Pre-Service Teachers' Assessment of 7th-Grade Students' Social Studies Learning

Derek Anderson[1,*], Holly Mathys[1] & Andrew Mills[1]

[1]School of Education, Leadership, and Public Service, Northern Michigan University, MI, USA

*Corresponding author: School of Education, Leadership, and Public Service, Northern Michigan University, MI, USA. Tel: 1-906-227-1873. E-mail: dereande@nmu.edu

Received: April 24, 2014 Accepted: May 8, 2014 Online Published: May 17, 2014

doi:10.5430/jct.v3n1p104 URL: http://dx.doi.org/10.5430/jct.v3n1p104

Abstract

The purpose of this study was to examine how 52 pre-service teachers (PSTs) assessed 7th-grade students' learning of social studies lessons they planned and taught. The PSTs provided assessment artifacts on 312 students, along with explanations for how they assessed and graded each student. Of the 429 coded assessment explanations, 240 (56%) related directly to students' social studies achievement, 141 (33%) related to non-achievement factors, and 48 (11%) related to achievement factors not connected with social studies. Consistent with previous research on the assessment practices of in-service teachers, the pre-service teachers in this study used a combination of student achievement, effort, behavior, and ability to make their grading judgments. Assessment of student learning is taking a greater role in education today, and much of the burden for helping new teachers to improve their assessment practices will fall on teacher educators, who should integrate explicit coursework on assessment with authentic field experiences for PSTs to practice planning and teaching lessons, as well as assessing and grading students.

Keywords: *assessment; grading; teacher education; field experience*

1. Introduction

1.1 Assessment of Student Learning

Assessment of students has, perhaps, never been more prominent in the US. The barrage of media attention given to American students' dismal performance on standardized tests co mpared to students across the world has, in part, lead to numerous initiatives to assess US students in more subjects and more often. What's more, the intent of these assessment policies has extended beyond ranking and sorting students. In many states, assessment of student performance is now being used to assess teacher competence.

Assessment is a big concept in education and incorporates a variety of actions and mechanisms for determining what students know and are able to do, with the ultimate purpose of improving teaching and learning. Teachers spend approximately 1/3 of their time on assessment-related activities (Stiggins, 1999; Wise, Lukin, & Roos, 1991), so it is imperative that teachers are competent in assessment. The Council of Chief State School Officers Model Core Standards state that an effective teacher "understands and uses multiple methods of assessment to engage learners in their own growth, to monitor learner progress, and to guide the teacher's and learner's decision making" (p. 9). Teachers need to be effective at assessing students because assessment is a major part of their job, but more importantly because effective assessment leads to increased student learning (American Federation of Teachers, 1990).

1.2 Teacher Accreditation

Teacher accreditation standards have long contained language about the importance of preparing teachers who are proficient at assessment. The Council on the Accreditation of Educator Preparation (2013) Standard 1:

Calls on candidates to understand appropriate uses of a variety of assessments and to be able to construct and employ formative and summative assessments that evaluate P-12 student learning of explicit instructional goals. Completers also should be able to construct and use assessments specifically designed to diagnose learner progress and determine, as appropriate, intervention needs. Completers need to know how to analyze and make use of results from

summative measures such as standardized state or district tests that are administered to their students.

A recent report from the National Council on Teacher Quality (2012) criticized teacher preparation programs for failing to teach assessment adequately. Although notorious for being overly critical of teacher education, the NCTQ echoes a contention expressed by others (Campbell & Evans, 2000; MacLellan, 2004; Mertler, 2004; Volante & Fazio, 2007) who have concluded that new teachers overwhelmingly report they are unprepared for the demands of effective assessment, including how to construct, administer, and score assessments of students' learning.

The National Council on Teacher Quality (2012) review of 180 teacher education programs identified three assessment domains needed by PSTs. In addition to *assessment literacy*, which they described as knowing how to measure student performance, they included *analytical skills* for interpreting performance data, and *instructional decision making*, which involves the use of student performance data to plan instruction. Their critical review of teacher preparation programs placed high emphasis on the analysis and display of quantitative data, which are intentionally left out of many teacher preparation programs that eschew standardized, multiple-choice testing, and especially in humanities courses like the social studies methods course in this study.

1.3 Teacher Preparation

As with most criticisms of teacher education, the disconnect between the attempts by teacher preparation programs to teach assessment and new teachers' reports of their lack of preparation stems from claims that teacher education on assessment is "theory-laden" and "disconnected from teachers' daily assessment practices" (DeLuca & Bellara, 2013, p. 2). The vast majority of published research on assessment focuses on teachers' knowledge and beliefs, not their actual practices. Nonetheless, "there is still relatively little research devoted to understanding the assessment literacy of classroom teachers" (Volante & Fazio, 2007, p. 750). What's more, the majority of published research on teachers' assessment practices has been conducted on in-service teachers (DeLuca & Klinger, 2010). Very little research has been published on pre-service teachers (PSTs) and assessment, and most of those studies have involved self-report and attitudinal data (Brookhart, 2004; DeLuca & Bellara, 2013, Volante & Fazio, 2007). Little is known about how PSTs reason and act related to their assessment of students in authentic contexts (Talanquer, Tomanek, & Novodvorksky, 2013). A great need exists to learn more about the assessment practices of PSTs, particularly how they construct assessments and apply those assessments in practice with actual students.

1.4 Purpose of the Study

The purpose of this study was to examine how 52 PSTs assessed 7th-grade students' learning of social studies lessons they planned and taught. Using PSTs' lesson plans, materials, and post-lesson reflections, as well as student work and classroom observation, we investigated how PSTs assessed learning, assigned grades, and compared student performance with their expectations.

2. Review of Literature

Mounting research on the importance of assessment competence in the role of effective teaching, coupled with the rising accountability movement, has compelled policy makers to mandate teacher preparation programs to teach PSTs to be assessment literate (DeLuca & Bellara, 2013). Despite the requirement from both state and national accrediting bodies that teacher preparation programs prepare PSTs to be competent assessors of student learning, several states, including the one in which this study was conducted, do not require teacher preparation programs to require a specific assessment course. Today, most states require a designated assessment course, a number that has oscillated since Noll (1955) first wrote about the absence of required assessment courses in teacher preparation programs. For example, in a 1972 study, Roeder concluded that only 42% of elementary programs required an assessment course; in 1991, Wise et al. ascertained that 25% of teacher preparation programs required an assessment course; and, in 1999, Stiggins reported that 10 states required an explicit assessment course.

