

Classroom Assessment, Student Motivation, and Achievement in High School Social Studies Classes

Susan M. Brookhart
School of Education
Duquesne University

Daniel T. Durkin
School of Education
Duquesne University

The purpose of this case study was to describe a variety of classroom assessment events in high school social studies classes. This study included data from 12 classroom assessment events in the classes of a teacher–researcher in an urban high school. Four assessments in each course of the teacher-researcher’s entire teaching load were studied. The courses were world cultures, honors U.S. history, and philosophy. The total number of students was 96; sample sizes for individual assessment analyses ranged from 11 to 39. Results supported the conclusion that even within the same classroom assessment environment, student perceptions of the assigned task and self-efficacy for the task, reported mental effort invested, goal orientations, and learning strategy use differed by assessment. The mean level of these variables differed by type of student. Observed correlations among these variables differed between paper-and-pencil tests and performance assessments. Potential implications for classroom assessment practices are discussed.

When researchers want to understand what students know, classroom assessment is not the tip but the bulk of the iceberg. Large-scale assessment is more carefully studied, better funded, and higher profile than is classroom assessment—but the lion’s share of assessment that students experience is classroom assessment. It is from frequent and regular classroom assessment and teacher feedback that

students get a sense of what they know and do not know, what they might do about this, and how they feel about it. Understanding the dynamics of classroom assessment is essential for improving education or even for understanding the current state of learning in classrooms.

Each time a particular assessment task is assigned, students experience expectations and the degree to which they meet these expectations; these segments can be called *classroom assessment events* (Brookhart, 1997b). A classroom assessment event may be considered a special case of an activity segment—a lesson or part of a lesson that has a focus, a beginning and an end, participants, materials, and goals (Stodolsky, 1988)—in which students realize that the degree to which their performance achieves the goal will be judged by the teacher. Many classroom assessment events occur in typical classrooms, all intertwined. The overall sense of expectations that these events build up, the meaning or sense that students make out of this aspect of their classroom, composes the *classroom assessment environment* (Stiggins & Conklin, 1992).

The classroom assessment environment, as Stiggins and Conklin (1992) originally described it, was more about teacher practices than about student perceptions. Important to creating a classroom assessment environment were the purposes for which teachers used classroom assessments; the assessment methods used, the criteria for selecting them, and their quality; the teacher's use of feedback; the teacher's preparation and background in assessment; the teacher's perceptions of students; and the assessment policy environment. All these except the last are under the teacher's control. Thus, a teacher's classes had an assessment "character" or environment that stemmed from the teacher's general approach to assessment. Haydel, Oescher, and Kirby (1999) called teachers' beliefs about testing and assessment practices "the evaluative culture of classrooms" (p. 1) and found that these beliefs were related to teacher efficacy. In turn, teacher efficacy is related to student efficacy and achievement.

Instruction and assessment come to the students in the form of activity segments (Stodolsky, 1988). Student behaviors at least partly depend on the kind of activity with which they are presented. Activity segments in general, including classroom assessment events in particular, occur at an interesting intersection of group instruction and individual psychology. Classroom assessments are typically assigned, presented, or administered to classes or groups of students. However, individual differences among students operate as each student perceives and responds to the assignment. This phenomenon of individual differences operating within group and classroom-environmental space is what makes classroom assessments at once so interesting and yet tricky to study (Ames, 1992). Nevertheless, design difficulties should not be an excuse for ignoring classroom assessment as a study topic. There are some theoretical "places to stand," as Archimedes would say, from which to begin inquiry.

FRAMEWORK FOR UNDERSTANDING CLASSROOM ASSESSMENT EVENTS

Variables that are important to student motivation and learning are identified in the cognitive psychology literature. These variables may operate somewhat differently for each classroom assessment event because both the group assignment and individual students' responses to it may differ, even in the same classroom assessment environment, from one assessment to another. An important consideration is that any classroom assessment event occurs within a particular classroom assessment environment (Brookhart, 1997b; Haydel et al., 1999). Thus, for example, an end-of-unit test presented after appropriate instruction in a supportive environment may be perceived differently and have different student responses than the same test presented after ambiguous lessons in a judgmental environment.

Within the classroom assessment environment, each assessment task is typically assigned to a group of students but perceived by each student individually. Each student's perception of the importance and value of the assigned task and each student's perception of his or her ability to accomplish this particular task affect effort (Salomon, 1983, 1984) and ultimately achievement (Gipps, 1994). The perceived importance, usefulness, and value of engaging in a task are motivators for student effort (Pintrich & Schrauben, 1992). For example, high school students for whom tests have serious consequences are more likely to perform well on test items that are mentally taxing than are students for whom test consequences are minimal (Wolf, Smith, & Birnbaum, 1995).

