

Teacher Education for Social Justice: What's Pupil Learning Got To Do With It?

Marilyn Cochran-Smith,¹ Ann Marie Gleeson, and Kara Mitchell

Boston College

Abstract

There are many controversies related to the increasingly widespread theme of “social justice” in teacher education, including debates about whether and/or how promoting pupils’ learning is part of this theme. This article briefly discusses the concept of teacher education for social justice in terms of pupils’ learning and then considers this notion in terms of the current press to hold teacher education accountable for learning. The article then presents the results of the “Teacher Assessment/Pupil Learning” (TAPL) study, an analysis nested inside a larger qualitative study about learning to teach over time in a preparation program with a stated social justice agenda. The purpose of the TAPL analysis was to evaluate the outcomes of teacher education for social justice by assessing the intellectual quality of assessments created or used by teacher candidates during the student teaching period and also to assess the quality of their pupils’ responses to those assessments. The project used Newmann and Associates’ (1996) framework of “authentic intellectual work” and the scoring system that emerged from that framework because of their general consistency with the idea of social justice. Drawing on scored examples of teacher candidates’ assessments and pupils’ work samples, the article shows that many teacher candidates created cognitively complex and authentic learning opportunities for their pupils and that when pupils had more complex classroom assignments, they produced higher quality work. The article concludes that although it is complex, it is possible to construct teacher education assessments, such as the TAPL, that focus on pupil learning outcomes in ways that are consistent with social justice, especially preparation for a democratic society.

Keywords: Teacher Education, Social Justice, Authentic Intellectual Work, Pupil Learning

Some critics reject the idea of “teacher education for social justice” because it emphasizes what they see as progressive and political goals at the expense of traditional academic learning goals (e.g., Crowe, 2008; MacDonald, 1998; Will, 2006). In direct contrast, many proponents argue that ensuring that all students have intellectually complex learning opportunities is a central part of teacher education for social justice (e.g., Cochran-Smith, 1999, 2006; Michelli & Keiser, 2005; Oakes & Lipton, 1999) and thus that promoting students’ learning is inherent and integral to the concept of social justice education. Exacerbating debates about whether teacher education for social justice is connected to academic learning goals is the fact that there are very few studies of teacher preparation programs committed to social justice where students’ learning is treated as a measurable outcome (Cochran-Smith, Davis, & Fries, 2004; Weideman, 2002).

¹ Corresponding author: Lynch School of Education, Boston College, Campion Hall, Room 111, 140 Commonwealth Avenue, Chestnut Hill, MA 02467; Tel: +1 (617) 552-4591; Fax: +1 (617) 552-1840; Email: cochrans@bc.edu

This article briefly discusses the concept of teacher education for social justice in terms of students' learning and then considers how this notion is or is not consistent with the current emphasis in teacher education on accountability for learning.² Next the article discusses the results of the "Teacher Assessment/Pupil Learning" (TAPL) analysis. This assessment was nested inside a larger qualitative study about learning to teach over time in a preparation program with a stated social justice agenda. The purpose of the TAPL analysis was to assess the intellectual quality of assessments created or used by teacher candidates during the student teaching period and to assess the quality of their pupils' responses to those assessments. The project used Newmann and Associates' (1996) framework of "authentic teaching and learning" because of its general consistency with the idea of social justice, which is elaborated below.

Building on one another, the multiple layers of analysis in this article have several purposes. Details about the TAPL analysis help to make the argument that, although complex, it is possible to construct teacher education assessments focusing on pupil learning outcomes in ways that are consistent with social justice. The results of this assessment help to demonstrate that many of the teacher candidates we studied did indeed create cognitively complex, authentic, and demanding learning opportunities for their pupils. Our analysis also indicates that K-12 pupils who were provided with more cognitively complex classroom assignments were more likely to produce higher quality work. Finally, by juxtaposing our evaluation of teacher candidates' assessments and pupils' work with additional interview and other qualitative data, we show that teacher candidates also held themselves accountable for pupils' learning. They did so by thoughtfully scrutinizing their practices and the larger conditions that supported or constrained pupils' learning. At the same time they also considered the implications for their future practice.

The bottom line of our argument in this article is that it is imperative that teacher education programs with social justice agendas include appropriately complex measures of pupils' learning outcomes—and not simply measures of teacher candidates' learning or changes in beliefs—among the repertoire of ways they assess teacher candidates' performance and evaluate program effectiveness. We are *not* suggesting here that any teacher education program that promotes authentic intellectual learning or critical thinking skills is by definition a social justice program. Nor are we suggesting that promoting pupils' learning is the only goal of social justice education. Rather we are arguing that taking responsibility for pupils' learning is a necessary, but not sufficient, aspect of teacher education for social justice. This means that promoting pupils' academic learning and developing the skills needed for critical deliberation and problem solving, which are basic to participation in a democracy, must be an essential *part* of the mission of teacher education programs committed to social justice and must be among the ways those programs evaluate their success.

Teacher Education for Social Justice

"Social justice" has become a catch phrase in many teacher education programs in the U.S. over the last decade. Despite its appeal, there is great variation in how the term is

² For the remainder of this article, for the sake of clarity, we use the terms, "teacher candidates" to refer to prospective teachers who were students in college or university-based teacher education programs and the term "pupils" to refer to the K-12 students they taught.

used (Grant & Agosto, 2008; North, 2006; Zeichner, 2006), and critics have rightly argued that the concept is under-theorized (McDonald & Zeichner, 2009). Even taking these variations into account, in much of the literature on this topic, a distributive notion of justice is either implicit or explicit (Cochran-Smith, in press; North, 2006). That is, it is assumed that the bottom line of teaching is enhancing pupils' life chances by challenging the inequities of school and society (e.g., Adams, Bell, & Griffin, 1997; Ayers, Hunt, & Quinn, 1998; Cochran-Smith, 1999, 2004; Darling-Hammond, French, & Garcia-Lopez, 2002; Michelli & Keiser, 2005; Oakes & Lipton, 1999; Villegas & Lucas, 2002; Zeichner, 2003). This perspective is based on recognition of significant disparities in the distribution of educational opportunities, resources, achievement, and positive outcomes between minority and/or low-income pupils and their white, middle-class counterparts. This is coupled with the position that teachers should be both educators and advocates, committed to the democratic ideal and to diminishing inequities in school and society by helping to redistribute educational and other opportunities.

From a social justice perspective, it is paramount that all pupils—with full appreciation of differences in linguistic background, class, culture, gender, ability, and race—have access to knowledge. However, this also involves questioning what counts as knowledge in the first place, whose interests are served, whose perspectives are represented, and who decides what knowledge counts (Castenell & Pinar, 1993; Michelli, 2005). From this perspective, teacher education for social justice encompasses many pupil learning goals, including thinking critically, connecting knowledge to real-world problems and situations, challenging received knowledge, understanding multiple perspectives, debating diverse viewpoints, unpacking underlying assumptions, and engaging productively in cross-cultural discussion. To meet these goals, all pupils need learning opportunities that develop basic skills as well as deep knowledge and the attitudes and values necessary for participation in a democratic society—opportunities that have historically been reserved for the privileged.

The social justice agenda in teacher education has been criticized on a number of grounds, many of which are related to knowledge and ideology (e.g., Cochran-Smith, 2006; Cochran-Smith, Barnatt, Lahann, Shakman, & Terrell, 2009). The research in this area has also been critiqued, particularly because it is primarily small-scale, short-term, qualitative and focused on attitudes and beliefs (Hollins & Guzman, 2005; Sleeter, 2001). One of the most serious criticisms is that this research has failed to account for teacher education outcomes, particularly the outcome of pupils' learning (Cochran-Smith et al., 2004; Grant & Secada, 1990). The TAPL project we describe in this article is small-scale and thus shares the limits of other such studies. However, this study focuses on teacher candidates' performance, rather than simply their attitudes and beliefs, by looking directly at the assignments and assessments they created or used in classrooms. This study also focuses on pupils' learning by assessing their performance in response to the assessments teachers create.

Teacher Education and Accountability for Pupils' Learning

It is conceptually and methodologically complex to demonstrate direct linkages between teacher education programs and pathways, on the one hand, and pupils' learning, on the other (Cochran-Smith & Fries, 2005; Kennedy, 1999; Schallock, Schallock, & Ayres, 2006; Wasley & McDiarmid, 2004). However there is no question that one of the current trends in teacher preparation in the U.S. is the expectation that teacher preparation

should be accountable for pupils' learning. Contemporary efforts to hold teacher education accountable for pupils' learning fall into four categories: (1) correlating preparation programs and entry pathways with pupils' achievement scores; (2) evaluating programs in terms of candidates' demonstration of classroom behaviors correlated with pupils' test scores; (3) assessing the learning opportunities teacher candidates create along with pupils' performance; and, (4) assessing the learning opportunities program graduates create along with their pupils' performance. These approaches vary according to the positionality of those making judgments about accountability (e.g., external critics, state or national accreditors, internal assessors), what is considered evidence of pupils' learning, temporal proximity of evidence to the preparation period, pedagogical proximity of evidence to the daily work of teaching and learning, and the intended or unintended consequences of accountability.

Studies in the first category include large-scale initiatives that assess the impact of entry pathways or preparation programs on pupils' achievement using multiple regression, value-added, and related statistical analyses. The New York City Pathways Project, for example (Boyd, Grossman, Lankford, Loeb, Michelli, & Wycoff, 2006), studied the impact of the multiple pathways into teaching on pupils' achievement and the composition of the teacher workforce. Similarly, Louisiana's value-added assessment of preparation programs (Noell & Burns, 2006; Noell, Porter, & Patt, 2007) used a multivariate longitudinal database and mixed linear models to assess the state's preparation programs in terms of estimated contributions to pupils' achievement.