Designated, or explicit, models are one of three different models for teaching assessment during teacher preparation (DeLuca & Bellara, 2013; DeLuca & Klinger, 2010). Integrated models weave assessment objectives into several courses, typically into methods courses. Blended models use a combination of explicit and integrated models. Explicit models are most common, though no rigorous research has compared the three models to determine which is most effective (Daniel & King, 1998; DeLuca & Bellara, 2013).

Most of the published research on the teacher preparation for assessment has examined explicit models (DeLuca et al., 2013). Researchers (ie. Campbell, Murphy, & Holt, 2002; DeLuca & Klinger, 2010; DeLuca et al., 2012; Mertler, 2004; Mertler & Campbell, 2005; Plake & Impara, 1997) often extol the positive impact of explicit assessment

courses; however, their research typically focuses on aspects such as assessment literacy and confidence, not the PSTs' actual assessment of students. Though numerous studies have reported PSTs' confidence in their assessment abilities, this inflated sense of optimism is likely due to their lack of awareness of the difficulties and complexities of actual classroom assessment (DeLuca, Chavez, & Cao, 2013).

Calls for purposeful application of teacher preparation course content into authentic field-based contexts pervades research on how to best prepare tomorrow's teachers (Darling-Hammond, 2006). Teaching about assessment should not an exception. PSTs should have opportunities to work with students in authentic contexts, yet teacher preparation courses on assessment are typically devoid of field experiences that give PSTs opportunities to assessment students (Greenburg & Walsh, 2012).

Teacher educators perpetually seek to bridge theory and practice, and assessment is not immune to such appeals. Shepard, Hammerness, Darling-Hammond, and Rust (2005) urged teacher educators to give PSTs opportunities create, administer, and analyze assessments related to actual student work. Volante and Fazio (2007) concluded after studying PTSs' assessment literacy over several years that the PSTs longed for practical knowledge about assessment, which would likely bolster their self-efficacy. Likewise, Talanquer, Tomanek, and Novodvorksky (2013) recommended teacher educators provide PSTs with increased practice opportunities related to student thinking and performance, as PSTs who participated in a field experience during their assessment course scored significantly higher on assessment knowledge, skills, and attitude than their counterparts who did not have a practicum (Alkharusi, et al., 2011). Having opportunities to apply assessment principles in authentic contexts appears to improve PSTs' assessment knowledge, skills, and attitude (Alkharusi, et al., 2011).

Despite calls for developing teachers' assessment skills, what comprises competence with assessment is not clear. Assessment serves many purposes for a variety of stakeholders. Primarily, assessment is used to inform both student and teacher about each student's achievement and growth. Parents, administrators, and the public at large, are also stakeholders.

Assessment is complex, which certainly contributes to the difficulty of improving PSTs' efficacy (DeLuca, Chavez, & Cao, 2013; DeLuca, Chavez, Bellara, & Cao, 2013). Assessment is largely a product of constant decision-making (McMillan & Nash, 2000), and the use of information gathered from assessment of student learning is essential trait of expert teachers (Berliner, 2004; Kohler, Henning, & Usma-Wilches, 2008).

Despite its complexity, assessment is typically delineated it into two distinct types: formative and summative. Whereas summative assessment is designed to audit or measure student learning, formative assessment serves to improve teaching and learning. Formative assessment guides future teaching, helping teachers to make instructional decisions (Popham, 2008). Traditionally, teachers emphasized assessment of learning, though research has more recently articulated the role of assessment *for* and *as* learning (DeLuca et al., 2013; Deluca & Bellara, 2013; Earl, 2012).

The term, Assessment Literacy, refers to knowing "what they are assessing, why they are doing so, how best to assess the achievement of interest, how to generate sound samples of performance, what can go wrong, and how to prevent those problems before they occur" (Stiggins, 1995, p. 240). Others have referred to assessment literacy as the ability to "design, implement, and discuss assessment strategies" (as cited in Mertler, 2003, p. 10) Teachers with high levels of assessment literacy have knowledge of theory and practice, and they are skilled at using a variety of assessment approaches (DeLuca & Klinger, 2010). Assessment literacy requires knowing and doing.

Measuring assessment literacy has typically been conducted in artificial contexts or by using instruments such as the Assessment Literacy Inventory (ALI), which has been criticized for its poor psychometric qualities (Mertler & Campbell, 2005). Overall, researchers have asserted that teachers' assessment skills are weak (Brookhart, 2001; Mertler & Campbell, 2005; Plake, 1993; Stiggins, 2001), though higher than their PST counterparts (Mertler, 2003)

Following their study of 498 PSTs using a revised ALI, which revealed remarkably low scores, Mertler and Campbell declared, "[T]he role of teaching experience may be too important to overlook" (2005, p. 14). They went on to urge researchers to "disentangle the effects on experience on acquiring assessment skills as applied in educational decision-making" (p. 14).

The descriptor, *hodgepodge*, coined by Brookhart (1994), is frequently cited in the literature related to how teachers assess students' learning. Assessment and grading are complex processes impacted by many factors, internal and external, related to teachers and students (Sun & Cheng, 2013). Nearly all teachers consider both effort and achievement when determining grades. The degree to which teachers weigh achievement factors and non-achievement factors is not clear; however, it is widely agreed that teachers use some combination of student

achievement, effort, behavior, and ability to make grading judgments (McMillan & Nash, 2000; Stiggings, Frisbie, & Griswold, 1989). What is not clear, is the extent to which teachers consider those elements and why.

