
Title	Observation on teacher quality: Critical analysis on observational instruments in pre-service teacher performance assessment
Author(s)	Samantha Caughlan and Heng Jiang
Source	<i>Journal of Teacher Education</i> , 65(5), 375-388. http://dx.doi.org/10.1177/0022487114541546
Published by	SAGE Publications

This is the author's accepted manuscript (post-print) of a work that was accepted for publication in the following source:

Caughlan, S., & Jiang, H. (2014). Observation on teacher quality: Critical analysis on observational instruments in pre-service teacher performance assessment. *Journal of Teacher Education*, 65(5), 375-388. <http://dx.doi.org/10.1177/0022487114541546>

Notice: Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source.

© 2014 SAGE Publications

Abstract

Teacher preparation programs commonly use observational instruments to assess the progress and the exit performances of teacher candidates. However, while these instruments have been described and several have been studied for effectiveness, the field lacks a close examination of how they position participants: teacher candidates, K-12 pupils, and teacher educators.

This paper closely examines three classroom observation instruments used in preservice programs. We use critical discourse analysis and systemic-functional linguistics to examine how the grammar of these instruments assigns agency and positions participants as teachers and learners, and define their larger discourses of professionalism and accountability.

We argue that instruments differ in the extent to which they grant participants agency, thus influencing the assumed pedagogical relations among the teacher educator, teacher candidate, and K-12 pupils. Instruments are not neutral, but reflect the values of the programs that use them, inflected by often contradictory discourses of teacher and student learning.

Introduction

The assessments that serve as gate-keeping mechanisms for any profession reflect the field's image of valued professional knowledge and practice, as well as the roles the professional takes in relation to clients and the larger social order. As assessments for new teachers shift from an old model of input measures (grade point average, credits taken, content knowledge examinations) to output measures (performance assessments, portfolios, and observations of teaching practice), what valued knowledge, performance, and professional roles are implied? What is our image of a novice professional? This study addresses these questions by comparing the valued knowledge, activities, and participant roles reflected in three observation instruments used as performance assessments in teacher education programs: the Christopher Newport University Student Teacher Observation Form (CNU), the Michigan State University Field Instructor Feedback Form (FIFF), and the Performance Assessment for California Teachers and Pre-service Teachers (PACT).

We use the framework of critical discourse analysis (Fairclough, 2003; Rogers, 2011) to examine these three instruments by carefully reading and noting the features that imply what is valued as quality teaching practice. We argue that the observational instruments employed as tools to assess the progress and the exit performances of teacher candidates are not neutral, but reflect the values of the programs that use them through particular (and sometimes contradictory) discourses of teacher learning and student learning. These values are also reflected in assumed relations among the teacher educator, teacher candidate, and the pupils in the classroom. Because performance assessments are regarded as part of the learning process for teacher candidates (Porter, Youngs, & Odden, 2001), it is important to examine how teacher candidates are positioned as learners, when they are performing the teacher's role, by performance assessments designed to guide and evaluate candidates'

teaching and expected teacher-student interactions. Thus, this study focuses on the positioning of the teacher candidate and her students as well as the pedagogical relationship between the teacher educator and the candidate afforded by the texts of performance assessment instruments.

Specifically, the research questions that guided this inquiry were:

1. To what knowledge and performances do observation instruments orient the observer's attention?
2. How do these instruments reflect the values of the program and larger educational discourses?
3. How are teacher candidates, their students, and teacher educators positioned in relation to each other by the textual features of observational instruments?
4. How are professionalism and agency implicitly defined by these instruments?

This study is significant for several reasons. First, we examine the expected performances described by observational instruments to reveal naturalized values about teaching and learning embedded in social, cultural and political contexts. Second, the instruments' projection of professionalism is revealed through examining how the grammar of performance assessment instruments positions participants and assigns agency. Third, our analysis provides a picture of what is assumed about teacher and student learning in observational instruments. Finally, although the findings of this study are particular to these instruments, recent changes in state licensure and national accreditation requirements have resulted in the wider use of observational instruments, thus making this study interesting to a larger audience.

Performance Assessment of Teacher Quality

Although many factors contribute to student academic achievement in schools, there is increasing agreement among educational researchers that teacher quality may be the most significant factor (Harris & McCaffrey, 2010; Sanders & Horn, 1998). Other researchers point out that what “quality” is and how best to measure it is debatable (Cohen, 2010; Kennedy, 2010a). In 2003, the US Department of Education defined “highly qualified” as teachers with bachelor’s degrees, certification and a major in the courses they teach. However, there is growing agreement that teacher quality as defined by those measures is only weakly linked to student achievement. Such “input measures” are currently being replaced or supplemented by “output measures” of teacher quality, such as performance assessments (Campbell et. al., 2000). That is, teacher quality is measured in terms of how a teacher performs in the classroom and how the teaching practice results in the students’ learning achievement. In the past decade, teacher performance assessments (TPAs) have gained popularity as both formative and summative assessment instruments in the process of teacher education and certification (Goe, Bell, & Little, 2008; Youngs, Odden, & Porter, 2003). The two national accreditation agencies, NCATE and TEAC, currently require that programs document their candidates' performance.¹ In addition, new state requirements for licensure have accelerated the adoption of TPAs (Wilson & Youngs, 2005).

Performance Assessment

A performance assessment specifies what teachers ought to know and be able to do in real teaching contexts. TPAs may have teachers teach a lesson to students, grade student papers, evaluate a textbook, and so on (Wilson & Wineburg, 1993); hence, they are regarded as authentic assessments that encompass the complexities of teaching and learning, holding novice teachers up to standards set by research and more experienced professionals (Danielson, 2007; Darling-Hammond, 2006).

Besides assessing teacher candidates' practice and contributing to teacher education program evaluation, performance assessments are considered valuable for learning to teach. Researchers suggest that performance assessments, when appropriately designed, can effectively help candidates improve their teaching (Wei & Pecheone, 2010; Wilson & Wineburg, 1993). Much previous scholarship on teacher performance assessments paid scant attention to observation protocols, as they were considered less reliable than other tests of teacher skill and knowledge (Wei & Pecheone, 2010; Youngs, Pogodzinski, & Low, 2010). However, protocols such as Danielson's Framework for Teaching and Stanford's edTPA require intensive training for scorers, which greatly improves the reliability of those instruments (AACTE, 2013; Danielson Group, 2011). Satisfactory performance on an observation protocol is one required step towards licensure in an increasing number of states.

Observation instruments. We consider evaluative observation of teacher candidates an essential component of performance assessment of pre-service teachers. The results of observation need to be approached with caution, as the practice is based on the assumption that competent performance is self-evident, that decisions can be inferred from action, that visitors understand situations, and that practitioners and observers will agree upon a limited range of appropriate responses to situations (Kennedy, 2010b). These assumptions underlie the very creation and use of observation instruments.