These and similar projects (e.g., Decker, Mayer, & Glazerman, 2004; Harris & Sass, 2007; Lasley, Siedentop, & Yinger, 2006) are intended to inform city, regional, or state policies (e.g., whether or not to allow "alternate" routes into teaching) or to mandate changes in under-performing programs. This approach generally relies on test scores as evidence of pupils' learning, which is removed from teacher preparation in time and space, and is also pedagogically distant from the daily work of teaching and learning. In fact, McNergney and Imig (2006) have argued that with this approach, external judges hold teacher education accountable for outcomes over which they have little control or immediate impact.

A second approach to teacher education accountability for pupils' learning is to assess the extent to which candidates demonstrate classroom behaviors that are correlated with pupils' achievement. This approach was robust from the 1960s to the 1980s when teachers were trained to display behaviors that process-product research on teaching had already correlated with test scores or other desirable outcomes (Cochran-Smith & Fries, 2005). A contemporary illustration comes from the University of Virginia, where teacher candidates are assessed via the Teaching Performance Record (McNergney, 2006) and the Classroom Assessment Scoring System (La Paro, Pianta, & Stuhlman, 2004; Rimm-Kaufman, La Paro, Downer, & Pianta, 2005) to determine the extent to which they "demonstrate teaching behavior that has been shown to relate to or cause PK-12 pupil learning" (McNergney & Imig, 2006). With this second approach, teacher educators themselves are generally the assessors who treat candidates' classroom behavior as evidence of accountability for pupils' learning. The indicator of teachers' classroom behavior has close temporal and spatial proximity to preparation as well as close pedagogical proximity to the daily work of teaching and learning. Results are used by preparation programs to inform institutional decisions (e.g., teacher education curriculum content or structural arrangements) or to evaluate individual candidates.

A third approach, which is also usually carried out within higher education institutions, is that of assessing teacher candidates in terms of their own analyses of pupils' learning. For example, with the teacher work sample method (Girod, 2002; Schalock, Schalock, & Girod, 1997), candidates teach a unit of study and assess pupils based on pre- and post-instructional measures. Newer assessments, such as teaching portfolios or classroom research projects, also require candidates to analyze lessons and interactions, using pupils' work as evidence of effectiveness. One example, which cuts across teacher preparation programs and institutions, is the Performance Assessment for California Teachers (PACT) (Darling-Hammond, 2006; Pecheone & Chung, 2006), which centers on the "teaching event" portfolio, scored by trained educators across institutions.

These and related performance assessments (e.g., Cochran-Smith, Barnatt, Friedman, & Pine, in press; Wasley & McDiarmid, 2004) use candidates' own analyses along with the evaluations of independent scorers as evidence of pupils' learning. This evidence has close temporal and spatial proximity to preparation and close pedagogical proximity to teachers' work. Unlike those in the second category, assessments in this third category require direct evidence of pupils' learning evaluated by candidates and others. This kind of assessment is generally used for high stakes decisions about candidates (e.g., program completion or recommendation for certification) as well as for identification and improvement of the strengths and weaknesses of programs.

Finally there are some efforts to hold teacher education accountable by evaluating the practices of program graduates during the early years of teaching, including the quality of learning opportunities they create and pupils' performance. Again, these are usually carried out by the agents of teacher education programs or pathways. At Bank Street College, for example, the Structure of the Observed Learning Outcomes (SOLO) instrument (Biggs & Collis, 1982) was used to assess both the cognitive demands of the tasks graduates assigned to pupils and their pupils' responses to those tasks. A second example is a Boston College project (Jong, Pedulla, Mitescu, Salomon-Fernandez, & Cochran-Smith, in press; Pedulla, Salomon-Fernandez, Mitescu, Jong, & Cochran-Smith, 2007), which analyzed graduates' practices and their pupil performance on math tests in comparison to those of graduates of an alternate program using the Reformed Teaching Observation Protocol (RTOP) (Piburn, Sawada, Falconer, Turley, Benford, & Bloom, 2000). These and similar approaches to holding teacher education accountable for pupils' learning (e.g., Darling-Hammond, 2006; Gore, Griffiths, & Ladwig, 2004; Schalock et al., 2006; Schalock & Schalock, 2004) regard graduates' practices and the performance of their pupils as evidence of pupils' learning. This evidence is distant from preparation in time (i.e., 1-3 years after teacher preparation) and space (i.e., classrooms that were not part of preservice preparation), but close to the daily work of teaching and learning since the focus is naturally occurring activity. These assessments are generally used to inform policy and practice at individual institutions, but have broader possibilities for local and larger policy.

In addition to variation in what counts as evidence of pupils' learning, these four approaches vary in terms of their consistency with social justice goals. The first two approaches use test scores, either directly or indirectly, as either the prime or one of a group of indicators of learning. In the U.S., given current federal testing requirements for sub-group accountability, some see this as a step toward educational equity and justice (e.g., Education Trust, 2008). However, there is also mounting evidence that under the current testing regime, poor and minority pupils may have more limited learning opportunities than before (Kantor & Lowe, 2006), and that there are serious unintended

consequences of testing related to drop-out and graduation rates, to the narrowing of the curriculum, and to schools ending up less accountable to their local communities (Center on Education Policy, 2006; Darling-Hammond, 2004; Meier & Wood, 2004; Orfield, Losen, & Wald, 2004).

Whether or not the third and fourth approaches are consistent with social justice agendas is contingent on how “quality of learning opportunities” and “quality of pupils’ learning” are defined and measured in each case. Although we have argued that a major goal of social justice education is providing access to rich knowledge for all pupils, this also includes questioning that knowledge, promoting critical thinking, and debating differing viewpoints. This means that the third and fourth approaches above may or may not be consistent with social justice agendas, depending on the specific assessments used.

Background and Description of the Study

The TAPL analysis is nested within a larger project, titled the Qualitative Case Studies (QCS) project. QCS is one of six studies in a portfolio created by the Evidence Team of the Boston College (BC) Teachers for a New Era (TNE) initiative.

Teacher Education at Boston College and “Teachers for a New Era”

A Jesuit university, BC serves some 15,000 undergraduate and graduate students with the Lynch School of Education preparing 250-275 undergraduate and graduate teachers per year. Its mission includes an explicit commitment to preparing teachers to teach for social justice by focusing on teachers’ and pupils’ learning (Boston College Lynch School of Education, 2008). Other program themes include: constructing knowledge with pupils through critical thinking, problem solving, and making connections across disciplines and perspectives; inquiring into practice through classroom research and developing an inquiry stance; affirming diversity as an asset and providing access to rich learning opportunities for all; and, collaborating with others to build support and diminish barriers to learning.

At BC, teacher education faculty members have engaged in deliberative inquiry into their own practice over the course of several years to articulate the social justice vision of the program. What emerged from their work was an understanding of teaching for social justice as “an activity with political dimensions . . . [in which] all educators [are] responsible for challenging inequities in the social order and working with others to establish a more just society” (Boston College Lynch School of Education, 2008). A central goal of teaching for social justice is to improve pupils’ learning and enhance their life chances by challenging school and societal inequities. In addition to methods courses and field experiences, candidates at the master’s level, who are the focus of this analysis, take foundations courses in the social contexts of education and teaching pupils with diverse and special needs, as well as an inquiry seminar focused on classroom research. All candidates have teaching placements in schools with diverse populations and also complete fieldwork with bilingual pupils. The capstone inquiry project requires candidates to pose a question about the impact of their teaching on pupils’ learning, collect multiple data sources, and interpret these in terms of guidelines for their own practice and commitments to social justice.

At the time of this study, BC was one of eleven sites of TNE, a national teacher education initiative funded primarily by the Carnegie Corporation to reform teacher

education in universities. As part of TNE at BC, an interdisciplinary Evidence Team³ was charged with developing outcomes measures and generating evidence about BC's teacher education programs.

The Qualitative Case Studies Project

The QCS project is a cross-case longitudinal study of learning to teach, which is based on 22 cases of teacher candidates' learning during the preservice period and early years of teaching. Specifically QCS was designed to explore relationships among: teacher candidates' entry characteristics; what they learned from coursework and fieldwork; their perceptions and understandings over time about teaching, pupils' learning, and social justice; their classroom practices and their pupils' learning; and, career paths over time. Participants included both elementary and secondary teachers in multiple subject areas.

The QCS project works from a framework based on critical sociocultural theory (Eisenhart, 2001) using methods drawn from critical ethnographic research (Erickson, 1986) and multi-site cross case analysis (Ayres, Kavanaugh, & Kraft, 2008; Stake, 2006). Data sources include: in-depth structured interviews over time with candidates, supervisors, mentors, and principals; detailed classroom observations and related class materials; assessments/assignments used by teacher candidates/teachers and the accompanying pupils' work; and, candidates' work and program materials during the preservice period. QCS uses standard methods of qualitative data analysis, and a "consensual" approach (Hill, Thompson, & Williams, 1997) where research is conducted using the same protocols across cases by a team of researchers who inductively arrive at "consensus judgments" (p. 521) about data coding and frameworks for analysis.

TAPL Research Design

The TAPL analysis used the rich detail from the QCS project as a backdrop for examining both the learning opportunities created by teacher candidates and their pupils' performance. This analysis is consistent with the third approach we described above to holding teacher education accountable for pupil learning in that it focuses on external evaluations of candidates' practices and their pupils' performance along with candidates' own analyses of these.

The purpose of the TAPL analysis was to examine the intellectual quality of the learning opportunities teacher candidates created for their pupils by evaluating the assessments and assignments they used in their student teaching classrooms through the lens of Newmann and Associates' (1996) framework of authentic intellectual work. In this study, we use the word "assessment" to mean any classroom activity or assignment that provides teachers with information about what their pupils are learning. Thus assessments might include worksheets for single lessons, culminating activities for units, or conventional tests. We also examined pupils' responses to these assessments. Specifically, the TAPL study addressed the following questions:

³ During the period when the work described in this article was carried out, the Evidence Team included Boston College faculty members and administrators Marilyn Cochran-Smith (chair), Alan Kafka, Fran Loftus, Larry Ludlow, Patrick McMullan, Joseph Pedulla, and Gerald Pine; TNE Administrators Jane Carter and Jeff Gilligan; and doctoral students Joan Barnatt, Robert Baroz, Lisa D'Souza, Sarah Enterline, Ann Marie Gleeson, Cindy Jong, Kara Mitchell, Emilie Mitescu, Aubrey Scheopner, Karen Shakman, Yves Salomon-Fernandez, and Dianna Terrell.

