge, skills, and dispositions" to include a more collaborative framing of circumstances, events, and problems teachers encounter (p. 147). Feiman-Nemser (2001) argued that central to continuation of this teacher learning process are five tasks that build on ideas about what teachers need to know and be able to do. To not confuse tasks as outlined in PACT with Feiman-Nemser's tasks, we refer to the latter as practices in describing her framework. Practices most significant to this study include Practice 3, introducing perspectives on development and learning to "provide necessary frameworks for understanding students, designing appropriate learning activities, and justifying pedagogical decisions and actions" (p. 1018), and Practice 5, providing teacher candidates with opportunities to observe, interpret, and analyze, as with "analyzing student work, comparing different curricular materials, . . . and observing what impact their instruction has on students" (p. 1019). Building on the significance of teacher learning, Cochran-Smith (2005) argued that the most defining goal of teacher education should be a focus on student learning. Furthermore, Darling-Hammond (2006) argued that the most important questions for teacher educators concern the relation between what teachers have learned and how it influences what their pupils learn.

The National Academy of Education Committee on Teacher Education (NAECTE; 2007) described effective teachers as those who use a variety of different tools to assess how students learn in addition to what students know. Effective teachers design lessons based on students' prior knowledge and level of development and adapt the curriculum to students' needs. They also engage students in active learning (as with debating, discussing, researching, experimenting, etc.). Aside from defining an effective teacher, the authors of NAECTE also explain that teacher education programs should be structured in ways that enable candidates to learn about practice in practice, by bridging learning experiences on campus to those taking place in the school classroom, to lay a foundation for lifelong learning. In other words, teacher research and performance assessments should relate teacher learning to classroom practice (see also Darling-Hammond, 2000, 2010) and should help candidates develop habits of reflection and analysis, which may be utilized once they have completed a particular preservice program (NAECTE, 2007). Specific characteristics that define teacher capacity and teacher effectiveness

are related to teacher candidates' abilities to understand their students, design appropriate learning activities, justify pedagogical decisions (Feiman-Nemser, 2001), and adapt the curriculum to students' needs (NAECTE, 2007). Building on these characteristics and others, we argue that central to a teacher candidate's success in the classroom is his or her ability to provide students with opportunities to develop the academic language in the specific discipline.

The academic language framework used in our teacher education program at the time these assessments were completed centered on the work of Dutro and Moran (2003). Dutro and Moran included a simplified description of academic language as the "language of texts, of academic discussion, and of formal writing" (p. 231). These may include justifying evidence, generating hypotheses, summarizing, evaluating information, defining causal relationships, and comparing and synthesizing information (see Chamot & O'Malley, 1994; Dutro & Moran, 2003). Academic language includes how (*forms*) we use language to accomplish academic purposes (*functions*) inside and outside of the classroom. Language *functions* are expressed through *forms*. Forms can include discipline-specific content vocabulary, which may take on different meanings depending on the discipline. Dutro and Moran (2003) distinguished between two different but interrelated types of forms. Using an architectural metaphor, they defined content-specific vocabulary terms as "brick terms" and the linguistic or grammatical structures that show relationships among words as the "mortar" terms. The brick and mortar terms and phrases work in tandem to express ideas.

Current research has indicated that teacher candidates are able to apply language objectives, functions, and language structures to their lesson plans (Scalzo, 2010) and that they are able to articulate different levels of understanding and advocate for a variety of instructional supports for English learners (Bunch et al., 2009). Furthermore, students can use academic language in the classroom, but only when instructional support is provided (Fillmore & Snow, 2002; Schleppegrell, 2004). However, as Grant (2008) argued, absent from much of the scholarship on teacher capacity is research on how teacher capacity relates to knowledge and skills for teaching diverse groups of students. Nevertheless, Dutro and Moran's (2003) approach takes a structural view of language, and no evidence exists that students studying explicit forms develop language fluency (see Valdés, Capitelli, & Alvarez, 2011). As such, we are exploring more recent and sophisticated approaches to studying academic language development in our work with preservice teachers (see, e.g., Bailey, 2007; Bailey & Butler, 2003; Bunch, 2013; Arias & Faltis, 2013). These include opportunities for teachers to develop pedagogical language knowledge (Galguera, 2011); to facilitate students' academic language development within the fabric of everyday classroom interactions, not separated from social language (Faltis, 2013); and to develop academic language proficiency tests to better understand language usage in academic settings (Bailey, 2007; Bailey & Butler, 2003). Specific approaches to students' development of academic language as a social

practice are defined in terms of "learning to talk science" (Lemke, 1990) and using language recognized by social scientists (Short, 1994), or what De Oliveira (2013) called *history discourse*, including presenting and interpreting historical events (Schleppegrell, 2004). In the field of mathematics, students use structures as well as language in developing a *mathematics register* (Middleton, Llamas-Flores, & Guerra-Lombardi, 2013). Educators have also used *systemic functional linguistics* to understand the importance of language forms for meaning making (Halliday, 1985; Halliday & Webster, 2004). In particular, according to Faltis (2013),

> if teachers could learn about language formations within different academic disciplines and teach students to recognize and use these patterns, students would have more access to the academic content because the features of language in academic contexts would become transparent. (p. 20)

Performance Assessments

Performance assessments are used for a variety of reasons and have been identified as a valuable tool for evaluating teacher preparation (Darling-Hammond, 2010). By utilizing performance assessments, educators can be more flexible in how they design and implement their teacher education programs. Furthermore, performance assessments most closely align with evaluating what teachers actually do (Arends, 2006; Darling-Hammond, 2010). Context is important—just as with students at any level, "the learning varies with individual learners and their aspirations and abilities" (Arends, 2006, p. 20). Self-reported data by teachers who found completing the PACT to be valuable indicated that the assessment was helpful for sequencing lessons, evaluating what students were learning (and not learning), and reflecting on how to use that understanding to prepare for the next lesson. Teachers reported that completing this assessment continued to influence their teaching practices during their first year of teaching (Darling-Hammond, 2010).