These cautions should be considered when looking at high- versus low-inference instruments, defined by Kennedy (2010a) as those that allow wide latitude for interpretation of both standards and behavior by an observer versus those that constrain interpretations by more carefully describing target performances (p. 231). Youngs, Pogodzinski and Low (2010) prefer low-inference rubrics as providing a common language for participants and more guidance for the teacher candidate, but overlook the extent to which both rubric

descriptions and observed behaviors can be misinterpreted. Such tendencies toward misattribution are mitigated in most observation instruments, where a conference with the candidate is mentioned or implied, lesson plans (an indication of intentions) are turned in for the observation or, as in PACT, the video is accompanied by the teacher candidate's written interpretation of the lesson. As psychometric properties of instruments have improved through more robust training on the instruments and insistence on repeated observations of individuals (Bell, 2012; McClellan, 2012), low-inference rubrics are gaining popularity in use for high-stakes assessments for both inservice teachers and for new teacher certification, making it even more critical to look at them closely.

While research on TPAs has focused on their use and effectiveness, their function as expressions of the underlying values of teacher education programs has been overlooked. Observation assessments in particular provide indications of what programs value in teaching, how they position the participants in the pedagogical interactions taking place in field supervision, and to what extent an instrument portrays the candidate as an agentic professional (one who uses his or her knowledge to make decisions and take action in ill-defined domains).

Conceptual Framework

Critical discourse analysis (CDA, Fairclough, 1992, 2001, 2003) provides a useful theoretical and methodological framework for examining observational instruments for the values associated with teacher quality, the agency afforded by their grammars, as well as the relationships implied in the texts of these instruments and underlying discourses of professionalism and accountability.

Fairclough's method of CDA was developed from linguistics and language studies (Halliday, 2004) and sociopolitical thought (Foucault, 1981; Habermas, 1984). For

Fairclough, discourses are not neutral; they do not just reflect or represent social entities and relations, but also construct them in a relationship of power within a sociopolitical context. Discourses “position people in different ways as social subjects, and it is these social effects of discourse that are focused upon in discourse analysis” (Fairclough, 1992, p.4). In this study, we view language and approaches used by selected documented observational instruments for pre-service teacher education as a discourse able to reveal and construct values about teacher performances by putting learners and educators in particular positions.

The “critical” aspect of Fairclough’s framework refers to the investigation of “connections and causes which are hidden” (Fairclough, 1992, p. 9). Chouliaraki and Fairclough (1999) state that, “all social practice is embedded in networks of power relations, and potentially subordinates the social subjects that engage in it, even those with ‘internal’ power.” (p.24). Further, while they agree with Foucault that power circulates and resides in daily actions, they think his definition needs to be supplemented with the idea of “power as domination,” (p. 24), those larger socio-economic forces imposing power externally to set the norms for individual performances and interpersonal relationships. In the case of observational instruments, domination is exercised by the institutions that create, distribute, and police the uses of these instruments, each of which refers to standards for performance defined by the program or the state.

We do not study these instruments in use, so do not have data regarding how they are interpreted, or the extent to which particular categories of performance are adopted or resisted. However, their grammar presents particular affordances and constraints, positioning participants in relation to each other in ways that assign or deny agency, and thus the power to exercise professional discretion in setting goals, planning, and accomplishing activities. The language features of the observational instruments reveal the assumption that participants

will act according to the legitimate performances stated in the rubrics, thus adopting values and practices prescribed for them.

CDA provides means for investigating the ideological underpinnings of educational discourses and accountability practices, whether they result from the imposition of business models on schools (Kliebard, 1994) or from neoliberal application of quality control and accountability measures (Baker et al., 2010; Harris & McCaffrey, 2010). The writing of content-area standards by most states, a political act intended to result in improved curriculum, has been extended to include teacher-preparation standards at the level of the state, the institution, and the leading subject-area organizations (e.g., NCTE, 2012). These standards are then reflected in standardized tests and other standards-aligned accountability measures to establish both that the individual teacher candidate and the teacher preparation program are meeting teacher preparation standards. Observation instruments are just one of these tools.

We consider the ways that the discourses of accountability in these assessment instruments position teacher candidates in relation to their students, the curriculum, and other aspects of their context. Such positioning is done not only through what is explicitly expressed as expected, but through what is presumed about the expected activities of teachers that is not stated (Fairclough, 2003). The concise wording required by observation instruments necessitates a reliance on shared, but unstated, assumptions regarding expectations.

This study critically examines several observation instruments for their underlying assumptions regarding valued teaching performances and teacher agency. In the following sections, we explain our selection of a larger and smaller sample, our analytical approach, and our findings.

Methodology

We began this two-step analysis by considering a larger sample of nine instruments for characteristics related to purpose, features, and participant roles. We then chose three from that larger group for a closer textual analysis.

Features and Types of Observational Instruments

Observational instruments come in a variety of forms, and reflect the commitments of the programs that use them. We searched for instruments referred to in the literature on preservice teacher assessment and also did an Internet search for instruments specific to various teacher preparation programs in the United States. All instruments we found were standards-based: they explicitly referred to the teacher preparation standards adopted by the state or by the individual program, even if only as a list for the observer's reference. We found with Kennedy (2010a) that instruments ranged in their ability to limit inference for the sake of reliability, from higher-inference (our notes-based and leveled checklist types, defined below) to lower-inference types (the more detailed rubrics). We sorted the instruments accordingly. We coded nine instruments in use by programs between 2010 and 2012, limiting ourselves to instruments that were referred to in the literature on assessment or where programs posted complete instruments and directions for use (Appendix A lists instruments surveyed and coding categories).

We surveyed the nine instruments to gain perspective on how they refer to teacher quality and how they position participants in the teacher education triangle: teacher candidate, student, and teacher educator. Where rubrics referred to performance assessment taking place outside the classroom (such as assessment of response to student written work or communication with parents), we did not code that portion of the rubric. Our only deviation from this principle included lesson planning: most observations are grounded in a lesson plan

turned in by the teacher candidate, and references to planning occur throughout all instruments.

Coding categories were devised using a variety of sources, including the literature on performance assessment (Danielson, 2007, Kennedy, 2010a; Youngs, Pogodzinski & Low 2010), literature on teacher knowledge (Ball, Thames & Phelps, 2008; Shulman, 1987), and our readings of a wide range of assessment instruments. These all informed our codes for purpose (formative versus summative), type of instrument, type of teacher knowledge accessed, and teacher efficiency.

We define three types of instruments. *Notes-based* instruments rely on the field instructors' notes on what has been achieved and what still needs work. They vary in the extent to which they nominate categories to be addressed. All that we found had program standards listed somewhere on the form for the observer's convenience. *Leveled checklists* list expected teaching skills and practices, but performances are checked according to the level at which the teacher candidate achieves them (e.g.: 1, 2, 3, or 4; basic, proficient, advanced). *Rubrics* are also leveled, but contain descriptions of behavior at the different levels of achievement for each standard.