1. What is the intellectual quality of (a) assessment tasks created or used by teacher candidates and (b) pupils' learning as demonstrated on those assessments?
2. How does the intellectual quality of assessments and pupils' learning vary across candidates, subject matter, grade level, and school contexts?
3. What is the relationship between quality of assessments and pupil learning (i.e., Do higher quality assessments result in higher quality pupil work)?
4. What is the relationship between teacher candidates' own analyses of their assessments and pupils' learning and the analyses of independent evaluators?

The TAPL Protocol

Administration of the TAPL Protocol was part of the data collection strategy for the larger QCS study.⁴ Participants in the study collected class sets of pupils' work for one assessment that had been used as a final or culminating task for a topic or unit, as well as two assessments that had led up to the final task. Assessments and pupils' work were evaluated in two ways: (1) an "internal evaluation" that occurred in an interview setting wherein candidates analyzed their assessments and also reflected on pupils' work, and (2) an "external evaluation" wherein the authentic intellectual quality of the assessments and pupil work were assessed by researchers. Figure 1 represents the two parts of the TAPL.

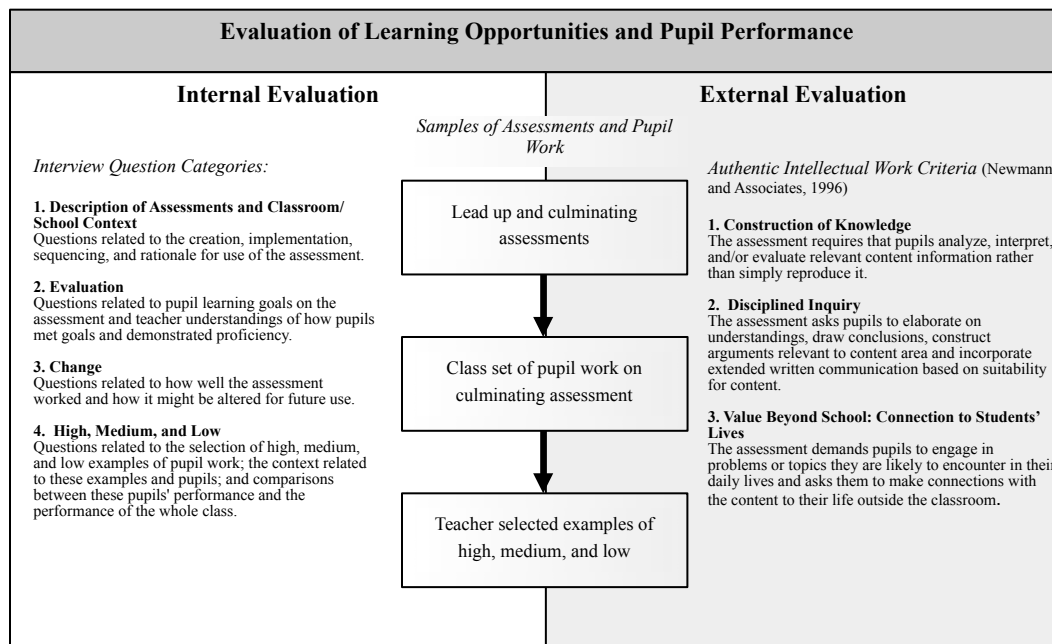


Figure 1. Teacher Assessment/Pupil Learning (TAPL) Protocol.

⁴ The TAPL protocol was used regularly as a repeated data collection strategy in the larger longitudinal study. This provided a way to examine changes in teachers' practices over time as they moved from student teaching into the first, second, and third years of teaching. This article focuses on data from the student teaching period only.

Internal evaluation. For the TAPL “internal evaluation,” candidates collected assessments and pupils’ work samples from the student teaching setting and reflected upon these in response to interview questions about their learning goals, teaching strategies, and construction of assessments. Teacher candidates also selected examples of “high,” “medium,” and “low” pupil work samples to describe how they evaluated pupil learning.

External evaluation. For the TAPL “external evaluation,” candidates’ assessments and pupils’ learning were evaluated by researchers using the framework of “authentic intellectual work” (Newmann & Associates, 1996). The concept of “authentic intellectual work” is derived from the ways adults use knowledge in the workplace and as citizens in a democratic society and is grounded in the idea that in contemporary society and professional communities, important types of knowledge are used beyond school. As schools prepare pupils for life in this society, then, school experiences should be as close to “real” world experiences as possible, so that pupils become “problem solvers and lifelong learners capable of adapting to changing economic and social conditions” (p. 28). To do this, all pupils ought to have meaningful and significant opportunities to engage in disciplined, rigorous, and critical educational experiences.

Newmann and Associates (1996) define authentic intellectual work through three criteria. The first criterion, *construction of knowledge*, refers to intellectual activities that involve applying information to new situations and constructing meaning and new knowledge. The second criterion, *disciplined inquiry*, involves drawing upon an extensive knowledge base, developing in-depth understandings of material, and expressing understandings through writing or other extensive communications. The third criterion, *value beyond school*, emphasizes connections to the larger world. All three criteria must be present for assessments and assignments to be considered exemplars of authentic intellectual work.

To further articulate the three criteria for authentic intellectual work, Newmann and Associates (1996) developed standards to evaluate teachers’ assessments and pupils’ work, which cut across elementary and secondary grades and apply to math, science, writing, and social studies. For our analysis, we used the rubric developed by the Research Institute on Secondary Education Reform (RISER) (2001), based on the authentic intellectual work standards in Newmann, Secada, and Wehlage (1995) and Newmann, Lopez, and Bryk (1998). According to the RISER rubric, scores can range from 3.0 to 10.0 for assessment tasks and from 3.0 to 12.0 for pupils’ work. Our decision to utilize the framework of authentic intellectual work was based on its general consistency with our conceptualization of teacher education for social justice at Boston College. As we conceptualize it, a necessary, but not sufficient, goal of teaching for social justice is enhancing the learning of all pupils and providing all with rich opportunities for critical evaluation, disciplined inquiry, and deliberation, traits that are consistent with the notion of authentic intellectual work. In addition, the authentic intellectual work framework has been used in other studies to measure teaching and school outcomes (e.g., Gore et al., 2004; King, Schroeder, & Chawszczewski, 2001; Newmann et al., 1998; Queensland School Reform Longitudinal Study, 2001), and our work aims to extend this research to teacher education.

For the external evaluation, 18 assessment tasks and 158 pieces of pupil work were scored according to the rubric by a team of researchers, each of whom had teaching experience and subject matter expertise in at least one content area. Team members were trained by Bruce King, one of the developers of the rubric. Two raters scored all of the

assessment tasks and were within one point of agreement 90% of the time. Pupils' work was scored by pairs of raters; pairs were within one point of agreement 91% of the time.

Participants. In this article we analyze data gathered with the TAPL protocol for 18 teacher candidates involved in the QCS study. Participants included elementary and secondary teachers across subject areas. More than half of the candidates taught in high-poverty schools comprised primarily of pupils of color, while, at the other end of the spectrum, four candidates taught in predominantly middle-class schools where the majority of students were white. Since the TAPL analysis was embedded within the QCS project, the sample for the TAPL study was limited to the participants already involved in the case studies project, which, by design, followed a relatively small group of teacher candidates over a long period of time. The resultant small sample generally prohibited us from establishing statistical significance for correlations between assessments and pupil learning scores.

Teacher Candidates' Creation of Learning Opportunities and Pupils' Performance

As noted, our aim for this study was to examine the intellectual quality of assessments created by teacher candidates along with their pupils' performance on those assessments. Our emphasis was on the external TAPL authentic intellectual work scores. We examined scores across candidates and by grade level, subject matter, and classroom context. We also looked at the relationship between assessment task scores and scores on pupils' work. Finally, we used the internal TAPL to compare candidates' own analyses of their assessment tasks and pupils' learning with scores on the external TAPL.

Intellectual Quality of Assessment Tasks and Pupil Performance

Creating learning opportunities: Teacher candidates' assessment tasks. Teacher candidates submitted a variety of "culminating" assessment tasks, which are organized by grade and subject and listed on Table 1. These ranged from multiple choice exams and fill-in-the-blank worksheets to written essays and less traditional classroom activities, such as a mock trial related to *Romeo and Juliet*. As Figure 2 indicates, authentic intellectual work scores for these assessments ranged from 5.0 to 10.0 across the 18 assessments, with a mean of 7.5 ($SD=1.66$). On average, the teacher candidates produced assessment tasks with "moderate to high" levels of authentic intellectual quality (RISER, 2001). This mean score is comparable to the mean scores of assessments created by experienced teachers in previous studies that have utilized this scale (see Bryk, Nagaoka, & Newmann, 2000; King et al., 2001).