McMillan and Nash (2000) determined that teachers' philosophies of education had the greatest impact on how they assess and grade students. For example, teachers who believe that student growth is most important tend to reward student effort and growth rather than performance by itself. Tierney et al. (2011) found that teachers overwhelmingly claimed that they did not consider students' behavior and attitude when determining grades, yet they typically factored in non-achievement aspects such as timeliness and effort. They concluded, "[T]eachers were driven in their practices by a sense of what was fair for students, which may have included a host of unexamined assumptions, rather than a sound understanding of grading principles" (p. 224). Bachor and Baer (2001) examined how 127 PSTs made assessment decisions about three hypothetical 5th-grade students. While the majority of the PSTs in their study made assessment decisions based on the students' performance on tasks, 33 (26%) students made decisions that went beyond the information provided, including suppositions about students' home lives and their cognitive capacity. A sociocultural perspective on grading suggests that teachers are not able to separate their assumptions, values, and beliefs about students and learning from their grading practices (McMillan, 2008; Sun & Cheng, 2013).

Numerous studies have pointed to teachers' inability to separate students' non-achievement factors from the quality of their work related to lesson objectives (Guskey, 2011; McMillan, 2008; Randall & Engelhard, 2010: Sun & Cheng, 2013; Zoekler, 2007). Non-achievement factors include attributes such as effort, motivation, cooperation, neatness, creativity, promptness. Consequently, grading practices across teachers are inconsistent. Teachers struggle to weigh non-achievement factors, yet valuing students' effort and dispositions is understandable since those characteristics contribute to students' comprehensive academic success (McMillan, 2008). Teachers regularly take into account student cognitive capacity, motivation, and work habits to adjust their expectations for individual students, which are sometimes recognized explicitly, such as by including non-achievement factors on a rubric, but are most often considered implicitly (Bowers, 2011; Lekholm & Cliffordson, 2008; Stiggins, Frisbie, & Griswold, 1989).

Brookhart (1993) found that teachers give low-achieving students inflated grades for trying hard, even if their performance was below what they were capable of, yet teachers graded average and above-average students based more on their actual achievement. Similarly, in several studies on teachers' grading practices, Randall and Engelhard (2010) noticed that teachers consider students' behavior when assigning grades, especially when students are on the borderline between two grades (ie – B+/A-). Though studied only via a simulation and not authentic work assessing real students, 31 PSTs in a study by Simon, Chitpin, and Yahya (2010) prioritized the use of assessment for classroom management, motivation, and social justice over measurement of student learning. Graham (2005) studied 38 secondary PSTs over three semesters and determined that the PSTs had a difficult time separating student motivation from grading and exhibited anxiety over how to assess students with special needs.

This study differs from nearly all the literature on the assessment literacy of PSTs in that it examines how PSTs assessed students they taught using lessons and assessment they created. Whereas most studies utilize inventories or artifacts from hypothetical students, this study investigated PSTs in authentic contexts.

3. Method

3.1 Participants and Setting

The participants in this study included 52 undergraduate elementary PSTs, from an approximately 9,000-student public university in the upper-Midwest. Consistent with national averages, the PSTs were predominately Caucasian, female (87%), and in their early- to mid-20s. As part of field-intensive undergraduate teacher education program that uses a cohort model to provide PSTs with increasing classroom responsibilities over four semesters, the PSTs in this study were enrolled in a block of four methods courses during the semester prior to student teaching. In one of the field experiences, the PSTs work in groups of three or four to deliver and assess lessons in all subjects to classrooms of 7th-grade students for one week. The cooperating site school is situated in a local rural school district that serves approximately 2,500 K-12 students, 91% of whom are Caucasian, 4% African American, 2% Native American, and 1% Asian. Students represent a wide range of socio-economic levels, including 33% who qualify for free- or reduced-lunch. Essentially, the local 7th-grade teachers turn over their classes to the PSTs for a week, where each PST is the lead teacher for at least one lesson in math, science, and social studies, with English-language arts integrated throughout.

Early each semester, the cooperating teachers identify the specific state content standards they would like the PSTs to teach in each subject during the field experience. Accordingly, each PST designed one or two original lessons and

assessments they taught to the 7th grade students. In this study we examined their social studies lessons specifically. The cooperating teachers require that the PSTs provide them with a grade for each student at the end of the week for a total of 25 points per subject for the week. The PSTs are not directed how to distribute those 25 points; however, nearly all PSTs assigned five points to each day's lesson. For purposes of this research study, as well as for general experiential purposes, we intentionally left the grading instructions ambiguous.

3.2 Data

The data sources for this investigation included the students' work on the assessments, as well as the PSTs' assessments of the students they taught, which included their written explanations and rationale. As part of a culminating reflection assignment, the PSTs were required to include two examples of students' work that were average or at their expectations, two examples that were above average or exceeded their expectations, and two examples that were below average or did not meet their expectations. The PSTs described their analyses of each of the six artifacts of student work, as well as their rationale for how or why they categorized each example accordingly. The 52 PSTs provided 429 pieces of data.

We began to conceptualize their assessment decisions by fracturing the PSTs' explanations into in vivo, or verbatim, codes using the PSTs actual language (Charmaz, 2006). Next, we arrayed the in vivo codes into core categories, saturating the data using constant comparison (Glaser & Straus, 1967) until all properties and dimensions of the PSTs' assessment practices were captured by the core categories (Holton, 2010). Throughout this process we coded collectively discussing and reconciling any discrepancies together (Campbell, Quincy, Osserman, & Pedersen, 2013). Finally, we grouped the core categories into two overarching codes: assessment decisions related to the 7th-grade students' achievement related to the lesson objective, and assessment decisions related to non-achievement factors.

4. Results

4.1 Overall Assessment Factors

The 52 PSTs provided assessment artifacts on 312 7th-grade students. For each student assessment, the PSTs included an explanation for how they assessed and graded the student. Because a number of PSTs included multiple explanations for some students, we coded 429 separate PST comments. Of those 429 comments, 240 (56%) related directly to students' social studies achievement, 141 (33%) related to non-achievement factors, and 48 (11%) related to achievement factors not related to social studies (see Table 1).