The teacher, student, and teacher educator roles categories were further shaped by our interest in participant positioning and agency in the three-way interaction of classroom observation. Defining these categories in terms of roles allowed us both to record what learning and teaching performances were valued and indicate how participants were positioned. Each of the authors read through each of the instruments independently, and coded for the categories listed. The codings were compared, and the authors met to resolve discrepancies.

This preliminary analysis revealed that the nine instruments share common values arising out of common themes in the standards that anchor the instruments (see Appendix A for the emphases noted in each instrument). For example, almost all beginning teacher standards refer to creating a supportive environment, planning using the appropriate content standards, using assessments effectively to monitor student progress, and engaging students as active participants. However, within these commonalities, different programs and/or instruments reflect their particular commitments. Some put more emphasis than other programs on the teacher as effective communicator; others refer more often to teacher knowledge of students than any other category of teacher knowledge. We expected to see differences in emphasis according to the type of instrument studied, and between formative and summative assessments, but did not.

Based on this preliminary analysis, we chose three instruments for discourse analysis: 1) The Elementary Literacy Teaching Event rubric from the Performance Assessment for California Teachers and Pre-service Teachers (PACT), 2) the Michigan State University Elementary Level Field Instructor Feedback Form (FIFF), 3) Christopher Newport University Student Teacher Observation Form (CNU). These instruments follow the basic pattern noted in the nine instruments of providing a main teacher role as facilitator of learning, for students as responding to teachers as active participants, and for teacher educators as observers and evaluators. Each is a representative example of its type (rubric, leveled checklist, and notes-based form). They differ from other available forms in either being researched (PACT), tested for reliability and validity (FIFF and PACT), or explicitly research-based (CNU). They represent examples of well-regarded tools useful for examining the values and assumptions adopted by teacher education programs in the United States.

Three Examples

The PACT summative assessment instrument was developed by the Performance Assessment for California Teachers consortium (consisting of 16 higher education institutions) and refined over several years (the version coded here is from PACT's third iteration). The PACT rubrics are aligned with the California TPE (Teaching Performance Expectations) for beginning teachers pursuing credentials. The PACT instrument is designed as a structured, subject-specific portfolio of teaching—the "Teaching Event" -- including planning, instruction, assessment and reflection around a series of lessons on a topic of the candidate's choice (Chung, 2007, p.3; PACT, 2010a; Pecheone & Chung, 2006). Teacher candidates choose a unit of instruction and submit descriptions of the teaching context, lesson plans, videotapes of instruction, an assessment plan and written reflection for evaluation. Teacher educators in edTPA states will recognize the model: PACT was one of the models used by Stanford researchers in developing the edTPA (Teacher Performance Assessment), currently distributed by Pearson (Stanford GSE 2013; AACTE 2013). Observation is one part of the PACT instrument, employed only when trained scorers (recruited from teacher education faculty, supervisors, and experienced teachers) watch the videotapes of instruction. We analyzed only those elements of the rubric related to planning and instruction: EL1-EL5. Teaching Events are not scored by the student teachers' own supervisors. The Elementary Literacy rubric can be accessed online (PACT, 2010b).

In contrast to PACT's significant emphasis on documentation, the Michigan State University (MSU) elementary Field Instructor Feedback Form collects most of its data for assessment through observation. The MSU elementary FIFF (MSU College of Education, 2010) was developed based on the *Entry-level standards for Michigan Teachers*, tested for reliability during 2006-2007, and put into use in 2007. MSU requires field instructors to observe their teaching candidate supervisees in the classroom every other week during the

internship year. The field instructors are experienced teachers or graduate assistants employed by the MSU teacher education program. The information collected by this instrument served as data on program quality for accreditation by TEAC as well as fostering the process of learning to teach. The observational instrument has two parts: a standards-based checklist to grade the teaching candidate's instruction, and a blank sheet for the field instructor to write down observation notes. The observation is complemented by the teacher candidate's lesson plans, and pre- and post-observation discussions between field instructor and teacher candidate.

Christopher Newport's University's Student Teacher Observation Form (CNU) was adapted from material from the New Teacher Center (University of California at Santa Cruz and the University of Virginia Curry School of Education), for use in their MAT program through Spring of 2012.² It was used as a formative assessment by university supervisors during their seven required observations of teacher candidates. It was distinct from the forms used for midterm and final evaluations; those were leveled checklists based on Danielson's (2007) Framework for Teaching. CNU's one-page observation form contained spaces to collect notes based on "What's Working," as well as current concerns and next steps; there was also a checklist of teaching standards observed.

The three instruments refer to a number of common domains: Planning, interaction with students, pedagogical skills, and assessment of student learning in process. These domains receive different emphases in the different instruments: CNU emphasizes classroom environment more than the others, and PACT has the most detailed description of assessment. Categories become more differentiated with the more detailed descriptions of behavior provided in FIFF and PACT, where performance is evaluated at four levels: for

PACT, these are 1 (does not meet standard) to 4 (exceeds expectations), while FIFF uses Beginning, Expanding, Connecting and Refining.

Formative and summative assessment in the three systems. The three instruments each serve differing purposes within their respective assessment systems. Both the California colleges and universities forming the PACT system and Michigan State University align their assessments with their respective state teaching standards, and both do so to satisfy requirements of larger policy bodies. California's Commission on Teacher Credentialing currently requires that any candidate must pass a valid, reliable Teacher Performance Assessment (TPA) prior to recommendation for a credential. MSU developed its performance assessment as part of its process of TEAC accreditation. However, the relationship between formative and summative assessment differs significantly among the three systems.

In the PACT system, formative assessment is the responsibility of each member program, in the form of ESAs, or Embedded Signature Assessments (PACT, 2010a). PACT acknowledges that institutions will differ in the extent to which their ESAs and field instruments align with the expectations of the PACT TPA. In contrast, the formative and summative assessments at MSU are integrated into a coherent assessment program engaged in by the teacher candidate, the field instructor, seminar instructors, and the mentor teacher. FIFF takes place at regular intervals, while further assessment of the candidate's progress towards the standards is provided by midterm and final conferences among the field instructor, teacher candidate, and mentor teacher, mediated by a rubric based on the same standards as FIFF (MSU College of Education, 2010). In contrast, the CNU instruments, borrowing from two different published programs, that of the New Teacher Center for the formative instruments and Danielson's Framework for Teaching for the summative

(Christopher Newport University, 2010), form a less coherent system, although they are used in ways similar to the MSU assessments.

Although the PACT is summative and FIFF and CNU are formative, these are the instruments used by each system to evaluate teaching performance in the classroom, the point at which the activity is judged against performance descriptors in an instrument, and thus one point at which valued teaching performances are defined. We use Fairclough's framework of critical discourse analysis to examine the embedded meaning in these instruments.