Research on what performance assessments make visible about effective teaching indicates that candidates use multiple representations to make language and mathematical concepts comprehensible, they promote and facilitate the use of mathematical vocabulary and discourse, they used a variety of participation structures, and they supported the use of students' native languages (Bunch et al., 2009). Although this study examined how the teacher candidates designed and implemented their lessons, other researchers have focused on the reflection task and have found that candidates made "a shift from inner reflection to a more critically reflective practice grounded on the examination of artifacts and reasoned discourse about such inquiry" (Nagle, 2009, p. 4). This process moved the discourse away from what Nagle called "war stories" or other personal stories toward more analytical and productive conversations about teaching practices. Teacher candidates also integrate aspects of these conversations into their work and create an expectation that reflective practice is part of everyday teaching practice.

Methodology and Data Collection

For this article, we qualitatively analyzed 12 performance assessments submitted by teacher candidates who were part of the secondary teaching cohort in the 2009-2010 academic school year. Table 2 charts the sum of each candidate's scores. So that pseudonyms are easy to distinguish, we used names that start with the letter H to represent candidates that received the "highest" set of scores. Names that start with L were chosen to represent candidates that received the "lowest" set of scores, and names that started with the letter N were chosen to represent candidates that received the "next lowest" set of scores. As represented in Table 2, candidates who received the highest scores earned between 20 and 25. Because there are seven rubrics, the teacher candidates who submitted strong assessments received, on average, a score of Level 3 for each rubric. Teacher candidates with the lowest scores received a total between 11 and 15, an average at or below a Level 2 on each of the seven rubrics. Nevertheless, all of the performance assessments we analyzed received passing scores (see section "The Performance Assessment for California Teachers" for the California passing standard).

After selecting the assessments we wanted to examine further, we read through the performance assessments and highlighted key characteristics and practices associated with effective teaching (see Darling-Hammond, 2000; Feiman-Nemser, 2001; NAECTE, 2007). Specifically, we coded the candidates' discursive choices about the pedagogical skills they incorporated into their lessons, how they monitored student learning, how they interpreted and used assessments, how they made content accessible, and how they scaffolded for English learners (ELs) and others they identified as struggling with the content. After coding the teacher candidates' descriptions, we charted them by task (Assessment, Academic Language, or Reflection) and then analyzed the charts to tease out patterns of differences between

Table 2
Highest and Lowest Set of Total Scores for Rubrics 6-12

	Subject area	*Sum of scores for Rubrics 6-12*
Holly	Science	23
Nancy	Science	14
Laura	Science	13
Hannah	Mathematics	25
Norah	Mathematics	15
Lucy	Mathematics	11
Heather	English language arts	20
Nikki	English language arts	14
Larry	English language arts	13
Henry	History/social science	21
Nadia	History/social science	14
Luke	History/social science	13

the strong and weak performance assessments for each subject and, subsequently, looked for evidence of these patterns across the TEs in the four different subject areas. During this analysis, we found five differences in how teacher candidates planned, assessed, and reflected on their teaching in the "strong" versus "weak/poor" performance assessments.

On the basis of these findings, we constructed telling cases (see J. C. Mitchell, 1984) of those five differences, which make transparent strong practices versus weaker practices in relation to characteristics of effective teaching. These telling cases make visible the spectrum of understandings constructed from the ideal or planned opportunities, as available through participation in the teacher education program, to the more situational opportunities made available in the different classroom contexts and through the feedback provided by the different school supervisors, among others. During the analytic process, we make visible how teacher candidates' actions and discursive choices are representative of teachers who either scored highest or scored lowest or next lowest on the PACT. In the second phase of analysis, we contrasted teaching practices of those who scored highest, specifically looking at how candidates engaged students in academic language opportunities to determine if differences across disciplines were evident.

Findings

Comparison of Teaching Practices

Table 3 presents the breakdown of the different class contexts, as described in the performance assessments we analyzed, according to grade level, subject area, number of students in the class, and number of students who were designated EL and who had IEPs or 504 Plans. From Table 3, it is important to note that these class contexts represent a range of grade levels and subject areas, and all classes had at least one student who was designated EL according to his or her performance on the California English Language Development Test. Furthermore, students in these classes represent a diverse population of students, both linguistically and culturally, and thus a rich source of data collection and complex learning environments.

Key teaching actions. Through analysis of the performance assessment documents, we found five key teaching actions that distinguished a strong TE from one that received the lowest or next lowest set of passing scores. Findings from this phase of analysis are as follows:

1. Teacher candidates who did well on the PACT used formative assessments to *monitor students' understanding toward meeting the standards and learning objectives* (see Practice 3 in Feiman-Nemser, 2001) *and language objectives*; whereas teacher candidates who did not do as well used formative assessments to *determine if students were on task* or to *monitor behavior*.

2. Teacher candidates who did well on PACT used assessment criteria that *focused on content and language objectives* (see Practice 5 in Feiman-Nemser, 2001), whereas teacher candidates who did not do as well *focused primarily on completion of the task and the grammatical or mechanical elements of writing*.

3. Teacher candidates who did well on PACT utilized scaffolding that *supported students' ability to build academic language fluency* (see Practice 5 in Feiman-Nemser, 2001), whereas teacher candidates who did not do as well often planned supports that *constrained what the students were able to discuss in their assignments*.