In our study, low-scored assessment tasks tended to be multiple-choice exams asking pupils to recall discrete pieces of information. For example, a ninth-grade physics assessment, which was scored a 6.0 (out of 10) and was one of the lowest-scored tasks in our study, consisted of 23 multiple-choice and three multi-part short answer questions. Two of the multiple-choice and two of the short answer questions are shown in Figure 3. The multiple choice questions, illustrated by Questions 1 and 2, were primarily recall questions, requiring pupils to remember the formula for distance and electrical force and identify the properties of an atomic nuclei. More authentic intellectual work would have included higher order thinking questions asking pupils to apply the formula or create a diagram of an atom. Although Questions 3 and 4 involve more complex levels of thinking by asking pupils to apply their knowledge of current and resistance, this is not the dominant expectation of the assessment task. The low score is explained by close

Teacher Candidate	Grade/Subject	Assessment Task
1	2 nd /Writing-ELA	Writing assignment where pupils wrote stories about objects that added up to 100
2	2 nd /Writing-ELA	“Story Buddy” writing assignment where pupils create their own story
3	2 nd /Writing-ELA	Worksheets on the conventions of non-fiction literature
4	4 th /Writing-ELA	Anthology of original poems
5	4 th /Social Studies	Essay on the Civil Rights Movement
6	4 th /Math	Math test on geometry
7	5 th /Math	Math test on numbers
8	6 th /Writing-ELA	Essay on the book <i>Blessed are the Peacemakers</i>
9	H.S./Writing-ELA	Test and essay on the novel <i>The Bean Trees</i>
10	H.S./Writing-ELA	Essay on the novel <i>A Lesson Before Dying</i>
11	H.S./Writing-ELA	Legal briefs and mock trial on the play <i>Romeo and Juliet</i>
12	H.S./Writing-ELA	Essay on the play <i>Death of a Salesman</i>
13	H.S./Writing-ELA	Essay on <i>All Quiet on the Western Front</i>
14	H.S./Social Studies	U.S. History Exam
15	H.S./Social Studies	PowerPoint on Renaissance artists
16	H.S./Math	Math test on inequalities
17	10 th /Science	Worksheet on “Complex Patterns of Heredity”
18	H.S./Science	Physics test on electricity

Table 1. Description of Assessment Tasks

examination of the RISER standards. For the standard “construction of knowledge,” the physics assessment scored a 2.0 (out of 3.0)—and not the lowest score of 1.0—because some of the questions required an application of knowledge beyond factual recall. On the standard, “elaborated written communication,” this task scored a 2.0 (out of 4.0) because the task contained a mixture of multiple-choice and short answer questions. On the final standard, “connection to pupils’ lives,” this task scored a 2.0 (out of 3.0). Although the majority of the questions were abstract and removed from the daily lives of pupils, a few were related to situations involving static and electricity that pupils could come into contact with. Here, and in other low scoring assessment tasks, there were few opportunities for authentic intellectual work, and pupils did not produce high quality intellectual work in response. In fact, the class mean score for pupils’ work on this physics assessment, which we explain in the next section, was 5.0 (out of 12.0), which was the lowest pupils’ work score in this study.

In contrast, higher-scored assessments offered more opportunities for authentic intellectual work, as illustrated by a fourth-grade writing assignment, which received a score of 10.0, the highest possible score. For this assessment, represented in Figure 4, pupils were asked to compose different types of poems, which were compiled into a poetry anthology. In addition, pupils wrote extended passages about poems they had read

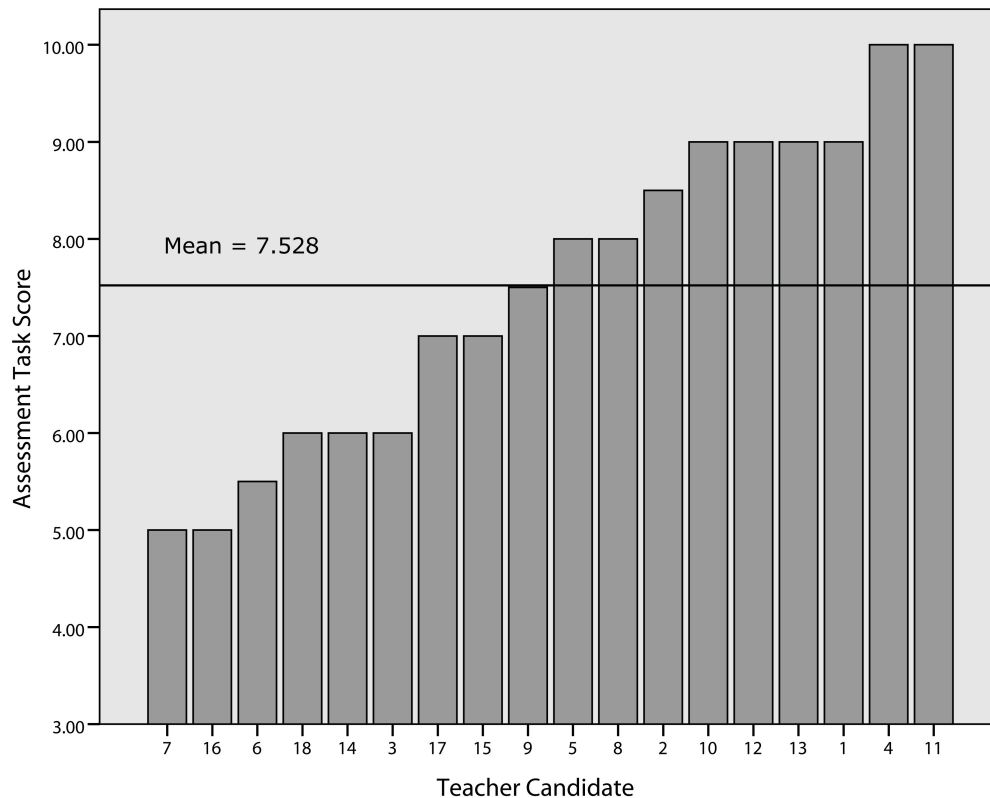


Figure 2. Assessment Task Authentic Intellectual Work Scores.

in class, providing supporting evidence about their favorites. On the standard “construction of knowledge,” this assessment received a 3.0 (out of 3.0). Pupils were required to interpret information and construct knowledge as they created poems. On the standard “elaborated written communication,” the task scored a 4.0 (out of 4.0) since pupils were asked to illustrate their understanding of poetry conventions by writing in distinct styles. On the final standard “connection to pupils’ lives,” this task scored a 3.0 (out of 3.0) since pupils wrote poems about their own experiences, as illustrated by the poem in Figure 4 about a pupil’s guitar. With this task, pupils had the opportunity to construct new knowledge, elaborate and support generalizations, and draw on experiences in their own lives. Pupil work on this task was scored higher than the physics example above; the mean pupil work score on the poetry assignment was 7.3 (out of 12.0), which was slightly above the mean score for all pupils’ work.

Response to opportunity: Pupils’ work. Like the scores on the assessments teacher candidates created, their pupils’ work in response to these assessments varied, but tended to fall within the middle of the authenticity scale. As Figure 5 indicates, the mean for all pupils’ work ($n=158$) was 7.0 (out of 12.0) ($SD=2.15$), ranging from 3.0, the lowest possible score, to 12.0, the highest. On average, pupils’ scores represent “moderate” levels of authentic intellectual work (RISER, 2001), which is comparable to levels obtained in other studies using this scale (King et al., 2001).

Description: 9th grade physics test on electricity, Teacher Candidate 18.

Multiple Choice

- | | |
|--|---|
| <p>1. As the distance between two charged objects increases, the electrical force between them</p> <p>A. decreases.</p> <p>B. increases.</p> <p>C. remains the same.</p> <p>D. usually cannot be determined.</p> | <p>2. Atomic nuclei of almost all atoms consist of</p> <p>A. only protons.</p> <p>B. protons and neutrons.</p> <p>C. protons and electrons.</p> <p>D. neutrons and electrons.</p> |
|--|---|

Short Answer

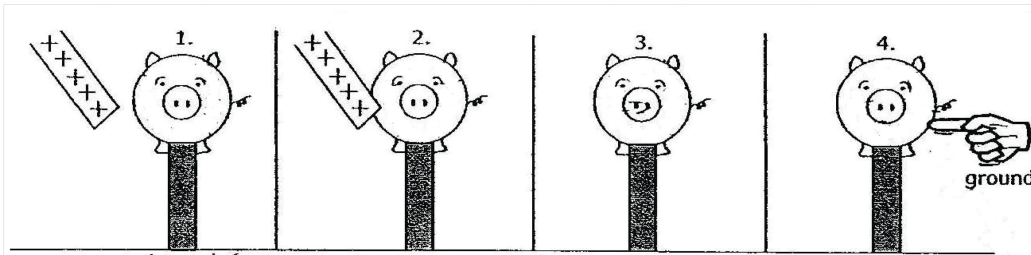
3. Mr. [N] is listening to AC/DC's 1974 debut album, High Voltage, on his 60-watt stereo.
- A. Calculate the current flowing through the stereo when it operates on 120 volts.
- B. Calculate the resistance of the stereo.

Show all of your work. Include units with your final answers.

4. A metal conducting sphere in the shape of Fuzzy the Pig is situated atop an insulating stand as shown below.

- (a) Draw the orientation and/or transfer of charge in Fuzzy for each step in the sequence below.
- (b) Below each step, briefly explain what is going on in terms of charge.

Note: Fuzzy is initially neutral; he is in contact with the positively- charged rod in step 2.



(c) What type of charging, if any, occurs in this sequence? _____

Assessment Task Score: 6.0 (out of 10.0)

Construction of Knowledge: 2.0

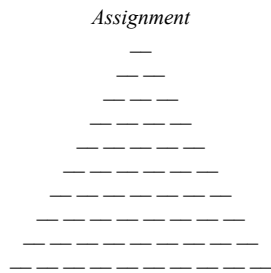
Elaborated Written Communication: 2.0

Connection to Students' Lives: 2.0

Figure 3. Physics Test on Electricity.

Description: Students wrote approximately six different types of poems and compiled their poems into a Poetry Anthology booklet. Two poem examples included a Pyramid Poem and Favorite Poem worksheet.

Pyramid Poem: Pupils were given a worksheet in the shape of a pyramid. Pupils were told to write a word on each line. The first line is the topic of the poem and each line has one more word than the previous line. On the left is the pyramid worksheet that pupils were given. On the right is an example of a pyramid poem written by a pupil for this assessment.



Pupil Example

Guitar
 Frets Chords
 red black white
 Dad listening Mom listening
 [Chad Bach Town] Music School
 Eat a darn good breakfast early
 Practicing over and over and over again
 I touch strings my guitar and my bed
 I smell my guitar music and my cool room
 Purple Haze, Fire, Sunshine of Your Love, Stairway to heaven

Poem Worksheet

Favorite Poem Worksheet*

Poem Title:

Poet:

Tell what the poem is about.

What did you notice about this poem?

Why did you choose this poem as one of your favorites?