The importance of interest to motivation has begun to receive attention. Brophy (1999) pointed out that more is known about the expectancy aspects of motivation, for example, self-efficacy theory and attribution theory, than about the value aspects. Valuing a task is critical not only for having continuing interest in a topic, but also for voluntarily electing to do a task (e.g., not just being able to read but choosing to read a book). Hidi and Harackiewicz (2000) found some evidence that student interest can be focused with situational factors (appealing material, "interesting" tasks); once interest is internalized, the student will continue to pursue mastery.

Bergin (1999) found that factors that influence student interest are of two types: individual and situational. Individual factors influencing interest include belongingness, emotions, competence, utility, and background knowledge. Situational factors influencing interest include hands-on use, discrepancy, novelty, food, social interaction, modeling, games and puzzles, content, biophilia, fantasy, humor, and narrative. Bergin recommended that teachers attend especially to situational factors because they are under teachers' control. Three of his individual factors relate to variables under study in this article as well. In this study, utility, or "goal relevance" as Bergin described it, is studied as the perceived value of the task. Feeling competent and having requisite background knowledge are part of perceived self-efficacy, also a variable in this study.

Self-efficacy to do a task refers to students' expectations that they can actually accomplish it. Therefore, self-efficacy is logically investigated at the level of individual tasks, or individual classroom assessment events. Empirically, measures of self-efficacy for specific tasks or kinds of tasks are more predictive of accomplishment than are generalized self-efficacy measures (Pajares, 1996).

Students differ in their perceived self-efficacy to accomplish a task as they perceive it (Pintrich & Schrauben, 1992; Schunk, 1994; Weiner, 1979). Self-efficacy is associated with effort, persistence, and performance. Students make judgments about their own self-efficacy by comparing their past accomplishments with standards, either relative or absolute (Lepper, 1988; Schunk, 1994). To make these judgments, students must weigh task characteristics such as difficulty, amount of effort required, and amount of assistance available against their perceptions of their past performances and accomplishments. Then, they expend the effort to do the assessment.

In a study of two third-grade language arts classes, in which four classroom assessment events were sampled in each, Brookhart and DeVoge (1999) investigated perceived task characteristics, perceived self-efficacy, and amount of invested mental effort for various classroom assessment events. The results of the study affirmed the general outline of the preceding theoretical framework but suggested that additional variables should be considered. Specifically, Brookhart and DeVoge found that the goal orientations that students bring to the assessment should be considered and that students refer directly to individual prior experiences with similar assessments when they are sizing up their work on a particular assessment. Third graders invoked past performance literally, saying things such as "We did two sentences like that already and I went over it."

The value of the addition of goal orientations to the theoretical framework is also supported by current literature. Goal orientations characterize students' approaches to learning. Students with a mastery goal orientation place value on the learning itself, whereas students with a performance goal orientation place value on others' approval of their performance (Ames & Archer, 1988). Theorists have divided performance goal orientations into two subsets: performance-approach, wanting to be perceived as smart, wanting to outperform classmates, and so on; and performance-avoid, wanting to avoid failure or being seen as incompetent (Elliot & Covington, 2001). As did Meece and Miller (2001), this study measured performance goals that focused on students' positive desires (i.e., performance-approach goals).

The relationship between mastery and performance goals is complicated. It is not simply a continuum along which students either want to learn for its own sake or need outside attention to motivate learning. It is not even as simple as a three-fold set of categories (mastery, performance-approach, performance-avoid) instead of a twofold set. Elliot and Thrash (2001) described a hierarchical model of achievement motivation, differentiating between how competence is defined and how it is valenced. By *valence*, they mean the distinction between approach and avoidance already noted. More interesting for this study is the insight that both

mastery goals and performance goals are ways of seeking competence, depending on how the student defines *competence*, whether according to an absolute standard (seeking mastery), an intrapersonal standard (seeking personal improvement), or a normative standard (seeking to outperform others). In measurement terms, this distinction is made according to whether the student's judgments are criterion referenced, self-referenced, or norm referenced, although this is not the language Elliot and Thrash used. If both mastery goals and performance goals are forms of seeking competence, it makes sense that they could be related and that they could be more related for some kinds of achievement tasks or classroom assessments than for others.