4. Teacher candidates who did well on PACT *provided different types of support for academic language development* and were *able to articulate why these strategies are likely to support the development of the students' understandings* of the course content (see Practice 3 in Feiman-Nemser, 2001), whereas the teacher candidates who did not do as well planned

Table 3
Breakdown of Class Contexts for Each Teaching Event Analyzed

	Science			Math	English language arts			History/social science				
	Holly	Nancy	Laura	Hannah	Norah	Lucy	Heather	Nikki	Larry	Henry	Nadia	Luke
Grade[a]	8	8	7	9 (22), 10 (5), 11 (1)	10 (1), 11 (24), 12 (5)	8	11 (28), 12 (2)	8	9	11	12	7
Subject area[b]	CP phys.	phs. sci.	Gate life sci.	CP geom.	pre-calc	alg. 2	English/ Am. lit.	CP Engl.	CP Engl.	CP U.S. hist.	CP econ.	CP world hist.
Focus of lessons	how forces contri-bute to velocity	stars and planets	struc./ func-tion of DNA	polygons/ parallelograms	combining functions	slope as repre-. senta-tion of rate	editorials	poetry inter-pre-tation	dia-logue	mino-rities and WWII	circu-lar flow and macro-econ-omics	justice and feudal life
No. students	32	27	28	28	30	29	30	29	24	34	34	32
No. designated ELs[c]	6 RF; 12 LP	1 AD; 13 RF	1 AD; 7 RF	1 AD; 4 RF	8 RF; 5 LP	3 AD; 1 EA	8 EL; 9 RF	7 AD; 5 EA	10 IA; 2 RF	5 AD; 2 IA	6 RF; 12 LP	9 AD; 3 EA
No. IEPs or 504 Plans	N/A	1 with IEP	1 w/ 504 1 w/ IEP	N/A	N/A	2 with 504	1 w/ 504 4 w/ IEPs	1 w/ IEP	N/A	7 w/ 504 1 w/ IEP	8 w/ IEPs	N/A
IEP	–	1	1	–	–	0	4	1	–	1	8	–
504	–	0	1	–	–	2	1	0	–	7	0	–

[a] CP = college preparatory.
[b] For multiple grades, number of students per grade in parentheses.
[c] EL = English learner; RF = reclassified fluent; LP = limited proficiency; AD = advanced; EA = early advanced; IA = intermediate advanced. Information provided is based the California English Language Development Test.

support that *focused on pronunciation and repetition of words and/or definitions*.

5. Teacher candidates who did well on PACT *discussed next steps that focused on reteaching, review, and using different strategies and/or assessments* (see Practice 5 in Feiman-Nemser, 2001), whereas the teacher candidates who did not do as well *reiterated what they already did or explained that they would spend more time presenting information in essentially the same way*.

On the basis of these findings, we shifted our analysis to constructing telling cases to further illustrate these differences. In constructing these telling cases, we described the differences in more detail and highlighted the discursive choices made by teacher candidates who earned higher scores on the PACT versus by those who scored lowest and next lowest on the PACT.

Use of formative assessments. Each of the 12 teacher candidates used the term *formative assessment* throughout his or her lesson plan. The teachers also addressed how they used formative assessments as prompted by a question in the Planning Commentary, which asks, "Explain how the collection of assessments from your plan allows you to evaluate your students' learning of specific student standards/objectives and provide feedback to students on their learning." Candidates who scored highest on PACT used a variety of different types of formative assessments, including a preassessment that informed the design of their lessons, an observation checklist or a description of how the candidate would circulate the room and ask specific questions to check for understanding, and strategies for requiring students to explain their answers. For example, Holly's students completed a vocabulary preassessment that required them to write definitions and draw pictures of science concepts. She used this preassessment to determine if and how students understood the concepts before she began her new unit. In addition, Heather preassessed her students' academic language development, reading comprehension, and writing skills by having them analyze a political cartoon (before moving to more complex editorials), and Hannah preassessed her students' ability to identify polygons and describe in writing why a particular shape was considered a polygon or not. Henry had his students complete a K-W-L chart to determine what they "knew" and what they "wanted to know" about the study of minorities during World War II (what students "learned" would be assessed at the end of the unit). Conversely, only one of the other eight teacher candidates (who scored lowest or next lowest) stated that she used a preassessment. Laura mentioned that she gave a pretest but provided no explanation of what she required students to do and what she wanted to know about the students' prior understanding before planning her lessons.

In addition, all of the 12 candidates stated that they would circulate the classroom during instructional time and/or would ask questions during the lessons as

part of formative assessments. Candidates who scored highest provided examples of the specific questions they planned to ask. For example, on Day 5, Hannah explains that she will be walking around the classroom observing students' work and looking for specific evidence of understanding:

> I will be looking for correct answers as indications that they [the students] are applying the properties of parallelograms. I also will be looking to see if students are labeling the parallelograms using the properties of the parallelograms. If students are writing incorrect measurements, I will know they require additional instruction with applying the properties of a parallelogram.

Hannah describes questioning strategies she will incorporate to make visible students' understanding, but also to extend their understanding of the concepts. On Day 2 of her lesson plans, she states,

> I will ask them to identify polygons and then follow their statements with "why" questions to dig into their thinking. I anticipate students will have questions about the concepts of irregular, regular, and specific names for polygons. I plan to address these questions during a PowerPoint presentation [of different photo examples of polygons in sporting, travel, recreation, and home contexts] and as they arise during class. During the angle sums investigation, I will be asking students to respond to "why did you draw the triangles that way," "how many triangles are there," and "what would the sum be if . . ." to assess their knowledge of interior angle sums and progress toward the content standard and cognitive objectives for this lesson.

Based on these excerpts, it is clear Hannah was monitoring students' understandings by looking at how they engaged in the math activities. She was looking to see if students were labeling and measuring the shapes correctly but was also asking students open-ended questions that required them to discuss how and why they were approaching the assignments or questions in certain ways.

Other teacher candidates who scored lowest and next lowest also stated that they would be circulating the room during instructional time, but they described this formative assessment process as a way to monitor if students were "on task." For example, Laura explained that she would assess students on their ability to work in groups and the amount of input from each student. Nancy stated that she would circulate the room giving "tickets" to students who were participating, a reward they could exchange for a prize later in the week. In addition, Lucy states she would ask students questions "to check understanding of previously covered topics" and "to see how each group is doing," but she does not describe the strategies she might use or the specific questions she would ask. Interestingly, however, she explains the importance of positive feedback during her daily formative assessments. In every day of her lesson plans, she writes,

> Throughout all stages of the formative assessments, positive feedback will be critical. When students volunteer answers in class, they need to be encouraged and

> praised. I also plan to make positive comments to the students who are working hard on the class work and are on the right track. The students who are struggling will need encouragement as well. Often they only need a small nudge to get them back on track, and it is helpful for them to see how close they already are.