Analysis

During the process of analysis, three dimensions of texts denoting agency and interpersonal choices were closely examined in addition to the analysis of the instruments' form, purpose, and positioning: *formality*, *transitivity*, and *modality* (Fairclough, 1992, 2001, 2003). We coded each of the instruments for these three at the level of the clause. Within any one clause there may be overlapping nodes of formality, transitivity, and modality. Coding was completed independently, and then the researchers met to negotiate any differences across codings.

In CDA, *formality* is one aspect of analysis related to access to discourse. Greater formality is a property implying "high social prestige and restricted access" (Fairclough, 2001, p.54). We evaluate the relative formality of the instruments by examining the complexity of the grammar used, the forms of address, and the use of formal, professional vocabulary versus conversational vocabulary.

Transitivity is analyzed to identify the ideational meaning of language, the way it "represents reality" (Fairclough, 1992, p. 27) of subjects, objects, social positions, how subjects and objects are positioned, and instances of relations of power in the use of language. We examine pronoun use to indicate positioning, then do an analysis of ergative

transitivity to investigate the assignment of agency to various human and nonhuman actors represented in the observational instruments. Ergative transitivity grammatically represents responsibility for action through types of agents present or not present in a clause (Halliday, 2004).

Modality is also an important resource for construing textual meanings, referring to the multiple ways available to a speaker for expressing commitment to a proposition (Fairclough, 2001, p.107). An analysis of modality should extend well beyond the modal verbs to include all wordings speakers/writers use to modulate their attachment to, or detachment from, a proposition (Halliday, 2004). In our analysis, we examine relational and expressive modality embedded in the texts of the selected instruments.

Limitations

We have chosen to examine only one among several instruments used to evaluate teacher candidates in each program, and we look at only three programs in the United States. In addition, their research literatures differ. As stated earlier, PACT is a highly developed instrument studied by several researchers (Chung, 2007; Pecheone & Chung, 2006), while FIFF's effectiveness and effects were studied locally. CNU lies somewhere in between, as it is based on the New Teacher Center's research on induction (New Teacher Center, n.d.). Furthermore, PACT was established as a summative evaluation tool (Chung, 2007) while FIFF and CNU provide teacher candidates with formative evaluation. However, each instrument was designed to serve the dual purposes of evaluation and helping teacher candidates to learn to teach (Chung, 2007; New Teacher Center, n.d.), enabling us to compare how the texts of these instruments are scripted for both purposes.

We must stress that these instruments are not the assessments, but are texts used in social and discursive practices of assessment. Observers use tools according to their own

preconceptions of quality teaching performance (Courtney Bell, personal communication), and even the most low-inference of these instruments, PACT, with its requirement that scorers take PACT training and be calibrated at intervals, is not immune to variations in application. We did not study these tools in use, so as not to limit our choices to geographical convenience in this unfunded study. Therefore, our findings are limited to a discussion of the affordances of these instruments implied by their grammar, and the underlying assumptions about teacher quality and learning to teach embedded in the language.

Findings

These three observational instruments reflect the larger values of teacher knowledge, effectiveness and positioning of teacher candidates vis-à-vis their own teachers and their pupils seen in our analysis of nine observation instruments. All three emphasize knowledge of pedagogy and knowledge of students over knowledge of content (although FIFF emphasizes pedagogical knowledge less, and PACT knowledge of students less than the others). All three value planning skills as evidence of teacher effectiveness. Their views of the teacher's and students' roles in the classroom differ somewhat: CNU puts more emphasis on the teacher's ability to manage and the students' compliance; FIFF focuses most strongly on the teacher as facilitator, and the students as actively responding; and PACT, although it also emphasized the teacher as facilitator, puts more emphasis on the role of the teacher as evaluator, and on students as high achievers (Appendix A shows the content analysis of the text of the instruments).

In the rest of this paper, we describe the results of our analysis of the three focal instruments: how the instruments reflect the values of their programs and larger educational discourses; how their textual features differentially afford agency to teacher educators, teacher candidates and their students; and how they position these participants in relation to

each other in pedagogical relationships. We then discuss the implications of these findings for teacher education programs.

Discourses and Values Reflected in Observational Instruments

The Performance Assessment for California Teachers (PACT), the Michigan State University Elementary Level Field Instructor Feedback Form (FIFF) and the Christopher Newport University Student Teacher Observation Form (CNU) reflect larger educational discourses as well as specific values in the local teacher education program.

We examine the ideological underpinnings of these instruments through two of Kliebard's (1994) four types of competing curricular discourses: *social efficiency* and *developmentalism*. Social efficiency, growing out of business and economic discourses, has been dominant in forming education practice and policy since at least the late 19th century. In more recent decades, business discourse has been appropriated anew by neoliberal promoters of competition and accountability. (Kliebard, 1994; Baker et al., 2010). It demands standards to enforce a kind of teaching that represents economically useful knowledge, and requires accountability to those standards. Developmentalism entails attending to individual students' needs and learning preferences. Developmentalist discourses have existed on a parallel track with social efficiency discourses since the establishment of child and adolescent psychology in the 1860s (Kliebard, 1994). We find traces of both social efficiency and developmental discourses in these instruments.

These instruments of social efficiency and accountability incorporate developmentalist discourses where individual pupils' opportunity to demonstrate understanding of content is referred to in observational instruments. For example, note the implied sequence of activities in standards-based teaching at PACT target level 3 (bolding in original):

PACT EL5: (Monitoring student learning during instruction: How does the candidate monitor student learning during instruction and respond to student questions, comments, and needs?) Level 3:

The candidate monitors student understanding of literacy by eliciting student responses that require thinking.

Candidate responses **build on student input to guide improvement** of students' use of literacy skills or strategies.

PACT EL 5 describes how the candidate is expected to respond appropriately to their pupils' needs as assessed in the course of classroom activity. Similarly, this attention to student development can be seen in FIFF item 6:

6. You included diverse students in class activity (Stand. 2):
a. provided developmentally appropriate instruction for the grade.
b. enabled many students to participate, contribute, be appreciated.
c. took steps to accommodate/include particular students.

While the standard indicates attention to “diverse students” (a *social meliorist* discourse, Kliebard, 1994), this standard relates more closely to a *developmentalist* model in demanding teacher candidates tailor teaching for pupils based on “developmentally appropriate instruction for the grade.” The references to “grade-level” developmental stages, and to engaging and persisting with “the” content (in item 2), elide explicit reference to cultural variables in learning, such as race, class and gender.

It is not our contention that assessment instruments, which serve instrumental ends, be ideologically pure, but that teacher educators should be aware when their criteria for successful teaching and learning are conceptually incoherent. We found a troubling incoherence in the different ways in which educators and learners at different levels were positioned. In the following sections, we analyze the textual features of these instruments to illustrate how teacher candidates are positioned as having agency (or not), and how educators

at different levels are positioned in their own teaching/learning relationship with their students.