Hidi and Harackiewicz (2000) pointed out that there has been a tendency to champion individual, intrinsic interest and motivation, and mastery learning goals, and to discourage situational interest, extrinsic motivation, and ego or performance goals. Their review concluded that these two constellations did not constitute opposite motivational states. They also found some studies that reported that performance and mastery goals can coexist and that students with both orientations had higher levels of self-regulation and higher grades than those of students with one or neither goal orientation.

Students who believe that they are capable of performing a task and who want to learn the concepts or skill involved because they desire mastery do not simply expend effort or try hard, but they try "in a planful, strategic fashion" (Covington, 1992, p. 203). In this study, separate variables were used to measure (a) the amount of invested mental effort (Salomon, 1984) or "trying" and (b) the use of active or superficial learning strategies (Meece, Blumenfeld, & Hoyle, 1988). In a study of interrelationships among selected conative (motivational and volitional) constructs relevant to learning and performance, Jackson (1994) reported a high correlation between "mindfulness," defined as the amount of invested mental effort as measured with Salomon's items, and a deep approach to learning. Meece and Miller (2001) found that changes in task-mastery goal orientations explained variation in students' reported use of active learning strategies in reading and writing activities. Meyer, Turner, and Spencer (1997) found significant differences between challenge seekers and challenge avoiders in learning goal orientations, ability goal orientations, and self-efficacy in project-based mathematics classes. In this study, the construct of students' seeking out challenging work is measured as the reported amount of invested mental effort.

RESEARCH QUESTIONS

The purpose of this study was to describe a variety of classroom assessment events in high school social studies classes. This study was one of a series designed to expand a description of classroom assessment events to other grade levels and subject

areas beyond third-grade language arts (Brookhart & DeVoge, 1999). It is important to describe classroom assessment events across grade levels and subjects because student developmental level and subject area should both matter (Stodolsky, 1988). This study adds one more piece to the puzzle. The research questions were as follows:

1. What are student perceptions of task, self-efficacy, effort, goal orientations, and cognitive strategy use, and what are achievement levels, for a variety of classroom assessment events in social studies classes in an urban high school?
2. Do these descriptions differ from one assessment event to the next? Are patterns related to student level (honors classes, regular classes) or assessment type (paper-and-pencil vs. performance, individual vs. group, teacher-written rubrics vs. student-written rubrics)?

These research questions are descriptive and were examined across several social studies classes taught by a single teacher.

METHOD

This design represents a single-case, descriptive case study (Yin, 1993). Working in one teacher–researcher’s classes afforded an opportunity to sample a variety of classroom assessment events within the same classroom assessment environment, or at least an environment that was very similar from class to class. Although this study’s design does not allow conclusions about the effects of the classroom assessment environment to be drawn, it does function as a case study of one classroom teacher’s assessment environment in fairly rich detail, because 12 classroom assessment events were studied.

The role of theory in case studies that have description as their main objective is to determine what data should be collected. The theoretical framework for this study identified classroom assessment events as the objects for review and identified a set of salient variables used to describe them. The status of the research agenda at this time makes a descriptive study appropriate; the precise function of the classroom assessment events is unknown. After enough descriptive evidence has been presented in studies such as this one, patterns in the descriptions may suggest hypotheses to be tested. This design does not treat the teacher as a sample from a larger, varied population of teachers to which generalization is made, but as an example of a good case in which to examine the theory. Descriptions from this single case are appropriate for making generalizations about the classroom assessment theory to similar cases, teachers in similar situations (Yin, 1993).

Participants

A social studies teacher in a large urban high school (1,500 students, of whom 42% were classified as low income, in a district with an enrollment of 40,000 students) requested to participate as a teacher–researcher in a study of his own classroom assessment. He assisted the researcher in planning the study and collecting the data. Permission to conduct the study was obtained from the high school principal. Informed consent letters were sent to parents. The classes observed included the teacher’s entire teaching load: two regular sections of 10th-grade world cultures, two honors sections of 11th-grade U.S. history, and one section of a philosophy elective that included mostly seniors. This study included data from 12 classroom assessment events, four in each course. A brief description of each assessment event can be found in Table 1.