Thus "working hard" and being "on the right track" indicate good behavior but also understanding of the material. Although providing positive feedback is important, it should not be the focus of the formative assessment description.

Assessment criteria. All 12 candidates provided a rubric or set of criteria they used to evaluate whether students met, did not meet, or exceeded the learning objectives. Differences between highest scoring and lowest scoring performance assessments illuminated differences in how criteria were measured and aligned with respect to the learning objectives and related standards. Candidates who scored highest detailed criteria for evaluating whether students properly used academic language (Hannah, Holly, Henry), were able to make strong claims and support those claims with evidence from the text (Heather, Henry), accurately represented the history of the time (Henry), and provided justification for their answers (Hannah, Heather, Holly). They also were able to explain how certain students demonstrated limited or partial understanding. Heather's lessons focused on the analysis of rhetorical devices, structure, and techniques by which authors and speakers convey meaning. In particular, she addressed standards on structural features of informational materials and expository critique. Heather's summative assessment criteria included (a) selecting an editorial cartoon or written editorial suitable for analysis; (b) understanding lesson concepts (persuasive techniques, rhetorical devices, point of view, strengths and weaknesses of arguments); and (c) structuring written responses. These criteria were based on a three-level scale, with a score of 1 for below standards, a 2 for meets standards, and a 3 for exceeds standards. Heather's specific criteria focused on whether the students were able to "identify a writer's stance," the "persuasive techniques that were used in the editorial," and if and "how students warranted their claims with evidence."

Teacher candidates who scored lowest or next lowest on the performance assessment rubrics listed criteria such as completion of the handouts (Norah, Laura, Lucy, Luke), length of writing assignments (Nikki, Larry, Luke), and whether answers were correct or not correct (Nadia, Norah, Nikki, Lucy, Laura, Luke). For example, Larry required students to create a dialogue about a particular experience they had had, and his four assessment criteria included (a) length of dialogue, (b) character traits, (c) character emotion, and (d) other information included.

Larry's criteria were based on a four-level scale. Although these levels were not labeled in the assessment documents, 1 typically denotes not meeting the standard and 4 is exceeding the standard, with scores of 2 and 3 meeting parts of the standards. To assess how students met each of these four criteria, Larry used quantitative measures for the first two criteria (length of dialogue and character traits). For

example, a Level 3 on the rubric was qualified as "the dialogue having 6-7 lines" (length) and "describing at least 1 character trait per character." The third criterion, character emotion, was measured explicitly by whether the student "identified character emotions through the use of tags, structure, and diction." Furthermore, when he discussed what he learned from analyzing students' assessments (a prompt in the Assessment Commentary), he explained that some of the students "did not meet his expectations and did not punctuate their dialogue correctly" or "did not write text that conveyed emotion" but were able to "show traits of the characters in their dialogue." Larry never explained how some students were able or not able to write text that conveyed emotion or that illustrated character traits.

Academic language framework. All candidates used a functional approach to academic language through the use of functions, forms, and fluency as discussed in Dutro and Moran (2003), but how they employed this approach and what they sought to accomplish differed. Each candidate listed key vocabulary students needed to know to demonstrate their understanding. One candidate, Nancy, actually differentiated between "brick" and "mortar" terms. Eleven out of the 12 candidates identified language demands, such as describe, explain, convert, summarize, and ask/answer clarifying questions, among others. The only candidate who did not identify a language demand was Lucy.

Most teacher candidates used sentence frames to support students' ability to build academic language fluency. Nine out of 12 candidates listed at least one sentence frame that students could use to construct arguments and/or provide explanations for what they understood about the content. Through our contrastive analysis, we found clear difference between how the candidates constructed the sentence frames. For example, Larry, Luke, Nancy, and Nikki provided sentence frames that were more like fill-in-the-blank sentences, which constrained what students could say or write and thus how the students could explain their understanding. Nancy listed her sentence frame as follows: "The structure of DNA is a _____ _____, which is shaped like a _____ _____." To address the blanks in this sentence, students needed to determine what two words fit into those two sets of blanks rather than being able to explain what DNA is. Conversely, Norah, Hannah, Heather, Holly, and Henry used sentence frames that required students to include brick terms or proper nouns but also to provide evidence to explain their understanding of the content. For example, Henry incorporated the following sentence frame: "While all minorities experienced a level of discrimination, I think that _____ suffered the most on the home front during WWII because of _____ and _____."

In this example, the sentence frame allowed students to craft a thesis statement, which they subsequently developed in their essays. This type of support encouraged students to decide and articulate which of three given minority groups (Mexican American, Japanese American, and African American) suffered the most discrimination during World War II and then justify their responses with evidence.

Luke introduced a paragraph frame that essentially provided the entire structure of the written assessment. By discussing "Wilma's" treatment, Luke provided a sample paragraph for the students to follow when crafting their essays and justifying their claims:

> I think Wilma would have received just treatment if the . . . [choose Henry II reforms or Magna Carta] has been in effect. First, she would have received [first piece of evidence] in order to counter the injustice of. . . . Second she would have received [second piece of evidence] in order to counter the injustice of. . . . I believe that if Wilma had lived during the [Reforms of Henry II or the signing of the Magna Carta—choose one] she would have received better treatment based on the evidence that was presented.

This highly structured paragraph frame offers limited opportunities for students to express themselves. Although the candidate's goal was to provide scaffolding, this support was oriented toward a fill-in-the-blank assignment. Students were asked to choose which one, Henry II's reforms or the Magna Carta, would have enabled just treatment of a particular case. This type of support does not require students to understand the laws and apply the information but rather essentially to include one of the laws that would have changed the outcome of the case. Luke did require students to justify their claims with specific evidence. In looking more closely at what the candidate was asking students to do, what becomes visible is how the candidate asked them to choose a specific law that would have changed "X" injustice. Here students needed to understand the injustice and apply which law could have changed the outcome of that injustice. This did require students to examine the laws and apply them to specific cases. Again, the closing sentence in the frame offered only two choices (Magna Carta or Henry II's reforms). This explanation is not a critique of using paragraph frames but rather is an example of how the frame can potentially constrain what can be stated. Furthermore, all candidates who scored highest on the PACT rubrics discussed the use of sentence frames as an "option" for students, whereas candidates who scored lowest and next lowest discussed sentence frames as a requirement for participation in the lessons.