*The original worksheet for this assessment had lines where pupils could write their answers to the questions. There was enough room for pupils to write their answers in paragraph form.

Figure 4. Poetry Anthology Assignment, Teacher Candidate 4.

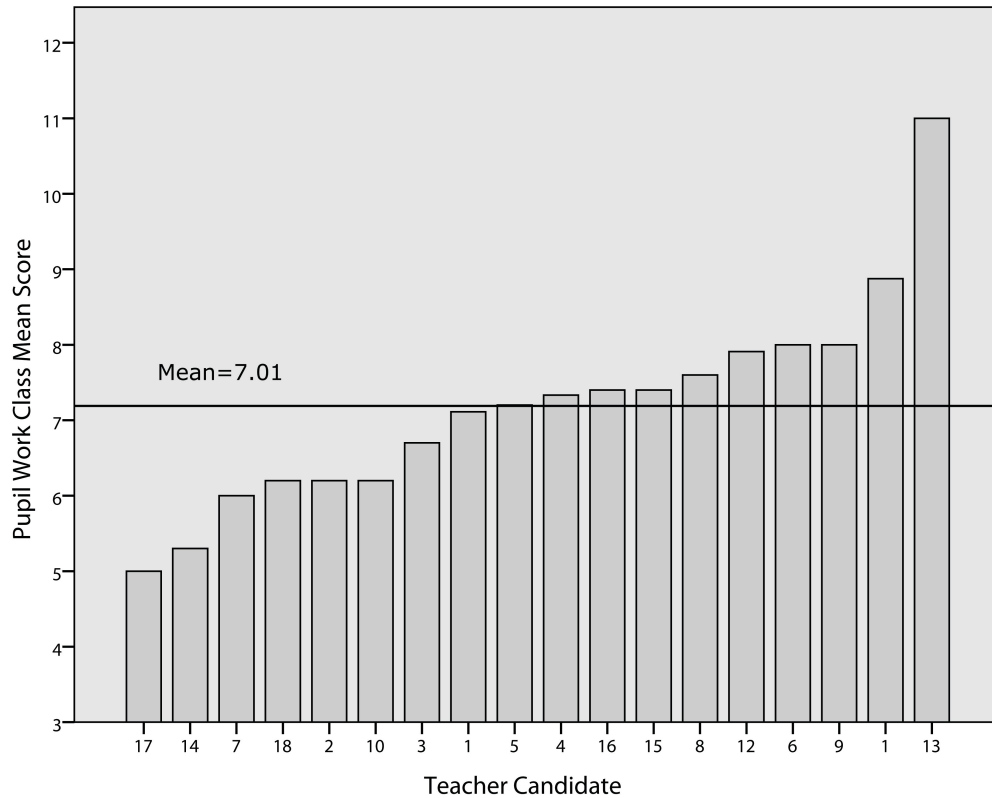


Figure 5. Pupils’ Work Scores.

Two examples of pupils’ work on a fourth-grade math assessment, which are shown in Figure 6, illustrate the range of scores. This math assessment task was one of the lowest-scored in our study, scoring a 5.5 (out of 10.0). It is worth noting here that low-scored assessments often result in low-scored pupils’ work because it is difficult for pupils to go beyond what the task calls for, although this is not necessarily because pupils are unable to perform at higher levels.

On this task, Pupil A received a score of 5.0 (out of 12.0), while Pupil B received a score of 9.0. The first standard, “construction of knowledge/mathematical analysis,” evaluates the ability to go beyond reproduction of information and algorithms. As Figure 6 shows, Pupil A, who received a score of 2.0 (out of 3.0), exhibited *some* analysis by interpreting similarities between shapes, but did not enumerate the differences. Pupil B, on the other hand, who received a score of 3.0, compared and contrasted the shapes. The second criterion, “disciplinary concepts,” measures pupils’ understanding of mathematical concepts. Again, Pupil A demonstrated *some* understanding of concepts by stating that triangles have 180° and quadrilaterals have 360° and thus received a score of 2.0 (out of 3.0). Pupil B, however, used the terms polygon, quadrilaterals, and trapezoid and noted differences in relation to a right triangle. Although Pupil B incorrectly labeled

Description: 4th grade math test about polygons submitted by Teacher Candidate 6. The assessment task scored a 5.5 (out of 10.0).

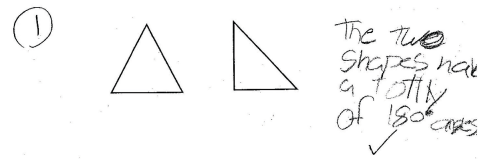
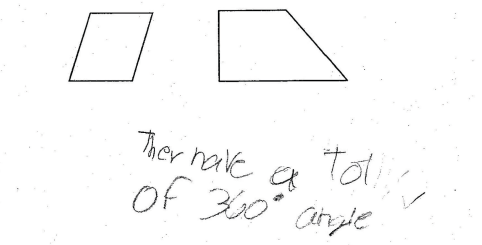
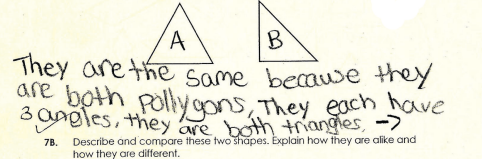
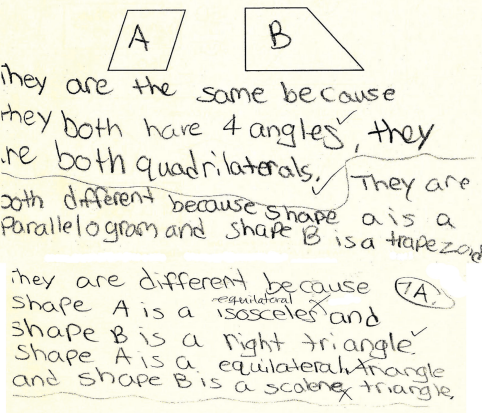
Pupil A	Pupil B
<p>Part III Open Response: Show all your work (drawing, tables, and/or computations) in the space provided. If you do the work in your head, explain in writing how you did the work.</p> <p>7A. Describe and compare these two shapes. Explain how they are alike and how they are different.</p>  <p>7B. Describe and compare these two shapes. Explain how they are alike and how they are different.</p> 	<p>Part III Open Response: Show all your work (drawing, tables, and/or computations) in the space provided. If you do the work in your head, explain in writing how you did the work.</p> <p>7A. Describe and compare these two shapes. Explain how they are alike and how they are different.</p>  <p>7B. Describe and compare these two shapes. Explain how they are alike and how they are different.</p> 
<p>Pupils' Work Score: 5.0 (out of 12.0)</p> <p>Analysis: 2.0</p> <p>Disciplinary Concepts: 2.0</p> <p>Elaborated Written Communication: 1.0</p>	<p>Pupils' Work Score: 9 (out of 12.0)</p> <p>Analysis: 3.0</p> <p>Disciplinary Concepts: 3.0</p> <p>Elaborated Written Communication: 3.0</p>

Figure 6. Pupils' Work Examples.

the triangle, the majority of the work shown in this figure as well as the work not shown here was correct; thus the pupil received a score of 3.0. On the third standard “elaborated written communication,” pupils are evaluated on their understanding of math through words or diagrams. Pupil A, whose answer was incomplete and unclear, scored a 1.0 (out of 3.0) while Pupil B scored a 3.0 because he/she introduced what was being referred to, enumerated similarities and differences, and used mathematical terms, even going beyond the task by using space on the back of the worksheet to continue the answer. The differences between the scores of these two pupils are defined by depth of conceptual knowledge and articulation of understandings.

Teacher candidate-created assessments. As we noted in our discussion of the TAPL research design, the assessment tasks in this study were selected by teacher candidates

themselves who were instructed to submit a “culminating” assessment. These varied considerably with 7 of the 18 candidates submitting tasks that were supplied by and/or mandated by the school or school district. For example, one candidate selected a math assessment from a highly-structured curriculum to which the school district subscribed. Since part of our interest in this study was to examine the learning opportunities that teacher candidates make available to their pupils, we did not include tasks that candidates had no role in creating in the next two analyses we offer below. However, we do include the mandated assessments in the final section of this article as part of our consideration of relationships between the internal and external evaluations of assessment tasks and pupils’ learning.

There were eleven assessment tasks created by teacher candidates themselves. These differ slightly from the larger group of 18 assessments. All of the math assessments selected by teacher candidates as part of this study were school- or district-mandated. In terms of overall authenticity scores, the mean score for candidate-created assessments was 8.1 (out of 10.0) ($SD=1.46$) out of a possible 10, which was higher, but not statistically significantly different from, the mean of all 18 tasks (mean=7.5; $SD=1.66$). A sample of 88 pieces of pupils’ work created in response to the 11 candidate-created assessments was analyzed. The mean score for pupils’ work was 7.1 (out of 12.0) ($SD=2.14$), which was not significantly different from the mean for all 18 tasks (mean=7.01).

Differences in Authentic Intellectual Work

In comparing assessment tasks and pupils’ work scores across grade levels, we found no statistical differences between elementary and secondary teacher candidates. However, the quality of authentic intellectual work did appear to be influenced by subject matter. It should again be noted that there were no math assessments in this part of the analysis because all of the submitted math assessments were supplied or mandated by the school or district. That said, there was a difference, although not statistically significant, between assessment tasks in writing and science, where writing assessments were scored higher than science assessments. Pupils’ work scores were also higher in writing than in science, statistically significant at $p<.05$. In both assessment tasks and pupils’ work, social studies scores fell between scores for writing and science. Our findings are consistent with similar research (King et al., 2001; Newmann et al., 1998) concluding that math and science assessments, as a whole, tended to emphasize lower order skills and did not always call upon pupils to support their findings in an elaborated way. In addition, “elaborated written communication” is one of the standards on the authentic intellectual work scale, and although the framework acknowledges that this may be demonstrated through “prose, graphs, diagrams, equations, or sketches” (Newmann, King, & Carmichael, 2007, p. 49), it may be more difficult to attain this standard in math and science, particularly on conventional standardized tests. We would also like to note here that, as a part of the QCS research design, we only collected pupil work samples where pupils produced a tangible artifact, and although we recognize that there are non-written ways to demonstrate authentic knowledge (such as oral presentations or dramatic performances), we limited our analysis to work samples that included a written component.