Agency and Power Relationships Afforded by Observational Instruments

The power and agency teacher candidates wield in the classroom in relation to their students and teacher educators is relevant to the concept of professionalism promoted by a program. Textual features (such as transitivity and formality) that indicate a passive position for the novice teacher would not likely contribute to agentive professionalism. We refer both to who determines what happens in the classroom, and who has power/agency in the evaluation relationship.

Implied agency in classroom activity. Transitivity is a distinct textual feature indicating the relation between participants and action. We focus on the use of ergative transitivity, as representing responsibility for how things happen, and pronoun use as indicating agency and relationships.

There are two types of clauses in ergative transitivity (Halliday, 2004): “middle” clauses, that involve a medium and a process, but no agent (we call these “agentless” clauses. Activity is referred to, but not the agent that sets it in motion), and “effective” clauses, that involve a medium, a process, and an agent. We make a further distinction in the effective clauses: those with human agents, and those with non-human agents.

In the FIFF, transitivity differs between the rubric and the performance score descriptors. The descriptors, an attachment to the rubric used by the field instructor to decide on the appropriate level of performance, contain slightly more non-human agents (18) than human agents (15 -- usually the teacher candidate). However, in the rubric that mediates the interaction between the field instructor and the teacher candidate, there are no agentless

clauses or non-human agents. The teacher candidate is most often stated or implied as the agent, with 29 out of 43 clauses featuring the candidate as agent (see Table 1).

(Table 1 about here)

Pronoun use also aids in FIFF's construction of the teacher candidate's agency. The FIFF identifies the teacher candidate as "you" in its action processes, and all of its standards items have "you" as the subject. By using the second person pronoun, the FIFF positions the teacher candidates as the agent of actions, such as "led", "prepared," "organized", etc. (see FIFF Item 6, above). The pupils, on the other hand, generally act under the leadership of the teacher candidate. The instance of "you" as the subject underlines the "you-centeredness" of the teacher candidate, in contrast to the objectivity of PACT.

The complex grammar of the PACT asserts its objectivity by presenting a situation in which performance is or is not evident. It is not until standard four of the entire PACT instrument for elementary literacy that the teacher candidate is grammatically credited with agency for making things happen in the lesson and more generally (the candidate "engages students," "monitors student learning," etc.). In the standards questions relating to planning and observation, the teacher candidate is an agent three out of five times. In the performance level descriptors, non-human agents dominate. While on a common-sense level it may be understood that the teacher candidate is the human agent responsible for the plans, performances, and opportunities described, this understanding is not reflected in the grammar. Reading the rubric, one is confronted by a series of nominalizations representing artifacts that are assessed by an unacknowledged scorer. Social efficiency's sense of objectivity both aids and is a product of these constructions.

For instance, the agency of both teacher candidates and students is elided through a lack of agents ("Opportunities are provided. . ."), non-human agents ("Learning tasks focus

on. . .” “Plans refer to. . .”), and nominalizations. These last include using nominalizations (underlined) as agents: “The candidate elicits explanations of student thinking, and uses these explanations to further the understanding of all students” (EL5, level 4). Students explain, they think, they understand, but their agency is hidden behind nominalizations. We were surprised by just how few times human agency is credited by PACT grammar. Of 66 clauses, 19 contained human agents (see Table 1).

Agency ascribed to the three participants in an observation is handled differently in CNU. There are no agents in the top half of the form; only in the direction to “include dates for completion” is an agent indicated: the teacher educator. The only category including a verbal (“What's working,”) elides agency through “what” used as a pronoun. In the bottom half of the form, only the teacher candidate is implied as an agent, once directions are given to the observer (“. . . you observed/discussed today.”). Each of the list of 24 performance components begins with a present-tense verb in the third person referring to the candidate (“Engages,” “Utilizes,” “Maintains,” “Assesses”). Students have no agency: they are either the objects of the candidate’s actions (three times) or in the grammatical position of modifier (“student learning,” “student needs”). The students' agency is elided through nominalizations, much as it was in PACT.

These assessment forms not only differ in their presentation of the teacher candidate as an agent in the classroom, but as a participant in assessing the teaching being observed.

Assessment as an indicator of relationship. The study of pronoun use in FIFF revealed both the roles played by teacher candidate and student in the classroom, and the relationship between the field instructor and the candidate. Observation protocols range in their use from bases for negotiation between educators to tools for communication of top-

down assessments by strangers. These roles are indicated by the features of formality and transitivity (pronoun use).

Formality. In examining the instruments, we found that all three documents are formal in comparison to vernacular speech: all incorporate professional vocabulary, needing to fit descriptions of complex performances into relatively little space. This necessitates a reliance on the use of compound predicates, lists of performances, and embedded clauses.

The following texts are example excerpts from the rubrics about “planning” from PACT and FIFF (emphasis in original):

PACT-EL1 (How do the plans structure student learning of skills and strategies to comprehend and/or compose text?) -Level 3

- Learning tasks *or* **the set of assessment tasks** focus on multiple dimensions of literacy learning through **clear connections** among facts/conventions/skills, and strategies for comprehending and/or composing text. A **progression** of learning tasks and assessments is planned to build understanding of the central literacy focus of the learning segment.

FIFF-Item 1:

1. You came prepared for standards-based instruction (Stand. 1):
a. prepared to conduct caring and organized class activity.
b. prepared to promote learning to recall and comprehend.
c. prepared to promote learning to apply, analyze, synthesize, or evaluate.

The difference in formality between PACT and the other two instruments stands out immediately. FIFF and CNU employ more speech-like forms, favoring short sentences and simple vocabulary, combined with formal vocabulary. For instance, the use of direct address of candidates in FIFF (“You came prepared for . . .”) in the example above comes across as informal, but the descriptions of performances generally use more formal diction and technical vocabulary (“. . . promote learning to apply, analyze, synthesize, or evaluate.”). CNU is slightly different. The top half of the instrument, prompting the observer to take notes, is less formal than the bottom checklist of observed practices. The labels on the note-

taking sections make use of colloquial metaphors to provide a sense of informality (“What’s Working;” “Student Teacher’s Next Steps”). In contrast, the metaphors used in the bottom list of components of performance (all set in the simple present tense) can be categorized as professional figures of speech (“*Designs* coherent instruction,” “Accepts *feedback* in a professional manner”).

In contrast, PACT uses more formal diction and more complex sentence types in general. By using more complicated grammar structures and selecting vocabulary usually reserved for more formal occasions and that codifies the knowledge of the field, PACT demands specific knowledge and skills in operating the evaluative procedures. In PACT, vernacular terms are never used. There is no direct form of address: everything is in the third-person indicative. Finally, the behavioral descriptions are longer and more complex. For instance, in the example above, the writers used embedded clauses (“... student learning of skills and strategies to comprehend. . .”) and a greater use of participial and infinitive phrases (“to build understanding. . .”) to incorporate more different types of action into a sentence. These descriptions display a repeated use of nominalization that, according to Fairclough, can be used to distance a person from action and “leave responsibility unclear” (Fairclough, 2001, p.103).