Table 1. PSTs' Overall Assessment Factors

Factors:		Total Comments	% of Category	% of Total
S.S. Achievement		240	100	56
	Thought/Depth	74	31	17
	Explanation	53	22	12
	Connections	42	18	10
	Detail	42	18	10
	Objective	29	12	7
Non-S.S. Achievement		48	100	11
Non-Achievement		141	100	33
	Effort	58	41	14
	Completion	46	33	11
	Participation	21	15	5
	Above & Beyond	16	11	4

4.2 Social Studies Achievement

The majority of the PSTs assessed their students' level of success at meeting the social-studies-specific lesson objectives. The PTSs' grading explanations related to students' social studies achievement included comments such as, "This student's reasons explained why it is useful for China to specialize (they mass produce efficiently) and why

it would be bad for China to make everything for itself (it would be its own little island)," as well as explanations like this:

> This student has really internalized the lesson and applied her own experiences to the assignment. She brings up the idea of online and the internet which has captivated the world and helps people communicate between great distances. She starts with that idea of internet and narrows her thinking to families that can also pass religion on from parents to children.

These PSTs skillfully assessed students' performance against their lesson objectives, which they had derived from their assigned social studies standards.

We further coded the 240 social studies-related assessment explanations into the following categories: thoughtfulness and depth of thinking (31%), explanation (22%), connections (18%), detail (10%), and lesson objective (12%) (see Table 2).

Table 2. PSTs' Assessment Factors by Student Performance Level

Factors:	Low	Average	High
S.S. Achievement	60	84	96
Thought/Depth	15	22	37
Explanation	20	18	15
Connections	2	11	29
Detail	17	13	12
Objective	6	20	3
Non-S.S. Achievement	14	16	18
Non-Achievement	61	40	40
Effort	35	15	18
Completion	18	16	12
Participation	8	7	6
Above & Beyond	0	2	12

4.2.1 Thoughtfulness/ Depth of Thinking

Nearly 1/3 of the PSTs' studies-related assessment explanations addressed students' thinking. For example, one PST wrote:

> The student extrapolated using the graph and the article as neither of the two resources had anything about the individuals of China paying higher taxes. She evaluated China's need for alternative [energy] sources then decided that it would mean China would trade more and therefore effect China in a positive way.

The PSTs recognized students' originality of thought ("She relied on her own opinions and thoughts to complete the assignment without solely relying on the discussions from the lesson/activity."), as well as their depth of thought ("They also show some deeper thinking when they mention different lands that might still be taken over."). As eager as the PSTs were to value students' thought, they were equally eager to critique students' lack of thought, as seen in this assessment comment: "There is no elaboration or complete thoughts shown on this assignment, indicating that the students did not apply their thoughts and knowledge to the current event or question presented to them."

How the PSTs interpreted their students' performance was by necessity normative. As part of their post-teaching reflections, they were required to include two examples of low, medium/average, and high student work. Thus, the PSTs frequently referenced comparisons. For example, in determining a student's depth of thought, one PST wrote: "I felt that this student's work should be put in the 'high' category because he was thinking about the questions more in depth than a lot of other students." Likewise, another PST stated, "Her work demonstrated well-thought out answers, and if you compare her answers to the majority of her peers, her work was of better quality."

4.2.2 Explanations

We coded 53 (22%) of the 240 Social Studies Achievement assessment comments under the Explanation sub-category, largely because the PSTs directly noted student's explanations. For example, PSTs wrote comments like: "Both of these students explained their opinions and ideas well" and, "His answers to the written responses are complete with correct explanations." Some of the PSTs expanded on their reasoning with comments like the following: "He answered all the questions with an explanation as to why he thought this way. He explained the article compared basketball to communism, and that he picked the same way to play his team;" and, "I gave Aden 6

out of 6 because I thought she answered the questions correctly. She explained that she couldn't image her life today without the Chinese inventions that she just learned about."

The PSTs also deducted points for students who failed to provide explanation, and included reasoning such as, "the student does not actually explain what having paper all around us means or how it affects us." Unlike the PSTs (described in later in this paper) who assessed students on non-achievment factors and often included conjecture, the PSTs in this category assessed student based only on what they produced on paper. For example, one PST wrote, "He could have included an explanation that would make his answers make sense but since he didn't I was not able to assume his reasoning for each."

4.2.3 Connections

The social studies methods course in which this field experience was situated is built around the National Council for the Social Studies conception of "powerful and purposeful" social studies teaching, which posits: "By building on students' skills and experiences, teachers can design learning events that challenge students to make meaningful connections and expand their knowledge and viewpoints (Berson, Bennett, & Dodson, 2009, ¶ 9). Accordingly, the PSTs are encouraged to design lessons that help student connect standards to their own lives.

In assessing their students, a number of PSTs noted student connections. For example, "This student was able to make the connection between information from their mind and the photos to relate the religions to holidays celebrated." The PSTs noted students' ability to make connections to the "real-world" through examples such as, "She made connections with prior knowledge and the real world that I didn't expect a student to make," and, "Rachel also demonstrated work above my expectations because she was able to not only take the higher-order thinking questions and answer them but apply a real-world example of why her explanation is correct."

Additionally, PSTs cited students' prowess at connecting lesson content to their own lives. For example, one PST explained why she gave this student full points, "This student took the concepts of the article and related them to their life in more than one way." Conversely, this PST explained why she gave a student four out of seven possible points, "This student did not make a connection to their life or the United States. He recalled the information without supporting why this was important or how it related or was important to his life."

The PSTs' comments related to this theme were much more common for high-achieving students (69% of the 42 comments) than for average (26%) or low-achieving students (5%), which indicates that the PSTs viewed student connections as positive and perhaps, unexpected.

4.2.4 Detail

In addition to commenting on students' explanations and depth of thought, 42 assessments specifically addressed students providing detail. For example, one PST explained, "This student answered the question, and she gave detailed reasons why she wouldn't change anything." Another PST commented, "He answered all of the questions completely and gave enough detail to demonstrate she understood the topic." Conversely, the PSTs also used students' use of detail to determine their grade, as seen in comments such as, "Becca received five out of seven points. I feel that she could have added much more detail to her answers."