Supports for developing academic language. In addition to the sentence frame, teacher candidates used a variety of other types of support for students' academic language development. Some teacher candidates modeled for their students how to engage with content. Others provided students with graphic organizers. A number of the candidates provided opportunities for students to share answers with partners or in small groups as a way for students to practice explaining, describing, and/or summarizing to a peer or peers, before doing so in front of the whole class. For example, Holly included a think/pair/share activity in her lesson plan on Day 1 and stated,

> I ask students to think of two more examples of forces on their own [after she provides an example] and then share their ideas with a partner—why?—to lower

> the affective filter before I call on them individually to answer the question. It also provides the opportunity for two-way interactions, which supports the need for my EL to build language proficiency.

In addition to building language proficiency, Holly also explained in her Planning Commentary,

> To give my students real hands-on experience with forces, one of the learning tasks involves students working in pairs to build a house of cards. During this time they will see forces in action. To move beyond simply identifying forces and to develop their academic language, students are required to write a paragraph about their experience building a house of cards. They are given key vocabulary words that they must include in their paragraph to explain the forces that are involved in constructing and destroying the structure.

Not only did Holly use different strategies to support her students' academic language development, she articulated why the strategies most likely would support her students. Nancy referenced academic language support in this way:

> Academic language is addressed primarily through repetition of terms. Students are unsure of pronunciation at first, but with repeated practice (I have the students repeat after me) they incorporate the new language.

What is interesting about this example is that Nancy had language objectives that required students to list, name, predict, and summarize, but then she discussed how students would develop academic language proficiency by repeating terms. Nancy was not the only candidate to mention the need for students to repeat terms and definitions as a way to build academic language fluency. In fact, all candidates who scored lowest and next lowest described the need for students to hear the academic language terms and to repeat the terms and/or definitions multiple times, but not necessarily in a sentence form or in the context of the particular lesson.

Adjustments to instruction. The final difference concerns how teacher candidates planned for next steps, as made visible in the Reflection Commentary. Candidates who scored highest on the PACT discussed next steps that described reteaching the lesson using different strategies and/or assessments. Thus the focus was on helping to facilitate their students' understanding. Candidates who scored lowest and next lowest reiterated what they did in the lessons and/or described next steps for teaching in essentially the same ways as before. For example, Hannah explains her next steps in this way:

> In my plans for this learning segment, I would do several things differently. During the constructions lab in Lesson 1, I would have planned to go through the final three tasks with the students in more guided exploration. I feel that doing so would have supported both EL and English-only students since the majority of students had trouble reading and following the written instructions on their own. . . . The second thing I would have planned to do differently is to give students different-

> sized parallelograms during the exploration of the properties of parallelograms in Lesson 3. I feel that this would have supported the idea that the properties work for all the parallelograms and not just a certain parallelogram for all students.

In this excerpt, Hannah describes the need for more guided exploration and giving students different sized shapes so they could make larger connections about properties of parallelograms. Henry discussed the need for students to have a graphic organizer that could be used to improve students' historical analysis and writing skills. Holly talked about using different assessments or incorporating more open-ended questions into the assessments so that students would be able to explain how they know the answer or concept. Heather described the need for her first lesson to have a different political cartoon that was more accessible to the students and having directions in written form, not just in oral form, so that students could better understand the steps for analyzing cartoons and editorials. Conversely, Lucy explained her adjustments this way:

> If I could go back and teach this learning segment again, I would want to go back one more day before Day 1. I think the fact that we started the segment late threw things off for the rest of the week. Because I was not able to thoroughly explain the homework assignment that was due on Day 1, we ended up spending a lot of time going over it, and getting behind schedule. This left less time over the next 3 days for the students to work on the assigned problems.

Although Lucy discussed issues of time management, her plan to adjust instruction focuses on explanation of the homework that was assigned before her lesson segment took place. Instead, she should have explained what students learned or did not learn from her lessons and how she could have made adjustments that provided more access to the content or better supported students' understanding of the concepts.

Differences Across Discipline

In the second phase of analysis, we shifted focus to only those who earned the highest scores in Assessment, Reflection, and Academic Language to determine if there were differences across disciplines—whether the candidates utilized different teaching practices, including planning for students' use of discipline-recognized language in everyday classroom interactions and then assessment of language use in the context of structured academic activities.

Everyday classroom interactions. Everyday classroom interactions are defined as opportunities for students to engage socially while practicing academic language vocabulary in action. Across all disciplines, higher scoring teacher candidates required students to work in pairs or small groups to elicit prior knowledge on concepts (math and science) or primary sources (English Language Arts and History/Social Science), which often occurred at the beginning of a lesson or

unit. Students in Hannah's class shared prior knowledge of polygons, including characteristics of polygons they had identified from home (or somewhere outside the classroom), and then created a definition based on this discussion. Students in Holly's class worked in small groups to share their definitions of speed, velocity, and acceleration in one science lesson and then to share examples of forces with a partner in a subsequent lesson. Students in Henry's class partnered to analyze and discuss primary sources related to the challenges minorities experienced during World War II. These discussions were documented on a graphic organizer that was used when students drafted their "Minority Reports." Students in Heather's class completed a give one/get one handout related to the editorials they chose to analyze for homework. In addition, she had them share the findings from their completed handouts with peers, stating, "I know that a number of students in my class can argue. I also know that many of them have social, political, and legal issues that they have a strong stance on." This activity allowed students to share ideas with multiple partners in a more social context than would be found by debating as a whole class in a more academically structured context.