TABLE 1
Description of Classroom Assessment Events

<i>Event</i>	<i>M Performance</i>	<i>SD</i>	<i>n</i>
U.S. history classroom assessment events:			
11th grade honors track			
Revolutionary America test (paper-and-pencil, individual exam)	75.96%	14.98	26
(Events leading to the) Civil War comic book (performance assessment, group project)	17.84 ^a	3.38	26
History game of events, 1877–1900 (performance assessment, group project, student-written rubrics)	95.52%	8.49	21
Evaluating JFK project (performance assessment, group project, student-written rubrics)	93.86%	19.64	22
Philosophy classroom assessment events: 12th grade elective			
Early philosophers test (paper-and-pencil, individual exam)	96.50%	4.46	12
Philosopher presentation (performance assessment, group project)	12.60 ^b	1.92	15
Hinduism presentation (performance assessment, group project, student-written rubrics)	95.77%	4.00	13
Current issue presentation (performance assessment, group project, student-written rubrics)	80.71%	17.74	14
World cultures classroom assessment events: 10th grade			
Renaissance quiz (paper-and-pencil, individual quiz)	65.25%	16.17	40
Hobbes/Locke conversation (written performance [imaginary conversation that Hobbes and Locke might have had if they had discussed the American Revolution], individual assignment)	18.44 ^c	4.74	27
Industrial age game (performance assessment, group project, student-written rubrics)	77.30%	33.43	37
World War II Time Line (performance assessment, individual assignment, student-written rubrics)	82.32%	17.10	25

^aOn a 20-point rubric scale. ^bOn a 15-point rubric scale. ^cOn a 25-point rubric scale.

The percentage of minority students in each class ranged from 0% to 32%. The attendance rate was 78% schoolwide but much higher in the honors and elective courses and lower in the regular sections. Therefore, the sample of students varied from event to event and was less stable for the regular course than for the honors and elective courses.

Instrumentation

Survey instruments were administered before and after each of the 12 classroom assessments were observed. Seven scales measured the central constructs from the theoretical framework. They were constructed by modifying items from two extant instruments: the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich, Smith, Garcia, & McKeachie, 1993) and the Student Activity Questionnaire (Meece et al., 1988; Meece & Miller, 2001). The primary difference between the original instruments and the measures used for this study was the unit of reference. The original items on both these inventories were generic, and what was needed for this study was a measure specific to an individual classroom assessment event (e.g., not “How hard do you try in school?” but “How hard did you try on the Revolutionary America test?”). Similar adaptations had been developed and used in previous studies (Brookhart, 1997a; Brookhart & DeVoge, 1999).

Students responded to each item by circling choices on 5-point Likert-type scales. The response choices were labels rather than numbers; that is, the responses were from “YES!” to “NO!” or “Very much” to “Not at all” or other words that were appropriate as answers to the particular item. The numbers 1 to 5 were not printed on the students’ version of the inventories; students simply circled the word or phrase that most closely matched their answers to the questions. Validity evidence for adapted versions of these scales was reported in Brookhart and DeVoge (1999).

Four additional scales measuring goal orientations and cognitive engagement were included as variables in this study as a way to help elucidate effort. This decision was based on the finding from Brookhart and DeVoge (1999) that when students explained the reasons for their efforts they often described their goal orientations.

The following subsections describe the presurvey and postsurvey instruments and scales as they were used in this study. They were based on the elementary version used in the previous study, formatted with smaller type, and expanded with the addition of the goal orientations and cognitive engagement scales. All subscale scores were computed as averages. Each subscale had a possible range of 1 to 5, in which higher scores represented more of the construct.

Presurveys. Perceived Task Characteristics (PTC) and Perceived Self-Efficacy (PSE) scales measured the perceived importance of the assessment and

the students' perceived ability to step up to its challenge, respectively. These perceptions were measured before the students participated in the classroom assessment event because the constructs were about students' reported conception of and approach to the assessments. The PTC scale was a five-item scale based on items from the Perceived Task Value scale from the MSLQ (Pintrich et al., 1993). An example of an item on this scale is "It is important for me to learn about [classroom assessment event topic]." Internal consistency reliability was calculated for each of the 12 administrations. The median α value for the PTC scale was .73.

The PSE scale included some items from the MSLQ Self-Efficacy for Performance scale and some items from the "difficulty of today's work" questions from the Student Activity Questionnaire (Meece et al., 1988; Meece & Miller, 2001). An example of an item on this scale is, "How difficult will the [classroom assessment] be for you?" The resulting five-item scale had a median α of .81.

Postsurveys. Perceptions of the amount and nature of student effort for any classroom assessment event must, by definition, be measured after this effort has been expended. Postsurveys for each classroom assessment event were administered after the students had turned in their work but before graded work was returned. The postsurveys included five scales:

First, Amount of Invested Mental Effort (AIME) was measured with modifications of the two items used by Salomon (1984), who first described this construct: "How hard did you try [on the classroom assessment]?" and "How much did you concentrate [when you did the classroom assessment]?" The median α for the AIME scale across the 12 events was .80.