Interestingly, as mentioned above, we found differences between teacher candidate-created assessments and assessments supplied or mandated by schools or school districts.

Assessment tasks and pupil's work were scored higher when they were created by candidates. The mean assessment task score for teacher candidate-created assessments was 8.1 (out of 10.0) ($SD=1.46$), while the mean for mandated assessments was 6.6 ($SD=1.65$). Although this difference is not significant ($\alpha=.05$ level), it does provide some indication that, when given the opportunity, teacher candidates created assessments that were more intellectually challenging and authentic than district-required assessments. The pupils' work scores for the candidate-created assessments also had a higher, but not statistically significant, mean—7.2 (out of 12.0) than the pupils' work scores from the mandated assessments (6.8 out 12.0). This small difference in the authenticity of pupils' work becomes more important as we examine the relationship between learning opportunities and pupils' performance.

Relationship Between Learning Opportunities and Pupils' Outcomes

As Figure 7 suggests, when we compared teacher candidate-created assessment task scores with pupils' work mean scores for the class, we found a positive correlation ($r=.52$). Although the correlation was not statistically significant at the $\alpha=.05$ level, this does suggest that pupils who were offered higher quality assessment tasks were more likely to produce work of higher intellectual quality and vice versa. Another way to consider this relationship is to look at pupil work from high scoring and low scoring assessment tasks. Pupil work in response to assessment tasks that scored above the mean were higher, in a statistically significant way ($p<.05$), than the pupil work scores in response to assessment tasks that fell below the mean. In short, when pupils were offered authentic learning opportunities, they were more likely to produce more authentic intellectual work. This is consistent with other studies using the Newmann scales, which have found significant positive correlations between quality of assessment tasks and pupils' work (King et al., 2001; Newmann et al., 1998) and with studies indicating a correlation between access to authentic learning opportunities and higher scores on standardized tests (Newmann, Bryk, & Nagaoka, 2001).

Teacher Candidates' Perceptions of Learning Opportunities and Pupils' Outcomes

A distinct feature of the TAPL protocol is the combination of external and internal evaluations of teachers' assessment tasks and pupils' work and the use of both qualitative and quantitative data. In the final section of this article, we concentrate on the internal evaluations—that is, teacher candidates' own analyses of assessments and pupil performance—which were constructed during interviews. The results of the internal analyses were similar to those of the external analyses; the former also provided insights into teacher candidate's efforts to support the learning of all of the pupils in their classrooms.

Consistency between teacher candidates' analyses and scores. We found a strong consistency between teacher candidates' own analyses of the assessments they created or used in their student teaching classrooms and their pupils' work in response to these assessments, on the one hand, and the external evaluations using the scales for intellectual work, on the other. In short, assessments and pupils' work that were scored highly by external evaluators were also regarded highly by teacher candidates themselves. By the same token, in instances where assessments and pupils' work were scored low by external evaluators, teacher candidates also identified weaknesses in the assessments or described classroom conditions that negatively influenced pupils'

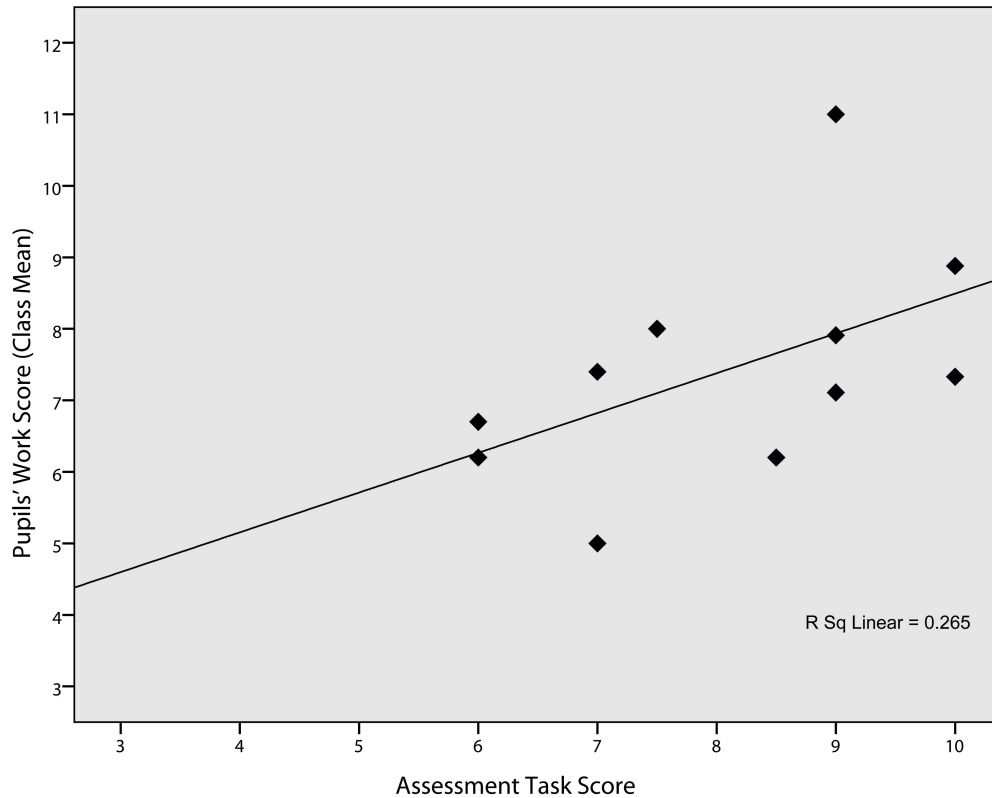


Figure 7. Correlation Between Assessment Tasks and Pupils' Work

performance. This consistency suggests that teacher candidates were working toward the goal of high levels of authentic intellectual work and could recognize when they did and did not meet this goal.

To illustrate further, we look at Teacher Candidate 11, a ninth-grade English teacher, whose culminating assessment of her class's study of *Romeo and Juliet* required pupils to write legal briefs and engage in a mock trial about the culpability of various characters in the deaths of the star-crossed lovers. The task received a score of 10.0 (out of 10.0) from external evaluators on the authentic intellectual work scale, and the mean pupils' work score was 8.88 (out of 12.0), the second highest class mean for pupils' work. During the TAPL interview, this teacher candidate highlighted the high quality of her pupils' work, stating, "The assignment in general was probably the best pupil writing all year. . . . And the best pupil presentations too." The teacher candidate indicated that her purpose for the assessment task was to have pupils "think like lawyers" by using evidence to support their arguments, suggesting a complex and nuanced understanding of pupil learning. Describing one pupil who had done uncharacteristically well, she said, "He's not the most productive, but he was actually really into this assignment and he worked the hardest that he's ever worked all year." She pointed out that the pupil's motivation was the competitive aspect of being a trial lawyer where real judges, whom the candidate invited to class, made decisions based on the soundness of his argument. The teacher candidate

commented, “For him it totally worked and for a lot of the pupils they knew that there were going to be real people there. Real people are going to see this. It was fun and there was something at stake, you know?” It was clear from this and other interviews about high scoring tasks that candidates had similar learning goals to the criteria for authentic intellectual work.

In contrast, second-grade Teacher Candidate 3 submitted a worksheet on non-fiction writing conventions as her culminating assessment; this (and the accompanying pupils’ work) received low scores from external reviewers. In the TAPL interview, this candidate spoke about her disappointment with the assessment, which she realized did not accurately capture her pupils’ learning or the type of work pupils had engaged in during the instructional sequence. She described the assessment as problematic and potentially confusing for pupils. The teacher candidate’s own evaluation of the mismatch between the instructional unit and the assessment was supported by the fact that her lead-up assessments scored higher on the external evaluation than her culminating assessment. In the interview, the candidate considered how she might alter her practice in the future; in doing so, she revealed accurate awareness of her pupils’ knowledge levels and confusions, and she had clear ideas about how to advance pupil performance in the future. The interview revealed that the candidate’s understanding of pupils’ learning was deeper than the low score on her assessment task might have suggested.

This case demonstrates the importance of using multiple measures of teacher candidates’ and pupils’ performance and the danger of using a single assessment when holding teacher education accountable for pupils’ learning. Low scores on assessment tasks and pupils’ work might suggest, at least on the surface, that a particular candidate was not a high performing teacher. However, as the interview above and other data sources illustrate, that same candidate might indeed be thinking critically and carefully about pupils’ learning. Along these lines, the candidate in the example above considered how to alter her problematic assessment, analyzed what was wrong with its content, and knew precisely what pupils did not understand. This and other examples support our larger argument that, in order to understand the process of learning to teach, a portfolio of evaluative measures is necessary. An instrument such as the TAPL protocol where both quantitative and qualitative data are used, can account for some of the complexities and conditions that influence teaching and learning and thus paints a more comprehensive picture of teacher candidates’ knowledge, skills, and practices than any single measure of teachers’ impact on pupils’ learning can.

Mandated assessments. TAPL’s internal-external approach was particularly useful in evaluating assessments that had been supplied or mandated by the school or district in which candidates were student teaching. Had we used an external evaluation alone for these, we would have had little information about teacher candidates’ own abilities to provide high quality learning opportunities. However, TAPL’s interview protocol provided a mechanism for getting at candidates’ understanding even when using the mandated assessments that are increasingly part of scripted curricula. For example, Teacher Candidate 7 submitted a fifth-grade mandated test from the scripted math curriculum used in her student teaching school. On the authentic intellectual work scale, the assessment received a score of 5.0 (out of 10.0), but this relatively low score (and the low pupils’ work score) reveal very little about the candidate since she did not create the assessment. However, during the interview, the candidate described a successful unit based on innovative instructional strategies where students had significant success on the lead-up assessments. She described the students’ excitement about their progress and

their eagerness to demonstrate this on the culminating assessment. She also explained that due to spring break and statewide testing schedules, the pupils had had a three week break between the unit and the final assessment. The candidate identified the large gap in time between instruction and assessment as a pivotal factor in pupils' performance.