Academic language in context. Academic language in context can be defined by how students use language in ways that members of the discipline may use language. In other words, in the context of structured lessons, students practice "talking science" or "history discourse," similar to how professionals understand concepts and innovate in their fields. This practice looked very different across disciplines. Hannah's assessment required students to identify and define characteristics of polygons and solve for angles; her lessons also required students to justify their answers and/or explain their solutions. For example, to receive full credit on the summative assessment, students needed to justify whether a figure was a polygon and/or whether it was irregular. Therefore academic language use was found written on the handouts as explanations. Holly planned an inquiry-based lab in which students "engaged in the scientific process by planning and conducting an investigation to test a hypothesis," related to the amount of force on a Hershey's Kiss and how it affects the distance it travels. The candidate assessed students based on "spontaneously using multiple vocabulary words without prompting" and providing accurate definitions. She tracked this participation on a spreadsheet, while circulating the room. Henry defined his assessment as a "constructivist based inquiry . . . engaging the students in inquiries where they are 'doing history,' using their individual strengths and previous knowledge to comprehend the content matter. . . . They were building literacy skills to improve their content knowledge." The rubric he designed assessed students' use of vocabulary words and if they were used in the "correct context." Heather's lesson engaged students in a *Jigsaw* activity where students were assigned to read and analyze one of the editorials provided in a packet, becoming the *expert* on that particular editorial. In groups, students shared individual understandings of the editorial. Students were assessed based on their ability to identify the writer's stance and persuasive technique

to meet the standards, but could *exceed* the standards if he or she "backs up the claims with evidence from the editorial." So academic language use was assessed in writing assignments in the disciplines of math, history, and English language arts, but ELA students could earn a proficient score by choosing the right stance or technique (potentially understanding the definitions of the vocabulary) without having to cite evidence to support their choice. Academic language was assessed verbally, based on students' active participation in the science lab and on written work based on their definitions.

Systemic linguistic functions. Teacher candidates used a variety of different language forms as structures for students to make meaning (Halliday, 1985; Halliday & Webster, 2004). All the candidates introduced sentence frames for students to use to support their construction of thesis statements or other written work. The higher scoring candidates, however, went beyond these scaffolds to support students' learning in other ways. For example, Hannah encouraged manipulative use for construction of meaning by having students use different-sized parallelograms when exploring the properties of these geometric shapes. Students in Henry's class analyzed primary sources and were asked to document evidence drawn from them on a graphic organizer that focused on contrasting events and perspectives. Heather focused on structural features of informational materials and expository critique. Students were introduced to editorials and political cartoons, were required to analyze them, and were ultimately required to identify a persuasive technique being used that "is particularly strong" and then explain how this technique adds to the writer's or artist's argument (for exceeding proficiency). Holly used a variety of formative assessments and open-ended questions to promote meaning making throughout her unit; yet her summative assessment was based on students' recall of definitions and properties of forces. She required students to perform calculations, choose a type of force based on a diagram and "explain how you know," demanding a reiteration of the definition of forces as opposed to proving or disproving a hypothesis with evidence, which is more in line with what scientists do.

Discussion

Through examination of performance assessments, we found that candidates who scored highest on Assessment, Reflection, and Academic Language rubrics included clearly stated formative assessment criteria they used to monitor students' understandings of the content and detailed rubrics that described various levels of proficiency toward meeting the standards or objectives. They also incorporated an Academic Language framework throughout their lessons to support the academic language development of their students and were able to clearly explain why the particular strategies were likely to support their particular classroom context and student demographics. Finally, teacher candidates who submitted the strongest

assessments were able to discuss how they would plan to adjust instruction in the future, based on analysis of what occurred during instruction and of students' formative and summative assessments.

Candidates who scored lowest on those rubrics focused on student behavior and completion of assignments rather than on evidence of learning. These candidates also included what they considered support for developing academic language, but that "support" often constrained what students were able to discuss in their writing, and the support strategies focused mostly on repetition of words and definitions. Finally, those candidates who did not do as well on the performance assessments struggled with the ability to discuss the changes they needed to make to be more effective and reflective teachers, a necessary expectation for everyday teaching practice (Nagle, 2009).

Through closer examination, we found that higher scorers were able to teach beyond vocabulary and mechanics and promote genuine discourse in their discipline to some extent. All higher performing candidates focused on academic language development through student-driven discussions, typically at the beginning of the lessons, when students were pair-sharing or, as in the case of Holly's lessons, during the curling lab conversations. Each of these activities focused on the development of disciplinary knowledge and skills (e.g., forces in science; polygons in math; editorials in ELA; historical primary sources in HSS). These teacher candidates promoted academic language development in contexts recognized as appropriate by professionals in their disciplines. For example, in the science lesson, students engaged in inquiry practices related to forces; in math, students explored the properties of polygons using manipulatives; in ELA, students identified a writer's stance and persuasive technique through examining editorials; and in history, students constructed arguments based on the analysis of political cartoons and other primary source documents. In addition, all teacher candidates used language forms to assist students in understanding discipline-specific content knowledge, but not separated from social language (Faltis, 2013). Henry had students participate in history discourse as a *linguistically responsive* history teacher (De Oliveira, 2013) by making the content accessible to ELs, not by simplifying the texts, but by providing scaffolding strategies for students to make meaning of the text, document evidence, and construct an argument with support from the texts. Hannah facilitated students' use of mathematical vocabulary and discourse (Bunch et al., 2009), while supporting development of their *mathematical register* by requiring students to communicate the reasoning behind mathematical solutions (Garrison, Amaral, & Ponce, 2006; Middleton et al., 2013). Holly did not always use forms that reflected what members of the scientific community would use in their own occupations. For instance, she had her students answer open-ended questions in her summative assessment, but many scientists learn through collaboration with one another and/or from developing a hypothesis and testing it via an iterative process. In fact, there was a discrepancy between what some of the teacher candidates assessed at the end of

the unit and how they planned activities during the unit, which may have occurred because mentor teachers required teacher candidates use a particular summative assessment at the end of a unit.

Okhremtchouk, Seiki, Gilliland, Ateh, Wallace, and Kato (2009) explained the importance of examining "the effects of these assessments [PACT] on teacher candidates in order to further understand and shape programs that prepare candidates for such evaluations" (p. 40). We agree, but just collecting student perspectives is not enough. By analyzing the performance assessments and the candidates' discursive choices, we were able to examine how candidates inscribe their understandings of working with linguistically diverse students (Bunch et al., 2009) and also add to the findings of how teacher candidates are able to use and interpret assessments and reflect on their teaching practices to inform next steps in their instruction. One limitation to this study is the number of performance assessments that we examined. In future research, we will use these findings to analyze more assessments from different content areas, specifically to see if there is a greater influence of the new Common Core State Standards and Next Generation Standards on summative assessment choices.