The differences in formality reflect the intended use and the assumed relationships between teacher candidates and evaluators. PACT must be scored by someone other than the teacher candidate’s supervisor, and videos are often scored by complete strangers. Scorers working with PACT have been trained to operate the evaluation formally and are to interpret and judge the videotaped teaching without considering a personal or pedagogical relationship with the teacher candidate. In contrast, field instructors and supervisors in MSU and CNU are to use the instrument *with* the teacher candidate immediately after observation. The more

colloquial style of FIFF well serves the collaboration between the field instructor and the teacher candidate.

Pronoun use also adds to formality and distance. Although CNU shares with FIFF its use in the field to “provide evaluative feedback” (CNU Student Teacher Handbook, 2012, p. 34), only the observer is directly addressed (“Please check the specific component under each domain that *you* observed/discussed today,” emphasis added). The teacher candidate is only referred to in the third person, as are students. While such reference does not determine a more formal and hierarchical relationship in the field, it appears to afford the supervisor the role of the one who delivers the evaluation, both to the candidate and to the unnamed final reader of the evaluation.

In PACT, the teacher candidate is referred to as “the candidate” throughout the form, indicating that the identity of the observed is unknown to the observer (referred to as the “scorer” in PACT literature: PACT, 2010a). For instance:

Strategies for intellectual engagement seen in the clip(s) offer **structured opportunities** for students to **actively develop and/or apply** specific literacy skills and strategies to comprehend and/or compose text.” (PACT, EL4, Level 3).

Except for the word “seen” in this example, the observer is never acknowledged in the grammar throughout the lengthy rubric.

Evaluation and the Pedagogical Relationship

The relationship between the evaluator and the teacher candidate is not only one of power, but it is a teaching relationship, analogous to that the teacher candidate has with his or her pupils. The instruments vary in their positioning of teacher educators, teacher candidates, and K-12 pupils within pedagogical relationships that function according to either developmental or social efficiency principles.

In the process of evaluating the teacher candidate according to the PACT standard above, the anonymous scorer is the one to interpret what s/he saw in the video clip and make judgments. What counts as evidence of effective teaching is the part that is “seen” in the clips. This wording gives the sense of candidates being subjected to anonymous and invariant evaluative procedures, without reference to the development of teacher/learners. The teacher candidate either performs proficiently or not. Although passive voice is also applied to the relationship between the teacher candidate and the pupil (“structured opportunity”), the pupil is present as an agent to “actively” comprehend/compose texts, activity that is facilitated by the successful candidate. The CNU checklist portion also indicates teacher candidate performances as evident (or not), to be checked off (or not), and the presence of evidence is determined by the observer.

In contrast, the focus on “you” in the MSU form indicates allowances for both perspectives in the more developmental conversation between the teacher candidates and the field instructors. However, the privileged perspective of the observer in CNU and PACT does not necessarily indicate greater power or agency for the teacher educator, a point we will discuss below.

Modality and possibilities for development. In these rubrics, relational modality assists in positioning participants in relation to the rubric and the pedagogical act of evaluation. In defining the speaker’s commitment to the proposition (e.g., in FIFF, “You talked or wrote about it” as stronger modality, vs. “You appeared organized, . . .” as weaker modality), relational modality helps to define the authority of one speaker in relation to others (Fairclough, 2001). Consider the following excerpts about teacher candidates’ ability to assess pupils’ learning and adjust instruction from each instrument:

PACT-EL5: How does the candidate monitor student learning during instruction and respond to student questions, comments, and needs?

Level 3: The candidate monitors student understanding of literacy by eliciting student responses that require thinking.

FIFF-Item 3: You took steps to learn whether students engaged the lesson, and adjusted.

CNU: Demonstrates flexibility and responsiveness to student learning needs.

As is shown in the excerpt from PACT, the verb (does) is in the simple present tense form. This is one terminal point of relational modality, a categorical commitment of the evaluator to the truth of the proposition regarding the performance of the teacher candidate. Throughout the PACT instrument, the verbs are all in non-modal present tense forms, supporting a view of the teaching as transparent—as if every teaching move signals its own meaning to any observer, without the need for interpretation. The checklist for CNU, from which this excerpt is taken, shows a similar categorical modality.

For the excerpt above from FIFF, “took steps to” indicates the idea of an attempt, thus acknowledging approximations to the goal of engaging students, a developmental stance. With the intermediate modalities suggested by “could” (FIFF Item 8) and phrases like “took steps to,” the complex meanings embedded in the established evaluation standards and the assumptions that field instructors bring to the process of evaluation are subject to interpretations from both field instructors and candidates.

The note-taking half of the CNU form shows variations in modality, from the strong expressive and relational modality of “What's Working” to the less strong “Current focus-- challenges – concerns.”

Modality in PACT helps define successful teaching performance as attending to the developmental needs of pupils, but fails to account for the developmental needs of teachers-as-learners. In FIFF, openings for teacher development are provided through the somewhat less categorical modality. Predictably, there are not many indications in either PACT or FIFF

of lack of commitment to an assertion. That is the nature of the genre, the logic of rubrics, to present a list of options for a given performance, but to commit to one so the performance can be evaluated.

Positioning the teacher educator. In general, formality, strong relational modality, and pronoun use establish the observer as authorized to judge the performance with certainty and place the teacher candidate on the continuum between an inadequate beginning and mastery of teaching behaviors and skills. However, they also imply at least one other agent outside the teacher candidate/teacher educator/pupil triangle. Who wrote the rubric, and who trained the scorer to exercise fidelity to the instrument? Who addresses the CNU supervisor as “you?” Who defined levels 1, 2, 3, and 4 on the PACT rubric? These instruments implicitly mark a process of holding teacher educators accountable for their students’ capabilities, masking the agency of the state and regulatory bodies.

Discussion

We chose to use critical discourse analysis in this comparison of observational instruments in order to examine how underlying discourses of social efficiency and developmentalism position the various participants in the preservice teaching context as agentive professionals and as learners through the grammar of these instruments. The language and the grammar of PACT, CNU and FIFF reveal the intended pedagogical interactions in their intended use: summative and formative assessment of teacher candidates.

Positioning educators as agents and professionals

If professionals are those whose specific expertise, derived from training and experience, enables them to use their knowledge to make decisions and take action in ill-defined domains, one would expect a performance assessment to allow for the messiness of the interaction of content, context, and participants, and leave room for judgment on the part

of the novice teacher and the teacher educator. Each instrument provides different openings for professional judgment on the part of participants.

At first glance, the anonymous evaluator in PACT would seem to have the power in that assessment process; however, the expectation implied by requiring standardized scoring training is that any observer should give the same score to a video. PACT does not provide openings to express individual professional judgments derived from differing experiences. In contrast, the transitivity features in FIFF and CNU show traces of state goals in the lists of the teaching standards, but ask the teacher educator/observer for her or his own judgment regarding what is going well and what still needs work. Teacher educators/observers have more opportunity to exercise discretion in decision-making. While all candidates are expected to meet state standards, what that looks like might not be identical between performances or performers, leaving possibility for negotiation as well as for growth.