4.2.5 Lesson Objective

A final subcategory of PST comments that fell under the theme of assessments related to social studies achievement dealt with PSTs' reference to their lessons' objectives. Notably, the PSTs mentioned students' meeting of the lesson objectives much more so for average work (20 out of the 29 comments; 69%) than for student work that was low (21%) or high (10%). For example, PSTs made a number of generic comments like: "This one meet the objective. Gave reason and supported those reasons" and, "This student achieved just what was being asked. He met the objective."

4.3 Achievement Factors Unrelated to Social Studies

In total, 67% of the 429 assessment comments pertained to students' achievement, with 56% relating directly to social studies, as described above, and 11% relating to non-social studies achievement. Non-social studies achievement assessments included comments like: "I appreciate the time the student spent to make sure that I could read his handwriting. This student used complete sentences and proper grammar/syntax." The 48 comments in this category all highlighted students' written language skills. For example, one PST explained, "She used good sentence structure," while another PST wrote, "He has excellent writing skills."

The PSTs' assessment of non-social students achievement were nearly equally spread across artifacts at the low, average, and high levels. The PSTs were as likely to reward students for their written language skills as they were to

deduct points for students' inadequacies with written language. This PST justified giving a student a low grade on her social studies assignment as follows: "Some of the student's spelling is wrong, or inaccurate, although they used vocabulary directing from their reading article." Other PST assessments included explanations like, "He did not use sentence structure or grammar," and, "Their work is very sloppy and hard to read."

4.4 Achievement Factors Unrelated to Social Studies

Of the 429 assessment comments, 141 (33%) related to non-achievement factors. In other words, the PSTs graded students' effort, behavior, and ability rather than their achievement. For example, one PST explained why she gave a student full points: "I gave her 7/7 because she took appropriate notes, followed directions, and put good effort into her responses. Throughout the entire week I noticed how well this student does." Another PST gave a student full points on an assignment because, "She asked a lot of questions and showed interest while doing the stations."

Whereas with factors related to social studies achievement where the PSTs were more likely to note social studies achievement for high-quality student work rather than low-quality work (96 vs 60 comments), for non-achievement factors, the PSTs were more likely to address low-quality work with effort or behavior comments than they were for high-quality student work (61 vs 40 comments). Nonetheless, one-third of the PSTs' assessment decisions were based on factors other than students' actual learning of the social studies objectives.

We further coded these 141 non-achievement assessment explanations into the following subcategories: effort (41%), completeness (33%), participation (15%) and above & beyond (11%). Below we explain and provide examples for each of these subcategories.

4.4.1 Effort

The PSTs made 58 assessment comments about their students' effort, though they were much more likely to use effort to justify taking away points (60% of comments) than to reward students based on their effort (14%). Comments like the following illustrate how the PSTs rewarded student effort: "This student put in a lot of effort in their notes" and, "I could tell that she didn't just answer my questions to answer them; she really put effort into it." Conversely, PSTs deducted points when they thought students lacked effort, as seen in assessment comments like the following: "This student got a 1 because their answers show that no thought was really put into the answers and they answered something just to complete the assignment" and, "I felt the student did not give the effort that was required of them. Some areas were indicated fully, but in a rushed manner just to complete the assignment."

In these cases, including those where the PSTs allocated full points and those in which the PSTs subtracted points, the grades were derived from student effort rather than their performance or achievement. The PSTs were straightforward about punishing students for lack of effort. For example, one PST wrote, "I wanted to give him 7/7 but looking at some of his previous work I knew he could have done better." Another PST explained, "These two students were below my expectations because I can tell that they didn't put much time or effort into their work, so they didn't receive full points." It is important to note that the PSTs had only worked with these students for one week, yet they felt compelled to note the discrepancy between students' ability and their achievement when assigning points for daily homework and coursework.

4.4.2 Completeness

One-third of the non-achievement comments pertained to whether students completed the assigned task, regardless of how well the task was completed or how much effort they perceived the students gave. For example, one PST wrote in her assessment justification, "These two students met my expectations and received a 7 out of 7 because they completed all of the parts on the rubric, they filled out the table completely, and they were working the entire time." Another PST made reference to the rubric criteria was looking for and how the student included all components: "This student had a complete climate graph with all required aspects. Their graphs included the following: title, key, labels, color, precipitation amounts, and temperature amounts. The graphs were also done in a nice, neat fashion." Conversely, the PST noted how she gave a low score to another student because her "graph is incomplete; no title, labels, key, precipitation amounts, color and the temperature amount aren't finished."

In these cases, PSTs were looking for compliance and amenability rather than student understanding. This phenomenon is likely related to the PSTs' worry over managing the classroom, particularly since many of the PSTs preferred to teach early elementary school and were nervous about teaching middle-level students. The PST who wrote, "This student was on task during the entire lesson. She completed the reading strategy and the graphic organizer," had made numerous comments prior to the field experience about how worried she was that the students wouldn't listen to her. After the experience, she remarked about how much better it went than she expected. As explained further in the next subsection, the PSTs were pleased when students engaged with their lessons and

rewarded the students accordingly.

4.4.3 Participation

Twenty-one assessment explanations connected students' assignment grade to their participation in lesson activities without reference to their achievement. In some cases, the PSTs simply awarded points for participation as seen in comments like, "He got points for participation," and, "I graded the students on participation, so he received full credit." Conversely, students who did not participate were docked points. One PST explained why she gave 2 out of 5 possible points to a student by stating simply, "He was not participating throughout this activity. In these instances, the students earned points based not on what they learned but based on their behaviors throughout the lesson. As noted in assessment comments like, "I really liked her enthusiasm and participation during the stations activity" several PSTs revealed their bias toward eager, compliant students.

4.4.4 Above & Beyond

The final subcategory of assessment explanations under the theme of non-achievement factors addresses PSTs' ambiguous justification for students who "went above and beyond." In some cases, the students' work went beyond what the PSTs were expecting. For example, on PST wrote, 'I assigned a grade of 6 out of 6 to Rachel because I thought her answers went beyond what I was looking for." Another PST explained, "This group of students exceeded my expectations, as they produced really phenomenal thoughts for their graffiti wall . . . these were things beyond what I expected from the age group. I was very impressed." In some instances, the PSTs cited a quantitative exceeding of what was asked: "The directions were to write 3 questions for the Skype interview with Sarah's parents. She wrote 10." Another PST wrote this about a student who earned full points: "This student exceeded my expectations. I told the students that they needed to use two of our word wall words from the lesson (import, export, and resource) but this student used three."