Conclusion

Although teacher education still has its share of harsh critics, the shift toward solid empirical evidence supporting its effectiveness is growing (Cochran-Smith, 2005). Interviews and surveys of teacher education candidates may provide evidence as to changes in thinking, growth in understanding, and reflection. However, examining the performance assessments constructed by teacher candidates during and about their classroom experiences may offer clearer evidence of changes in their beliefs and understandings about the teaching and learning process in relation to the contexts in which they are working. The use of performance assessments in teacher education programs is not new, but research tends to focus on teacher candidates' *perceptions* of the assessments or the *process* of completing the assessments. This study adds much to the literature on what performance assessments make visible about whether teacher candidates can engage in effective teaching practices and what elements of the teacher education program design need to be further revised and/or developed to strengthen preservice candidates' ability to plan engaging and effective lessons.

We believe the implications of our study are manifold. By understanding what types of teaching practices are more effective than others and how teacher candidates inscribe their understandings of these practices, teacher educators are better able not only to assess teacher candidates but also to model and facilitate highly effective teaching practices. In fact, we argue that any educator responsible for evaluating teacher quality at the preservice level could benefit from these findings, including those who are teaching courses on instructional design, lesson planning, and as-

sessment and those who are supervising and giving feedback to teacher candidates during their fieldwork. Also, experienced teachers working toward national board certification could benefit, as the PACT is partly modeled on the national board of professional teaching standards.[4] In conclusion, while Arne Duncan has stressed the need for a national assessment of teacher candidate readiness and highlighted AACTE efforts with the edTPA, his recent call for the National Council on Teacher Quality (NCTQ) to rate institutions of teacher education has drawn intense criticism. As Linda Darling-Hammond (2013) explained, "NCTQ's methodology is a paper review of published course requirements and course syllabi against a check list that does not consider the actual quality of instruction that the programs offer, evidence of what their students learn, or whether graduates can actually teach." By assessing evidence of teacher learning and performance, one can look beyond the criteria, beyond what is outlined in a lesson plan or syllabus, and better recognize what teachers are understanding about the teaching and learning process, how students are engaging in lessons, and how teachers are determining what students are learning or not learning—evidence that ultimately can be used to improve teacher education programs and classroom learning.

Notes

[1] See http://edtpa.aacte.org/ for more information.

[2] For more information on PACT and its history and the other assessments, visit http://www.pacttpa.org/. See also Darling-Hammond (2010, Table 1) for more details on the dimensions of PACT.

[3] This represents the passing standard for PACT in California. See http://www.pacttpa.org/ for more clarification concerning this passing standard. Also, this does not represent the passing standard for those states implementing TPAC. Each state may set its own standard based on analysis of pilot data.

[4] For more information, visit http://www.nbpts.org/.

References

Arends, R. I. (2006). Performance assessment in perspective: History, opportunities, and challenges. In S. Castle & B. D. Shaklee (Eds.), *Assessing teacher performance: Performance-based assessments in teacher education* (pp. 3-22). Lanham, MD: Rowman & Littlefield Education.

Arias, M. B., & Faltis, C. (2013). *Academic language in second language learning*. Charlotte, NC: Information Age.

Bailey A. (Ed.). (2007). *The language demands of school: Putting academic English to the test*. New Haven, CT: Yale University Press.

Bailey, A. L., & Butler, F. A. (2003). *An evidentiary framework for operationalizing academic language for broad application to K-12 education: A design document* (CSE Technical Report No. 611). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis, 31*(4), 416-440.

Bunch, G. C. (2013). Pedagogical language knowledge: Preparing mainstream teachers for English learners in the new standards era. *Review of Research in Education, 37*, 298-341.

Bunch, G. C., Aguirre, J. M., & Téllez, K. (2009). Beyond scores: Using candidate responses on high stakes performance assessment to inform teacher preparation of English learners. *Issues in Teacher Education, 18*(1), 103-128.

Chamot, A. U., & O'Malley, M. J. (1994). *The CALLA handbook*. Reading, MA: Addison-Wesley.

Clift, R., & Brady, P. (2005). Research on methods courses and field experiences. In M. Cochran-Smith & K. Zeichner (Eds.), *Studying teacher education: The report of the AERA Panel on Research and Teacher Education* (pp. 261-308). Mahwah, NJ: Lawrence Erlbaum Associates.

Cochran-Smith, M. (2005). Studying teacher education: What we know and need to know. *Journal of Teacher Education, 56*(4), 301-306.

Cochran-Smith, M. (2006). Taking stock in 2006: Evidence, evidence everywhere. *Journal of Teacher Education, 57*(1), 6-12.

Cochran-Smith, M., & Zeichner, K. (2005). *Studying teacher education: The report of the AERA Panel on Research and Teacher Education*. Mahwah, NJ: Lawrence Erlbaum Associates.

Darling-Hammond, L. (2000). *Studies of excellence in teacher education*. Washington, DC: American Association of Colleges of Teacher Education.

Darling-Hammond, L. (2006). Assessing teacher education: The usefulness of multiple measures for assessing program outcomes. *Journal of Teacher Education, 57*(2), 120-138.

Darling-Hammond, L. (2008). The case for university-based teacher education. In M. Cochran-Smith, S. Feiman-Nemser, J. D. McIntyre, & K. E. Demers (Eds.), *Handbook of research on teacher education* (pp. 333-346). New York: Routledge.

Darling-Hammond, L. (2010, October). *Evaluating teacher effectiveness: How teacher performance assessments can measure and improve teaching*. Center for American Progress. Retrieved from http://www.americanprogress.org/issues/2010/10/teacher_effectiveness.html

Darling-Hammond, L. (2013, June 18). Why the NCTQ teacher prep ratings are nonsense. *Washington Post*. Retrieved from http://www.washingtonpost.com/blogs/answer-sheet/wp/2013/06/18/why-the-nctq-teacher-prep-ratings-are-nonsense/

Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment of teaching in context. *Teaching and Teacher Education, 16*(5-6), 523-245.