Positioning participants as learners

The developmentalist stance toward learning implied by these observational instruments is largely reserved for the K-12 pupils, not for the new teacher as learner. While the teacher candidate prepares a lesson plan that attends to the developmental needs of the particular elementary students in her class, the field instructor uses the same institutional protocol for all candidates in all contexts. The rubrics for beginning teachers are developed and used by teacher educators in accordance with state standards as a process of norming towards standard approaches to delivering curriculum, and as a gatekeeping device: those who do not reach a minimal standard will not proceed to a credential. It is the case that the repeated use of FIFF and CNU as formative assessments allows for change in the descriptions and ratings of the teaching practices over time. However, the teacher educator's active role in instruction and mentoring is mostly implicit. Our content analysis of the three instruments

(Appendix A) indicates that the role of the teacher educator is less elaborated than that of the teacher candidate as educator.

Inconsistencies between the theories of learning ascribed to children's learning and those ascribed to teacher learning in these instruments raise questions about whether or not teacher educators see both stages of learning through the same theoretical lens. These inconsistencies may be justified, if adults do undergo different learning processes and can be held to different standards. However, if learning to be a teacher is also a process marked by individual and cultural differences based on variations in experience, background, and learning styles, and taking place by constructivist processes of reflection and building on previous experiences, then evaluation instruments that take a standardized view of learning and that place the evaluator in the position of marking where or whether the learner measures up to standard contravene that model of learning. In an age where learning theory and assessment theory are often at odds, teacher educators in particular need to be vigilant in demanding assessments that provide them, as well as their students and the state, with useful information that can assist them in supporting the learning of the teacher candidates in their care, and that respect the professionalism and agency of teacher candidates and teacher educators.

Implications

We acknowledge that the textual features we studied do not determine the interpretation and the use of these instruments, but they do constrain interpretations and usages, making certain actions more likely and others less likely. More research is needed on how such instruments mediate the relationship between teacher educators and teacher candidates. Teacher credential programs and state policy-makers should carefully examine

the goals and expected outcomes of teacher education and choose formative and summative assessments that align with their goals and values.

However, teacher educators do not always get to decide what protocols are used to evaluate their students. For example, EdTPA or an equivalent assessment instrument is currently required by eight states as part of their certification requirements, and programs in 24 other states are exploring edTPA (Lu, 2013). We have shown that observation protocols can reflect the values and the commitments of a teacher preparation program, and thus must be scrutinized for fit. However, where a state requires all new teachers pass the same instrument, it is even more important to take note of how it positions teachers, students, and teacher educators, and how it defines teacher professionalism. We urge teacher educators to look critically at the instruments used to define teaching and professionalism for their programs, and to advocate maintaining their place in defining teaching and learning within their own programs, and with their novice professionals.

We teach in an age when the state is increasing its mediation of the pedagogical relationships between students and teachers, and between teacher candidates and teacher educators. An awareness of the affordances of the observation protocols used allows for adaptation of their use. Where a program cannot develop its own instruments, or where there is no room for modification, joining teacher candidates in a critical reading of the state-approved rubric allows for discussion in how to prepare for and frame the performance that is assessed.

References

- AACTE. (2013). edTPA. Washington, D.C.: American Association of Colleges for Teacher Education. Retrieved from <http://edtpa.aacte.org/>
- Baker, E.L., et al. (2010) *Problems with the use of student test scores to evaluate teachers*. Economic Policy Institute Briefing Paper #278. Washington, D.C.: Economic Policy Institute. Retrieved from <http://www.epi.org/page/-/pdf/bp278.pdf>
- Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59 (5), 389-407, doi:10.1177/0022487108324554.
- Bell, C. (2012, February). Toward an understanding of teacher quality. Paper given at Michigan State University, East Lansing, Michigan.
- Campbell, D. M., Melenyzer, B. J., et. al. (2000). *Portfolio and performance assessment in teacher education*. Needham Heights, MA: Allyn & Bacon.
- Chouliaraki, L., & Fairclough, N. (1999). *Discourse in late modernity: Rethinking critical discourse analysis*, Edinburgh: Edinburgh University Press.
- Christopher Newport University. (2010a). Field Experiences. Retrieved from <http://www.cnu.edu/mat/fieldexp.asp>.
- Christopher Newport University. (2012). *Teacher internship handbook*. Retrieved from <http://cnu.edu/teacherprep/studentteaching/cnu%20stud%20teach%20handbook%2013.pdf>
- Chung, R.R. (2007, April). Beyond the ZPD: When do beginning teachers learn from a high-stakes portfolio assessment? Paper presented at the Annual Meeting of the American Educational Research Association. Chicago, Illinois.

- Cohen, D. K. (2010). Teacher quality: An American educational dilemma. In M. M. Kennedy (Ed.), *Teacher assessment and the quest for teacher quality: A handbook* (pp. 375-401). San Francisco: Jossey-Bass.
- Danielson, C. (2007). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Danielson Group. (2011). Promoting teacher effectiveness and professional learning. Retrieved from <http://www.danielsongroup.org/Default.aspx>.
- Darling-Hammond, L. (2006). Assessing teacher education: The usefulness of multiple measures for assessing program outcomes. *Journal of Teacher Education*, 57(2), 120-138, doi:10.1177/0022487105283796.
- Fairclough, N. (1992). *Discourse and social change*. Cambridge, England: Polity.
- Fairclough, N. (2001). *Language and Power* (2nd ed.) London: Longman.
- Fairclough, N. (2003). *Analysing discourse: Text analysis for social research*. New York: Routledge.
- Foucault, M. (1981). *The history of sexuality: An introduction. Volume 1*. (Trans. R. Hurley). New York: Vintage.
- Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved from <http://www.gtlcenter.org/sites/default/files/docs/EvaluatingTeachEffectiveness.pdf>
- Habermas, J. (1984). *Theory of communicative action. Vol. 1: Reason and the rationalization of society*. (Trans. T. McCarthy). Boston: Beacon Press.
- Halliday, M. (2004). *An introduction to functional grammar* (3rd. ed.). London: Arnold.