5. Discussion

The term *hodgepodge* (Brookhart, 1994) could not be more apt to describe how the 52 PSTs assessed 312 student assignments and provided 421 explanations for how they assessed and graded those students. Only slightly more than half (56%) of their assessment comments related to the students' performance on social studies content standards the cooperating classroom teachers asked them to teach. While 11% of their explanations related to students' written language performance, 33% of the PSTs' assessment decisions were based on students' effort, participation, and compliance.

Assessment lies at the intersection of instruction, classroom management, and evaluation (Brookhart, 2004). The relationship between assessment and grading is multifaceted. Both assessment and grading incorporate students' achievement and behavior (McMillan, Myran, & Workman, 2002). Grades communicate student achievement, ranking, and progress to students and parents. Grading can be individual-referenced (comparing a student's performance to her previous achievement), norm-referenced (comparing a student's performance to her peers'), or criterion-referenced or standards-based (comparing a student's performance to an objective) (Dalbert, Schneidewind, & Saalbach, 2007). Grading practices are highly variable, more so within schools than across schools. Furthermore, the extent to which teachers use non-achievement behaviors varies greatly (McMillan et al., 2002; Talanquer, Tomanek, & Novodvorksky, 2013).

NCLB (2002) and RTT (2011) have placed more emphasis on using a variety of classroom assessment practices, as well as on transparency (DeLuca & Klinger, 2010). Gone are the days when teachers taught and assessed what and how they wanted. Today, teachers face much more scrutiny from colleagues and stakeholders alike. If, as Brookhart maintains, the hodgepodge approach to assessment persists, teachers are going to be ill-equipped to handle the increased expectations for transparency regarding their assessment actions.

Advocates of standards-based grading urge teachers to decouple student effort and attitude from their mastery of course objectives, which they claim should be the only factors in determining student grades (McMillan, 2008; O'Connor, 2007). Standards-based grading is the most reliable method and reported by students to be the most just (Dalbert et al., 2007).

Conversely, differentiation advocates suggest teachers alter their expectations for student achievement based on individual student potential, teaching in each student's zone of proximal development (Vygotsky, 1978). Individual-referenced or differentiated assessment best fosters students' motivation and effort (Dalbert et al., 2007).

Teachers, especially at the elementary levels, tend to be highly concerned with students' self-concept and

consequently consider students' attitude and effort in addition to achievement in nearly all cases. Teachers often "pull for students" by giving them extra chances and softer grading practices (McMillan & Nash, 2000). Yet, when teachers reward students' effort and attitude, they deprive students of receiving accurate feedback about their performance. (McMillan et al., 2002). Grades should inform stakeholders of what each student knows and is able to do. In their study of more than 900 elementary teachers, McMillan et al. (2002) found that teachers who gave more A's used fewer objective assessments, emphasizing non-achievement factors.

Ultimately, teachers seek to make fair assessments, yet their perspectives on fairness vary widely (Tierney, Simon, & Charland, 2011; Yung, 2001). Randall and Engelhard (2010) urged teacher educators to spend more time preparing teachers beyond how to assess students, and to include discussion and practice related to assigning course grades. Graham (2005) offered a warning to teachers that if they are unable to reliably tie assessment practices to student learning, policymakers will turn to more objective measures.

On a positive note, the PSTs in this study frequently assessed students using questions and tasks that required more than merely recall. Several of the PSTs posed compelling questions, which are a key component of the new C3 Framework. Grant (2013) asserted that teachers should ask students questions that are "intellectually meaty", meaning that the task involves "an enduring issue, concern, or debate in social studies and it has to draw on multiple disciplines" (p. 326). Furthermore, Grant described how teachers should use the C3 to make student-friendly questions, or questions that matter to their lives. A central component of the PSTs' elementary social studies methods course was designing lessons that reflection what the National Council of Social Studies calls, *Powerful and Purposeful Teaching and Learning in Elementary Social Studies* (http://www.socialstudies.org/positions/powerfulandpurposeful). Accordingly, the PSTs are encouraged to teach lessons that are meaningful, integrative, value-based, challenging, and active.

6. Recommendations

We must recognize that today's teachers are trapped between the accelerating high-stakes testing culture, and their personal and teacher-educator-recommended orientations to assessment that value differentiation and student-centeredness (Graham, 2005).

Field experiences are vital to effective teacher development, and mentor teachers are at the center of those experiences (Anderson, 2007). PSTs frequently cite their cooperating teacher as having had the greatest influence on their learning about assessment (Graham, 2005; Kusch, 1999); and, DeLuca and Klinger (2010) concluded that PSTs who did not take an assessment course learned most about assessment from student teaching practicums, yet they warned that the PSTs' learning about assessment is "ultimately dependent on the knowledge and skills of the cooperating teachers" (p. 434).

While the field experience in this paper gave PSTs authentic opportunities to design, administer, and analyze assessments, they were not in any substantive way mentored in the process by the cooperating teacher. Furthermore, it is likely that cooperating teachers themselves lack the necessary assessment literacy to mentor PSTs (DeLuca & Klinger, 2010; Earl, Freeman, Lasky, Sutherland, & Torrance, 2002). Thus, teacher preparation programs that depend on PSTs learning about assessment from practicum are not likely to produce assessment literate teachers (DeLuca & Klinger, 2010). Teacher preparation programs in states without mandates for specific assessment courses need to consider requiring assessment courses anyway.

Much of the burden for helping new teachers to improve their assessment practices will fall on teacher educators, who must both orchestrate authentic field experiences for PSTs to practice authentically, and they must model to the PSTs the types of assessment they expect classroom teachers to use (Graham, 2005).