De Oliveira, L. C. (2013). Academic language in the social studies for English learners. In M. B. Aria & C. Faltis (Eds.), *Academic language in second language learning* (pp. 149-170). Charlotte, NC: Information Age.

Duncan, A. (2009, October 22). Teacher preparation: Reforming the uncertain professional. Remarks of Secretary Duncan at Teachers College, Columbia University.

Dutro, S., & Moran, C. (2003). Rethinking English language instruction: An architectural approach. In G. Garcia (Ed.), *English learners: Reaching the highest level of English literacy* (pp. 227-258). Newark, NJ: International Reading Association.

Faltis, C. (2013). Demystifying and questioning the power of academic language. In M.

Arias & C. Faltis (Eds.), *Academic language in second language learning* (pp. 3-26). Charlotte, NC: Information Age.

Feiman-Nemser, S. (2001). From preparation to practice: Designing a continuum to strengthen and sustain teaching. *Teachers College Record, 103*(6), 1013-1055.

Fillmore, L. W., & Snow, C. E. (2002). What teachers need to know about language. In C. T. Adger, C. E. Snot, & D. Christian (Eds.), *What teachers need to know about language* (pp. 7-54). Washington, DC: Centered for Applied Linguistics and Delta Systems.

Galguera, T. (2011). Participant structures as professional learning tasks and the development of pedagogical language knowledge among preservice teachers. *Teacher Education Quarterly, 38*(Winter), 85-106.

Garrison, L., Amaral, O., & Ponce, G. (2006). UnLATCHING mathematics instruction for English learners. *MCSM Journal for Mathematics Education Leadership, 9*(1), 14-24.

Grant, C. A. (2008). Teacher capacity: Introduction to the section. In M. Cochran-Smith, S. Feiman-Nemser, J. D. McIntyre, & K. E. Demers (Eds.), *Handbook of research on teacher education* (pp. 127-133). New York: Routledge.

Grossman, P. (2008). Responding to our critics: From crisis to opportunity in research on teacher education. *Journal of Teacher Education, 59*(1), 10-23.

Halliday, M. A. K. (1985). *An introduction to functional grammar*. London, UK: Edward Arnold.

Halliday, M. A. K., & Webster, J. J. (2004). *The language of science*. London, UK: Continuum.

Hollins, E., & Guzman, M. (2005). Preparing teachers for diverse populations. In M. Cochran-Smith & K. Zeichner (Eds.), *Studying teacher education: The report of the AERA Panel on Research and Teacher Education* (pp. 477-548). Mahwah, NJ: Lawrence Erlbaum Associates.

Lemke, J. (1990). *Talking science: Language, learning, and values*. Norwood, NJ: Ablex.

McDiarmid, G. W., & Clevenger-Bright, M. (2008). Rethinking teacher capacity. In M. Cochran-Smith, S. Feiman-Nemser, J. D. McIntyre, & K. E. Demers (Eds.), *Handbook of research on teacher education* (pp. 134-156). New York: Routledge.

Middleton, J. A., Llamas-Flores, S., & Guerra-Lombardi, P. P. (2013). English language learning and learning academic language in mathematics. In M. B. Aria & C. Faltis (Eds.), *Academic language in second language learning* (pp. 201-224). Charlotte, NC: Information Age.

Mitchell, J. C. (1984). Typicality and the case study. In R. F. Ellens (Ed.), *Ethnographic research: A guide to general conduct* (pp. 238-241). New York: Academic Press.

Mitchell, K. J., Robinson, D. Z., Plake, B. S., & Knowles, K. T. (2001). *Testing teacher: The role of licensure tests in improving teacher quality*. Washington, DC: National Academy Press.

Nagle, J. (2009). Performance assessment for California teachers (PACT) Teaching Event: Using a state accountability measure of teaching to enhance critically reflective practice. Paper presented at the annual meeting of the American Association of Colleges for Teacher Education. Retrieved from http://www.allacademic.com/meta/p35871_index.html

National Academy of Education Committee on Teacher Education. (2007). *A good teacher in every classroom: Preparing the highly qualified teachers our children deserve* (L. Darling-Hammond & J. Baratz-Snowden, Eds.). San Francisco, CA: Jossey-Bass.

National Research Council. (2001). *How people learn: Brain, mind, experience, and school*.

Washington, DC: National Academy Press.

National Research Council. (2010). *Preparing teachers: Building evidence for sound policy*. Washington, DC: National Academy Press.

Okhremtchouk, I., Seiki, S., Gilliland, B., Ateh, C., Wallace, M., & Kato, A. (2009). Voices of pre-service teachers: Perspectives on the performance assessment for California teachers (PACT). *Issues in Teacher Education, 18*(1), 39-62.

Pecheone, R. L., & Chung, R. R. (2007). *Technical report of the Performance Assessment for California Teachers (PACT): Summary of validity and reliability studies for the 2003-2004 pilot year*. Stanford, CA: PACT Consortium.

Scalzo, J. N. (2010). The social construction of academic language in teacher education: Preparing preservice teachers to work with English language learners (Unpublished doctoral dissertation). Retrieved from *ProQuest Dissertations & Theses* database. (UMI No. 3437286)

Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistic approach*. Mahwah, NJ: Lawrence Erlbaum Associates.

Short, D. (1994). Expanding middle school horizons: Integrating language, culture, and social studies. *TESOL Quarterly, 28*(3), 581-608.

Snyder, J., Lippincott, A., & Bower, D. (1998). The inherent tensions in the multiple uses of portfolios in teacher education. *Teacher Education Quarterly, 25*(10), 45-60.

Valdés, G., Capitelli, S., & Alvarez, L. (2011). Realistic expectations: English language learners and the acquisition of "academic" English. In G. Valdés, S. Capitelli, & L. Alvarez (Eds.), *Latino children learning English: Steps in the journey* (pp. 15-41). New York: Teachers College Press.