- Harris, D. and McCaffrey, D. (2010). Value-added: Assessing teachers' contributions to student achievement. In M.M. Kennedy (Ed.) *Teacher assessment and the quest for teacher quality: A handbook* (pp. 251-282). San Francisco, CA: Jossey Bass.
- Kennedy, M.M. (2010a). Approaches to annual performance assessment. In M. M. Kennedy (Ed.), *Teacher assessment and the quest for teacher quality: A handbook* (pp. 225-250). San Francisco: Jossey-Bass.
- Kennedy, M.M. (2010b). Attribution error and the quest for teacher quality. *Educational Researcher*, 39 (8), 591-598, doi:10.3102/0013189X10390804.
- Kliebard, H. (1994). *The struggle for the American curriculum, 1893-1958* (2nd ed.). New York: Routledge.
- Lu, A. (2013, October 7). Making sure teachers are classroom-ready. *Stateline*. Washington, D.C: Pew Charitable Trusts. Retrieved from <http://www.pewstates.org/projects/stateline/headlines/making-sure-teachers-are-classroom-ready-85899509951>
- McClellan, C. A. (2012, April). Video-scoring working group. Paper presented as part of the symposium *Measures of Effective Teaching Project*. American Educational Research Association Annual Meeting, Vancouver, British Columbia.
- MSU College of Education. (2010). Field Instructor Feedback Form. Retrieved 1/3/2011 from Elementary Education Teacher Preparation: Forms and Documents. <http://www.education.msu.edu/te/Elementary/Forms-Documents.asp#field-instructors>
- NCTE (2012). NCTE/NCATE standards for the initial preparation of teachers of secondary English language arts, grades 7-12. Retrieved from

http://www.ncte.org/library/NCTEFiles/Groups/CEE/NCATE/ApprovedStandards_111212.pdf

New Teacher Center. (n.d.). New Teacher Center Home Page. Retrieved from <http://www.newteachercenter.org/>

Pecheone, R., & Chung, R. (2006). Evidence in teacher education: The performance assessment for California teachers (PACT). *Journal of Teacher Education*, 57(1), 22-36, doi:10.1177/0022487105284045.

Performance Assessment for California Teachers (PACT). (2010a). A brief overview of the PACT assessment system. Retrieved from: <http://www.pacttpa.org/>.

Performance Assessment for California Teachers (PACT). (2010b). Elementary Literacy Performance Rubric. Retrieved from: http://www.pacttpa.org/_main/hub.php?pageName=Rubrics#Rubrics.

Porter, A., Youngs, P., & Odden, A. (2001). Advances in teacher assessments and uses. In V. Richardson (Ed.). *Handbook of research on teaching* (4th ed., pp.259-297). Washington, DC: American Educational Research Association.

Rogers, R. (2011). Critical approaches to discourse analysis in educational research. In R. Rogers (Ed.). *An introduction to critical discourse analysis in education* (2nd ed., pp. 1-20). New York: Routledge.

Sanders, W., & Horn, S. (1998). Research findings from the Tennessee value-added assessment system (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education*, 12(3), 247-256, doi: 10.1023/A:1008067210518.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review* 57, 1-22.

Stanford Graduate School of Education. (2013). edTPA. Retrieved

from <https://scale.stanford.edu/teaching/edtpa>.

Wei, R.C., & Pecheone, R.L. (2010). Assessment for learning in preservice teacher education: Performance-based assessments. In M.M. Kennedy (Ed.), *Teacher assessment and the quest for teacher quality* (pp. 69-132). San Francisco: Jossey-Bass.

Wilson, S., & Wineburg, S. (1993). Wrinkles in time and place: Using performance assessments to understand the knowledge of history teachers. *American Educational Research Journal*, 30(4), 729-769, doi:10.3102/00028312030004729.

Wilson, S., & Youngs, P. (2005). Research on accountability processes in teacher education. In M. Cochran-Smith. & K. Zeichner (Eds.). *Studying teacher education: The report of the AERA panel on research and teacher education* (pp. 591-644). Washington, D.C.: American Education Research Association.

Youngs, P., Odden, A., & Porter, A. (2003). State policy related to teacher licensure. *Educational Policy*, 17(2), 217-236, doi:10.1177/0895904803017002002.

Youngs, P., Pogodzinski, B., & Low, M. (2010). The role of formative assessments in new teacher induction. In M.M. Kennedy (Ed.), *Teacher assessment and the quest for teacher quality: A handbook* (pp. 165-200). San Francisco: Jossey-Bass.

Footnotes

¹ NCATE and TEAC recently merged to form the Council for Accreditation of Educator Preparation (CAEP), but as of this writing, institutions are still abiding by NCATE or TEAC rules.

² Currently, CNU is using a modified form of this instrument for student teacher observation.

See <http://cnu.edu/mat/studentteaching/Teacher%20Intern%20Observation%20Form.docx>,

p.42, for their current form.

Appendix A: Dominant emphases in teacher candidate observation forms

1. CNU	2. San Diego	3. Northern Arizona U.
4. MSU FIFF	5. Purdue	6. Rutgers Camden
7. St. Scholastica	8. PACT	9. Danielson's Framework for Teaching

Instrument title	1	2	3	4	5	6	7	8	9
Purpose: formative									
Purpose: summative									
Form: notes-based									
Form: leveled checklist									
Form: low-inference Rubric									
Teacher Knowledge(TK)									
TK: content									
TK: pedagogic									
TK: students									
TK: context									
Teacher Efficiency (TE)									
TE: planning skills									
TE: time management.									
TE: differentiates instruction									
Teacher Roles: TR									
TR: manage classroom									
TR: transmit knowledge									
TR: facilitate interaction									
TR: facilitate learning									
TR: learner									
TR: communicative efficiency									
TR: evaluate									
TR: nurture									
Student Roles: SR									
SR: intrinsic motivation									
SR: respond to teacher									
SR: participate actively									
SR: compliant receivers									
SR: high achievers									
SR: work consistently									
Teacher Educator Roles: TeR									
TeR: observe									
TeR: evaluate									
TeR: mentor or model									
TeR: facilitate TC learning									

Dark gray: most strongly emphasized in text of instrument
 Medium gray: less strongly emphasized
 Light gray: present in instrument
 White: barely mentioned or absent

Observation Instruments Coded:

1. Christopher Newport University Student Teacher Observation Form (CNU)
2. San Diego State University Student Teaching Observation Form B
3. Northern Arizona University Student Teacher Observation Form*
4. Michigan State University Field Instructor Feedback Form
5. Purdue University Student Teacher Observation Form and Guide to Student Observation Form
6. Rutgers University Campus at Camden Student Teaching Classroom Observation Form
7. The College of St. Scholastica Lesson Observation Rubric for Preservice Teachers
8. Performance Assessment for California Teachers Elementary Literacy Rubric
9. Danielson's Framework for Teaching, 2nd ed.

*Since we collected and scored Northern Arizona's notes-based observation form, they have switched to an instrument more similar to the leveled checklists in this sampling.

	No agent	Non-human	Human agents	TC	Students	Observer	Hyp. teacher
PACT	9	38	19	6	13	0	0
FIFF	0	18	58	35	19	3	2
CNU	2	0	27	25	0	2	0

Table 1: Agency in PACT and FIFF. PACT: data taken from descriptors, EL1-EL5. 66 total clauses. FIFF: Data taken from checklist and directions for implementation. 76 total clauses.

CNU: Data taken from directions for note-taking and accompanying checklist of 24 standards. 29 total clauses.