References

Alkharusi, H., Kazem, A. M., & Al-Musawai, A. (2011). Knowledge, skills, and attitudes of preservice and inservice teachers in educational measurement. *Asia-Pacific Journal of Teacher Education*, *39*(2), 113-123. http://dx.doi.org/10.1080/1359866X.2011.560649

American Federation of Teachers, National Council on Measurement in Education, and National Education Association. (1990). Standards for teacher competence in educational assessment of students. *Educational Measurement: Issues and Practice*, *9*(4), 30-32. http://dx.doi.org/10.1111/j.1745-3992.1990.tb00391.x

Anderson, D. (2007). Cooperating teachers' power in the student teaching practicum. *Education*, *128*(2), 307-323.

Bachor, D., & Baer, M. (2001). An examination of preservice teachers' simulated classroom assessment practices. *The Alberta Journal of Educational Research, 47*(3), 244-258.

Berliner, D. C. (2004). Describing the behavior and documenting the accomplishments of expert teachers. *Bulletin of Science, Technology & Society*, *24*(3), 200-212. http://dx.doi.org/10.1177/0270467604265535

Berson, I., Bennett, L., & Dodson, D. (2009). Powerful and purposeful teaching and learning in elementary school social studies. *Social Studies and the Young Learner*, *22*(1), 31-33.

Bowers, A. J. (2011). What's in a grade? The multidimensional nature of what teacher-assigned grades assess in high school. *Educational Research and Evaluation*, *17*(3), 141-159. http://dx.doi.org/10.1080/13803611.2011.597112

Brookhart, S. (2004). Classroom assessment: Tensions and intersections in theory and practice. *The Teachers College Record*, *106*(3), 429-458. http://dx.doi.org/10.1111/j.1467-9620.2004.00346.x

Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement*, *30*(2), 123-142. http://dx.doi.org/10.1111/j.1745-3984.1993.tb01070.x

Brookhart, S. M. (1994). Teachers' grading: Practice and theory. *Applied Measurement in Education*, *7*, 279-301. http://dx.doi.org/10.1207/s15324818ame0704_2

Brookhart, S. M. (2001). Successful students' formative and summative uses of assessment information. *Assessment in Education: Principles, Policy & Practice*, *8*(2), 153-169. http://dx.doi.org/10.1080/09695940123775

Campbell, C., & Evans, J. A. (2000). Investigation of preservice teachers' classroom assessment practices during student teaching. *The Journal of Educational Research*, *93*(6), 350-355. http://dx.doi.org/10.1080/00220670009598729

Campbell, C., Murphy, J. A., & Holt, J. K. (2002, October). Psychometric analysis of an assessment literacy instrument: Applicability to preservice teachers. In *annual meeting of the Mid-Western Educational Research Association, Columbus, OH*.

Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding In-depth Semistructured Interviews Problems of Unitization and Intercoder Reliability and Agreement. *Sociological Methods & Research*, *42*(3), 294-320. http://dx.doi.org/10.1177/0049124113500475

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. London: Sage.

Council for the Accreditation of Educator Preparation. (2013). CAEP 2013 Standards for Accreditation for Educator Preparation. Retrieved from http://caepnet.org/accreditation/standards/

Council of State School Officers (2011). InTASC Model Core Teaching Standards: A Resource for State Dialogue. Retrieved from http://www.ccsso.org/Documents/2011/InTASC_Model_Core_Teaching_Standards_2011.pdf

Dalbert, C., Schneidewind, U., & Saalbach, A. (2007). Justice judgments concerning grading in school. *Contemporary Educational Psychology*, *32*(3), 420-433. http://dx.doi.org/10.1016/j.cedpsych.2006.05.003

Daniel, L. G., & King, D. A. (1998). Knowledge and use of testing and measurement literacy of elementary and secondary teachers. *The Journal of Educational Research*, *91*(6), 331-344. http://dx.doi.org/10.1080/00220679809597563

Darling-Hammond, L. (2006). Constructing 21st-century teacher education. *Journal of teacher education*, *57*(3), 300-314. http://dx.doi.org/10.1177/0022487105285962

DeLuca, C., & Bellara, A. (2013). The Current State of Assessment Education Aligning Policy, Standards, and Teacher Education Curriculum. *Journal of Teacher Education*, *64*(4), 356-372. http://dx.doi.org/10.1177/0022487113488144

DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, *17*(4), 419-438. ttp://dx.doi.org/10.1080/0969594X.2010.516643

DeLuca, C., Chavez, T., & Cao, C. (2013). Establishing a foundation for valid teacher judgment on student learning: the role of pre-service assessment education. *Assessment in Education: Principles, Policy & Practice*, *20*(1), 107-126. http://dx.doi.org/10.1080/0969594X.2012.668870

DeLuca, C., Chavez, T., Bellara, A., & Cao, C. (2013). Pedagogies for Preservice Assessment Education: Supporting Teacher Candidates' Assessment Literacy Development. *The Teacher Educator*, *48*(2), 128-142.

http://dx.doi.org/10.1080/08878730.2012.760024

DeLuca, C., Luu, K., Sun, Y., & Klinger, D. A. (2012). Assessment for learning in the classroom: Barriers to implementation and possibilities for teacher professional learning. *Assessment Matters*, *4*, 5-29.

Earl, L. M. (2012). *Assessment as learning: Using classroom assessment to maximize student learning*. Corwin.

Earl, L., Freeman, S., Lasky, S., Sutherland, S., & Torrance, N. (2002). *Policy, politics, pedagogy and people: Early perceptions and challenges of large-scale reform in Ontario secondary schools.* Toronto: Report commissioned by The Ontario Secondary Teachers' Federation.

Glaser, B. G., & Straus, A. L. (1967). *The discovery of grounded theory.* Chicago: Aldine.

Graham, P. (2005). Classroom-based assessment: Changing knowledge and practice through preservice teacher education. *Teaching and Teacher Education, 21*(6), 607-621. http://dx.doi.org/10.1016/j.tate.2005.05.001

Grant, S. G. (2013). From inquiry arc to instructional practice: The potential of the C3 framework. *Social Education, 77*(6), 322–326.

Greenberg, J., & Walsh, K. (2012). *What teacher preparation programs teach about K-12 assessment: A review.* Washington, DC: National Council on Teacher Quality. Retrieved from http://www.nctq.org/edschoolreports/assessment/report.jsp

Guskey, T. (2011). Five obstacles to grading reform. *Educational Leadership, 69*(3), 16-21.