This candidate's designation of high, medium, and low pupil work samples was also revealing. For example, she provided a detailed explanation of the strengths and weaknesses of a "medium" level piece of work (See Figure 8), clarifying that the pupil understood division but chose an inappropriate method to solve the problem by using a number that could not be easily divided. She also connected the innovative instructional methods that had been used in the unit to some of the "high" level work samples from students who did not usually receive high grades in math.

Teacher Candidate 7: And then this is someone who got into the medium group. She was able to do the multiples and she could count backwards. She understands subtraction, although she got the wrong answer. Same thing with division. She kinda gets clusters* but not the right answer....

Interviewer: Ok, so this problem is 357 divided by 21.

Teacher Candidate 7: And so here she tried to break down the 357 into 300 and 57 and then you divide those. But the problem with what she did is that she just didn't do good clusters, like this isn't a friendly cluster. 300 divided by 21 doesn't equal a round number, it has a remainder.

* Cluster problems are related problems that highlight flexible approaches to computation (Van de Walle, 2007). They provide students with opportunities to think about problems with different starting points and encourage varied solutions.

Figure 8. "Medium" Example of Pupils' Work.

Conclusions and Directions Forward

Currently many teacher education programs in the U.S. and elsewhere are involved in efforts to become more accountable for pupils' learning in order to meet their own standards and also meet the requirements of various external accrediting and evaluating bodies. The ideas in this article are intended to address many of the current debates about teacher education accountability, particularly as this relates to preparation programs with social justice agendas. In short, our position is that all teacher education programs and pathways—including those with social justice agendas—should indeed hold themselves accountable for the outcomes and impacts of their work, including their efforts to prepare teachers who promote pupils' learning. Indeed, we suggest that unless all pupils have

both rich learning opportunities and positive learning outcomes, then there is no social justice. Perhaps more importantly, however, we believe that we need very careful examinations of what it actually means to hold teacher education accountable for learning, and we need thoughtful analyses of the assumptions underlying the various arguments that are made in this area. We conclude with four points.

First, this article makes the case that it is possible—and desirable—for a teacher education program with a social justice agenda to measure pupils' learning outcomes in ways that are consistent with that agenda and that take into account some of the complexities of teaching and learning in schools and classrooms. By examining the quality of the opportunities pupils have to engage in authentic intellectual work, along with their responses to those opportunities, we have attempted to assess whether and how pupils are engaged in critical, higher order thinking and rigorous and relevant intellectual work. This kind of engagement is necessary for democratic participation in our society and is, in turn, central to teaching for social justice. This approach to constructing and measuring social justice as an outcome of teacher education knowingly fuses ideology and accountability, and in doing so, demonstrates that these two are *not* irreconcilable (as some critics would suggest) but are in fact, resonant in that they build on one another. This approach acknowledges head-on that teacher education for social justice is ideological in that it is based on particular ideas, ideals, values, and commitments regarding the kind of knowledge and thinking skills needed for democratic participation.

Second, the TAPL protocol's combination of internal self-reflection and independent external evaluation is one example of multiple measures of teacher candidates' performance. This combination provides a more complete understanding of how teacher candidates think about and try to support pupils' learning than any single measure could and thus underscores the importance of using multiple indicators to measure outcomes. The TAPL's interplay between internal and external evaluations underscores the importance of having a variety of measures to assess teacher candidate and pupils' learning. We found that the internal evaluation often contextualized the external evaluation by providing otherwise unavailable information about candidate's perceptions, practices, and skills in supporting pupils' learning.

Third, the TAPL protocol is one example of how individual teacher education institutions can measure and follow teacher candidates and graduates over time even in state policy contexts where there are not yet extensive data systems that link teacher and pupil data. Although labor intensive, administration and analysis of the TAPL protocol does not require any additional access to information outside of the institution, such as data from standardized achievement tests, to follow teacher candidates and graduates. Rather TAPL can be implemented within the institution and the resulting data can be used to directly feed back into the preparation program. Over time samples of candidates and then graduates can be assessed using the protocol to examine how teachers learn to teach over time and how the learning opportunities they provide to students change and develop.

Fourth, we want to comment on the fact that the work related to the development and analysis of the TAPL protocol with social justice as a focus was carried out at a private Jesuit university where there is a historical commitment to social justice and where, although state-mandated requirements concerning program approval apply, some of the strictures that circumscribe teacher education at public universities may not pertain. There is no question that the university's commitment to social justice, coupled with its emphasis on personal formation, make it a felicitous space for the development of social

justice as an outcome of teacher preparation. However, this should not be interpreted to mean that social justice cannot or should not be pursued as an appropriate and legitimate outcome of teacher education at public institutions.

When it is suggested that learning to teach for social justice is not an appropriate outcome or goal of teacher preparation at public institutions, the underlying assumption is that this is inappropriate because it has to do with values, beliefs, and ideals, which are assumed not to be the proper purview of teacher education. The concomitant assumption is that teacher education ought to be neutral and apolitical when it comes to values and beliefs. However, the idea behind social justice as a teacher education outcome is that teaching is a profession with certain inalienable purposes, among them challenging the inequities in access and opportunity that curtail the opportunities of some individuals and groups to obtain a high quality education and, at the same time, recognizing and respecting the values and knowledge of marginalized social groups. From this perspective, teaching is a profession that—by definition—has social responsibilities that include challenging the barriers that constrain access to educational opportunities and resources and, at the same time, challenging the cultural hegemony of curriculum, educational policy, and the arrangements and norms of schools. This means that learning to teach for social justice is integral to the very idea of learning to teach, and thus, teaching for social justice is not an outcome only for those prepared at private universities, but a crucial and fundamental outcome of teacher preparation in general.

References

- Adams, M., Bell, L. A., & Griffin, P. (Eds.). (1997). *Teaching for diversity and social justice*. New York: Routledge.
- Ayres, L., Kavanaugh, K., & Kraft, K. A. (2008). Within-case and across-case approaches to data analysis. *Qualitative Health Research, 13*(6), 871-833.
- Ayers, W., Hunt, J. A., & Quinn, T. (Eds.). (1998). *Teaching for social justice*. New York: New Press.
- Biggs, J., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.
- Boston College Lynch School of Education. (2008). *Teacher Education/Special Education, Curriculum & Instruction*. Retrieved August 15, 2008, from <http://www.bc.edu/schools/lsoe/academics/departments/teseci.html>.
- Boyd, D. J., Grossman, P., Lankford, H., Loeb, S., Michelli, N. M., & Wycoff, J. (2006). Complex by design: Investigating pathways into teaching in New York City schools. *Journal of Teacher Education, 57*(2), 155-166.
- Bryk, A. S., Nagaoka, J. K., & Newmann, F. M. (2000). *Chicago classroom demand for authentic intellectual work: Trends from 1997-1999*. Improving Chicago's schools: A data brief of the Chicago Annenberg research project. Chicago: Consortium on Chicago School Research.
- Castenell, L., & Pinar, W.F. (Eds.). (1993). *Understanding curriculum as a racial text*. New York: State University of New York Press.
- Center on Education Policy. (2006). *From the Capital to the classroom: Year 4 of the No Child Left Behind act*. Washington, D.C.: Author.
- Cochran-Smith, M. (1999). Learning to teach for social justice. In G. Griffin (Ed.), *The education of teachers: Ninety-eighth yearbook of the National Society for the Study of Education* (pp. 114-144). Chicago, IL: University of Chicago Press.

- Cochran-Smith, M. (2004). *Walking the road: Race, diversity and social justice in teacher education*. New York: Teachers College Press.
- Cochran-Smith, M. (2006). Teacher education and the need for public intellectuals. *The New Educator*, 2(3), 181-206.
- Cochran-Smith, M. (In press). Toward a theory of teacher education for social justice. In M. Fullan, A. Hargreaves, D. Hopkins, & A. Lieberman (Eds.), *The international handbook of educational change* (2nd ed.). New York: Springer Publishing.
- Cochran-Smith, M., Barnatt, J., Friedman, A., & Pine, G. (In press). Inquiry on inquiry: Practitioner research and students' learning. *Action in Teacher Education*, Fall 2009.
- Cochran-Smith, M., Barnatt, J., Lahann, R., Shakman, K., & Terrell, D. (2009). Teacher education for social justice: Critiquing the critiques. In W. Ayres, T. Quinn, & D. Stovall (Eds.), *The handbook of social justice in education* (pp. 625-639). London: Taylor and Francis.
- Cochran-Smith, M., Davis, D., & Fries, K. (2004). Multicultural teacher education: Research, practice, and policy. In J. Banks (Ed.), *Handbook of research on multicultural education* (3rd ed., pp. 931-975). San Francisco: Jossey Bass.
- Cochran-Smith, M., & Fries, K. (2005). Researching teacher education in changing times: Paradigms and politics. In M. Cochran-Smith & K. Zeichner (Eds.), *Studying teacher education: The report of the AERA panel on research and teacher education* (pp. 69-110). Mahwah, NJ: Lawrence Erlbaum Associates.
- Crowe, E. (2008). Teaching as a profession: A bridge too far? In M. Cochran-Smith, S. Feiman-Nemser, D. J. McIntyre, & K. E. Demers (Eds.), *Handbook of research on teacher education: Enduring questions in changing contexts* (3rd ed., pp. 991-1001). Mahwah, NJ: Lawrence Erlbaum Associates.
- Darling-Hammond, L. (2004). From 'Separate but Equal' to 'No Child Left Behind': The collision of new standards and old inequalities. In D. Meier & G. Wood (Eds.), *Many children left behind: How the No Child Left Behind act is damaging our children and our schools* (pp. 3-32). Boston, MA: Beacon Press.
- Darling-Hammond, L. (2006). No Child Left Behind and high school reform. *Harvard Educational Review*, 76(4), 642-667.
- Darling-Hammond, L., French, J., & García-Lopez, S. P. (2002). *Learning to teach for social justice*. New York: Teachers College Press.
- Decker, P., Mayer, D., & Glazerman, S. (2004). *The effects of Teach for America on students: Findings from a national evaluation*. Princeton, NJ: Mathematica Policy Research, Inc.
- Education Trust. (2008). *Core problems: Out-of-field teaching persists in key academic courses and high-poverty schools*. Washington, DC: Author.
- Eisenhart, M. (2001). Changing conceptions of culture and ethnographic methodology: Recent thematic shifts and their implications for research on teaching. In V. Richardson, (Ed.), *Handbook of research on teaching* (4th ed., pp. 209-225). Washington, DC: American Educational Research Association.
- Erickson, F. (1986). Qualitative methods in research on teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 119-161). New York: Macmillan.
- Girod, G. (2002). *Connecting teaching and learning: A handbook for teacher educators on teacher work sample methodology*. Washington, D.C.: American Association of Colleges for Teacher Education.