Holton, J. A. (2010). The coding process and its challenges. *Grounded Theory Review*, *9*(1), 21-40.

Kohler, F., Henning, J. E., & Usma-Wilches, J. (2008). Preparing preservice teachers to make instructional decisions: An examination of data from the teacher work sample. *Teaching and Teacher Education, 24*(8), 2108-2117. http://dx.doi.org/10.1016/j.tate.2008.04.002

Kusch, J. W. (1999). The Dimensions of Classroom Assessment: How Field Study Students Learn to Grade in the Middle Level Classroom. *Journal of Educational Thought/Revue de la Pensee Educative*, *33*(1), 61-81.

Lekholm, A. K., & Cliffordson, C. (2008). Discrepancies between school grades and test scores at individual and school level: effects of gender and family background. *Educational Research and Evaluation*, *14*(2), 181-199. http://dx.doi.org/10.1080/13803610801956663

Maclellan, E. (2004). Initial knowledge states about assessment: Novice teachers' conceptualisations. *Teaching and Teacher Education, 20*(5), 523-535. http://dx.doi.org/10.1016/j.tate.2004.04.008

McMillan, J. H. (Ed.). (2008). *Assessment essentials for standards-based education.* Corwin Press.

McMillan, J. H., & Nash, S. (2000). *Teacher Classroom Assessment and Grading Practices Decision Making*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans.

McMillan, J. H., Myran, S., & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *The Journal of Educational Research, 95*(4), 203-213. http://dx.doi.org/10.1080/00220670209596593

Mertler, C. A. (2003, October). *Preservice versus inservice teachers' assessment literacy: Does classroom experience make a difference?* Paper presented at the annual meeting of the Mid-Western Educational Research Association, Columbus, OH.

Mertler, C. A. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American secondary education*, *33*(1), 49-64.

Mertler, C. A., & Campbell, C. (2005 April). *Measuring teachers' knowledge and application of classroom assessment concepts: Development of the "Assessment Literacy Inventory."* Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, QC.

National Council on Teacher Quality (2012). What Teacher Preparation Programs Teach about K-12 Assessment: A Review. Retrieved from http://www.nctq.org/dmsView/What_Teacher_Prep_Programs_Teach_K-12_Assessment_NCTQ_Report

Noll, V. H. (1955). Requirements in educational measurement for prospective teachers. *School and Society*, *80*, 88-90.

O'Connor, K. (2007). *A repair kit for grading: 15 fixes for broken grades.* Portland, OR: Educational Testing Service.

Plake, B. S. (1993). Teacher Assessment Literacy: Teachers' Competencies in the Educational Assessment of

Students. *Mid-Western Educational Researcher, 6*(1), 21-27.

Plake, B. S., & Impara, J. C. (1997). Teacher assessment literacy: What do teachers know about assessment? In G. D. Phye (Ed.), *Handbook of classroom assessment* (pp. 55-68). New York: Academic Press. Stiggins, R. J., & Bridgeford, R. J. (1985). The ecology of classroom

Popham, W. J. (2008). *Transformative assessment in action: An inside look at applying the process*. Alexandria, VA: ASCD.

Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education, 26*(7), 1372-1380. http://dx.doi.org/10.1016/j.tate.2010.03.008

Roeder, H. H. (1972). Are today's teachers prepared to use tests?. *Peabody Journal of Education, 49*(3), 239-240. http://dx.doi.org/10.1080/01619567209537858

Shepard, L., Hammerness, K., Darling-Hammond, L., & Rust, F. (2005). Assessment. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 275-326). San Francisco: Jossey-Bass.

Simon, M., Chitpin, S., & Yahya, R. (2010). Pre-service teachers' thinking about student assessment issues. *International Journal of Education, 2*(2), 1-19.

Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan, 77*, 238-238.

Stiggins, R. J. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice, 18*(1), 23-27. http://dx.doi.org/10.1111/j.1745-3992.1999.tb00004.x

Stiggins, R. J. (2001). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice, 20*(3), 5-15. http://dx.doi.org/10.1111/j.1745-3992.2001.tb00065.x

Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. (1989). Inside high school grading practices: Building a research agenda. *Educational Measurement: Issues and Practice, 8*(2), 5-14. http://dx.doi.org/10.1111/j.1745-3992.1989.tb00315.x

Sun, Y., & Cheng, L. (2013). Teachers' grading practices: meaning and values assigned. *Assessment in Education: Principles, Policy & Practice*, (ahead-of-print), 1-18. http://dx.doi.org/10.1080/0969594X.2013.768207

Talanquer, V., Tomanek, D., & Novodvorsky, I. (2012). Assessing students' understanding of inquiry: What do prospective science teachers notice? *Journal of Research in Science Teaching, 50*(2), 189-208. http://dx.doi.org/10.1002/tea.21074

Tierney, R. D., Simon, M., & Charland, J. (2011, July). Being fair: Teachers' interpretations of principles for standards-based grading. In *The Educational Forum* (Vol. 75, No. 3, pp. 210-227). Taylor & Francis Group. http://dx.doi.org/10.1080/00131725.2011.577669

Volante, L., & Fazio, X. (2007). Exploring Teacher Candidates' Assessment Literacy: Implications for Teacher Education Reform and Professional Development. *Canadian Journal of Education, 30*(3), 749-770. http://dx.doi.org/10.2307/20466661

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. In M. Cole, V. John Steiner, S. Scribner, & E. Souberman (Eds.). Cambridge, MA: Harvard University Press.

Wise, S. L., Lukin, L. E., & Roos, L. L. (1991). Teacher beliefs about training in testing and measurement. *Journal of Teacher Education, 42*(1), 37-42. http://dx.doi.org/10.1177/002248719104200106

Yung, B. H. W., & Yung, B. H. W. (2001). Three views of fairness in a school-based assessment scheme of practical work in biology. *International Journal of Science Education, 23*(10), 985-1005. http://dx.doi.org/10.1080/09500690010017129

Zoeckler, L. G. (2007). Moral aspects of grading: A study of high school English teachers' perceptions. *American Secondary Education, 35*(2), 83-102.