- Gore, J. M., Griffiths, T., & Ladwig, J. G. (2004). Towards better teaching: Productive pedagogy as a framework for teacher education. *Teaching and Teacher Education, 20* (4), 375-387.
- Grant, C., & Agosto, V. (2008). Teacher capacity and social justice in teacher education. In M. Cochran-Smith, S. Feiman Nemser, J. McIntyre, & K. Demers (Eds.), *Handbook of research on teacher education: Enduring questions in changing contexts* (3rd ed., pp. 175-200). Philadelphia: Taylor and Francis.
- Grant, C., & Secada, W. (1990). Preparing teachers for diversity. In W. R. Houston, M. Haberman, & J. Sikula (Eds.), *Handbook of research on teacher education* (pp. 403-422). New York, NY: Macmillan.
- Harris, D., & Sass, T. (2007). *Teacher training, teacher quality, and pupil achievement*. National Center for Analysis of Longitudinal Data in Education Research, Working Article #3.
- Hill, C.E., Thompson, B.J., & Williams, E.N. (1997). A guide to conducting consensual qualitative research. *The Counseling Psychologist, 25*(4), 517-572.
- Hollins, E., & Guzman, M.T. (2005). Research on 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.
- Jong, C., Pedulla, P., Mitescu, E., Salomon-Fernandez, Y., & Cochran-Smith, M. (in press). Exploring the link between reformed teaching practices and pupil learning in elementary school mathematics. *School Science and Mathematics Journal*.
- Kantor, H., & Lowe, R. (2006). From New Deal to no deal: No Child Left Behind and the devolution of responsibility for equal opportunity. *Harvard Educational Review, 76* (4), 474-502.
- Kennedy, M. (1999). The problem of evidence in teacher education. In R. Roth (Ed.), *The role of the university in the preparation of teachers* (pp. 87-107). Philadelphia, PA: Falmer Press.
- King, M. B., Schroeder, J., & Chawszczewski, D. (2001). *Authentic assessment and pupil performance in inclusive schools, brief no. 5*. Madison, WI: Research Institute on Secondary Education Reform (RISER) for Youth with Disabilities.
- La Paro, K., Pianta, R., & Stuhlman, M. (2004). The classroom assessment scoring system: Findings from the pre-kindergarten year. *The Elementary School Journal, 104*(5), 409-426.
- Lasley, T., Siedentop, D., & Yinger, R. (2006). A systemic approach to enhancing teacher quality. *Journal of Teacher Education, 57*(1), 13-21.
- MacDonald, H. (1998). Why Johnny's teacher can't teach [Electronic Version]. *City Journal, 8*(2). Retrieved February 1, 2009, from http://www.city-journal.org/html/8_2_a1.html.
- McDonald, M., & Zeichner, K. (2009). Social justice teacher education. In W. Ayers, T. Quinn, & D. Stovall (Eds.), *Handbook of social justice in education* (pp. 595-610). Philadelphia, PA: Taylor and Francis.
- McNergney, R. F. (2006, April 8). *Judging the value of teacher education*. Article presented at the Annual Meeting of the American Educational Research Association, San Francisco.

- McNergney, R. F., & Imig, S. (2006). Challenges of assessing value added to PK-12 education by teacher education [Electronic Version]. *Teachers College Record* Retrieved February 1, 2009, from <http://www.tcrecord.org/Content.asp?ContentID=12814>.
- Meier, D., & Wood, G. (Eds.). (2004). *Many children left behind: How the No Child Left Behind act is damaging our children and our schools*. Boston: Beacon Press.
- Michelli, N. M. (2005). Education for democracy: What can it be? In N. M. Michelli & D. Keiser (Eds.), *Teacher education for democracy and social justice* (pp. 3-30). New York: Routledge.
- Michelli, N. M., & Keiser, D. (Eds.). (2005). *Teacher education for democracy and social justice*. New York: Routledge.
- Newmann, F. M., & Associates. (1996). *Authentic achievement: Restructuring schools for intellectual quality*. San Francisco: Jossey-Bass Publishers.
- Newmann, F. M., Bryk, A. S., & Nagaoka, J. K. (2001). *Authentic intellectual work and standardized tests: Conflict or coexistence?* Chicago, IL: Consortium on Chicago School Research.
- Newmann, F. M., King, M. B., & Carmichael, D. L. (2007). *Authentic instruction and assessment: Common standards for rigor and relevance in teaching academic subjects*. Des Moines, Iowa: Iowa Department of Education.
- Newmann, F. M., Lopez, G., & Bryk, A. S. (1998). *The quality of intellectual work in Chicago schools: A baseline report*. Chicago, IL: Consortium on Chicago School Research.
- Newmann, F. M., Secada, W. G., & Wehlage, G. G. (1995). *A guide to authentic instruction and assessment: Vision, standards, and scoring*. Madison, WI: Wisconsin Center for Education Research, University of Wisconsin (Center on Organization and Restructuring of Schools).
- Noell, G., & Burns, J. (2006). Value-added assessment of teacher preparation: An illustration of emerging technology. *Journal of Teacher Education*, 57(1), 37-50.
- Noell, G., Porter, B., & Patt, R. M. (2007). *Value-added assessment of teacher preparation in Louisiana: 2004-2006, interim technical report*. Louisiana State University.
- North, C. (2006). More than words? Delving into the substantive meaning(s) of 'social justice' in education. *Review of Educational Research*, 76(4), 507-535.
- Oakes, J., & Lipton, M. (1999). *Teaching to change the world*. Boston: McGraw Hill Publishers.
- Orfield, G., Losen, D., & Wald, J. (2004). *Losing our future: How minority youth are being left behind by the graduation rate crisis*. Cambridge, MA: The Civil Rights Project at Harvard University, Contributors: Advocates for Children of New York, The Civil Society Institute.
- Pechone, R., & Chung, R. (2006). Evidence in teacher education: The performance assessment for California teachers (PACT). *Journal of Teacher Education*, 57(1), 22-36.
- Pedulla, J. J., Salomon-Fernandez, Y., Mitescu, E., Jong, C., & Cochran-Smith, M. (2007). *Lessons learned from a study comparing classroom practices and pupil learning for teachers from two pathways into teaching*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

- Piburn, M., Sawada, D., Falconer, K., Turley, J., Benford, R., & Bloom, I. (2000). Reformed Teaching Observation Protocol (RTO). ACEP IN-003. http://PhysicsEd.BuffaloState.Edu/AZTEC/rtop/RTO_full/PDF.
- Queensland School Reform Longitudinal Study. (2001). *School reform longitudinal study*. Brisbane: The State of Queensland (Department of Education).
- Research Institute on Secondary Education Reform (RISER). (2001). *Standards and scoring criteria for assessment tasks and pupil performance*. Madison, WI: Author.
- Rimm-Kaufman, S., La Paro, K., Downer, J., & Pianta, R. (2005). The contribution of classroom setting and quality of instruction to children's behavior in kindergarten classrooms. *The Elementary School Journal*, 105(4), 377-394.
- Schalock, D., Schalock, M., & Ayres, R. (2006). Scaling up research in teacher education: New demands on theory, measurement, and design. *Journal of Teacher Education*, 57(2), 102-119.
- Schalock, D., Schalock, M., & Girod, G. (1997). Teacher work sample methodology as used at Western Oregon State College. In J. Millman (Ed.), *Grading teachers, grading schools* (pp. 15-45). Thousand Oaks, CA: Corwin Press.
- Schalock, M., & Schalock, H. D. (2004). *The longitudinal effects of teacher preparation on the practices and beliefs of new teachers and the learning of their students: Final project report*. Teaching Research Institute, Western Oregon University.
- Sleeter, C. (2001). Preparing teachers for culturally diverse schools: The overwhelming presence of whiteness. *Journal of Teacher Education*, 52(2), 94-106.
- Stake, R. E. (2006). *Multiple case study analysis*. New York, Guilford Press.
- Van de Walle, J. (2007). *Elementary and middle school mathematics: Teaching developmentally* (6th ed.). Boston: Pearson Education, Inc.
- Villegas, A. M., & Lucas, T. (2002). *Educating culturally responsive teachers: A coherent approach*. Albany, NY: SUNY Press.
- Wasley, P., & McDiarmid, G. W. (2004). *Connecting the assessment of new teachers to pupil learning and to teacher preparation: National commission on teaching and America's future national summit on high quality teacher preparation*. Austin, TX: National Commission on Teaching and America's Future.
- Wiedeman, C. (2002). Teacher preparation, social justice, equity: A review of the literature. *Equity and Excellence in Education*, 35(3), 200-211.
- Will, G. (2006, January 16). Ed schools vs. Education. *Newsweek Magazine*, 147(3), 98.
- Zeichner, K. (2003). The adequacies and inadequacies of three current strategies to recruit, prepare, and retain the best teachers for all students. *Teachers College Record*, 105(3), 490-515.
- Zeichner, K. (2006). Reflections of a university-based teacher educator on the future of college- and university-based teacher education. *Journal of Teacher Education*, 57(3), 326-340.