The other four scales on the postsurvey were modifications of the goal orientations and learning strategies scales from the Student Activity Questionnaire (Meece et al., 1988; Meece & Miller, 2001). Item selection and item modification were based on factor analysis results reported in the original research and on the relevance of item content to the classroom context. Mastery Goal Orientation (MGO) was measured with a four-item scale. An example of an item on this scale is, "For [classroom assessment event], I wanted to learn as much as possible." The median α for the MGO scale was .63. Performance Goal Orientation (PGO) was measured with three items. An example of an item on this scale is "I wanted others to think I was smart when I did the [classroom assessment event]." The median α for the PGO scale was .68. Active Learning (ACTL) Strategy Use was measured with a five-item scale. An example of an item on this scale is "I went back over things I didn't understand for [classroom assessment event]." The median α for the ACTL scale was .74. Superficial Learning (SUPL) Strategy Use was measured with five items. An example of an item on this scale is "I checked to see what others were doing for [classroom assessment event] and did it, too." The median α for the SUPL scale was .63.

Validity evidence. Construct validity of the scales as used in this study was investigated by seeking convergent and discriminant evidence. It was reasoned that the correlations among the same construct from assessment to assessment should be stronger than the correlations between different constructs from assessment to assessment. Thus, for example, the correlations among the ACTL scales from each of the four U.S. history assessments should be stronger than the correlations between the ACTL scale for one U.S. history assessment and each of the other scales from each of the other three assessments in that class. Because the theoretical argument pursued in this study held that, even for the same students, different classroom assessment events might invoke different student responses, even the convergent validity coefficients were not expected to be very high. Nevertheless, the pattern of similar scales being more similar to one another than to different scales from other assessment events in the same class seemed a reasonable expectation. The two larger data sets were used to study validity in this way; in the philosophy data set, these correlations had a maximum sample size of 13, too small for stable correlation estimates. Correlations for the U.S. history (max $n = 25$) and world cultures (max $n = 39$) classes were based on maximum available sample sizes. Table 2 presents the results of this validity check. For each scale, the median correlation of the scale with the same scale from other assessments was higher than the median correlation of the scale with different scales from other assessments.

TABLE 2
Convergent and Discriminant Validity Evidence for Scales

Scale	U.S. History ^a		World Cultures ^b	
	Convergent ^c	Discriminant ^d	Convergent ^c	Discriminant ^d
Perceived Task Characteristics 5 items	.60	.08	.31	.11
Perceived Self-Efficacy 5 items	.42	.05	.55	.13
Amount of Invested Mental Effort 2 items	.30	.07	.36	.02
Mastery Goal Orientation 4 items	.28	.11	.20	.12
Performance Goal Orientation 3 items	.56	.02	.23	.18
Active Learning Strategy Use 5 items	.41	.07	.37	.20
Superficial Learning Strategy Use 5 items	.36	-.10	.33	.00

^amax $n = 25$. ^bmax $n = 39$. ^cMedian r (Pearson product-moment correlation) with same scale for different assessments. ^dMedian r with different scales for different assessments.

Another validity concern was whether the repetition of similar measures would affect responses. Informally, this question was investigated by asking a few students, after they had been surveyed several times, if they thought their answers were different depending on which assessment they were asked about. They were sure (“Oh, yes!”) that their considerations, and their answers, were different each time on the basis of the assessment in question. Empirically, Tables 3 through 5 show that responses did vary from event to event, but not in a predictable trend upward or downward, as would be expected if the students were experiencing general fatigue from exposure to the measures.

Procedure

The university researcher conducted an initial observation in each class to note general instructional practices and classroom routines and to familiarize her and the students with each other. After the researcher’s initial observation, the teacher–researcher

TABLE 3
Scale Means and Standard Deviations for U.S. History Classroom Assessment Events

Scale ^a	<i>Revolutionary America Test</i>			<i>Civil War Comic Book</i>			<i>History Game</i>			<i>Evaluating JFK Project</i>		
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>
Perceived Task Characteristics 5 items	3.00	0.67	26	3.19	0.53	28	2.71	0.84	19	3.41	0.53	21
Perceived Self-Efficacy 5 items	2.68	0.81	26	3.79	0.59	28	3.89	0.59	19	3.96	0.43	21
Amount of Invested Mental Effort 2 items	4.27	0.77	24	3.58	0.84	25	3.83	0.78	20	3.40	0.92	21
Mastery Goal Orientation 4 items	2.57	0.79	24	2.99	0.75	25	3.09	0.62	20	2.71	0.74	21
Performance Goal Orientation 3 items	3.15	1.21	24	3.56	1.03	25	3.50	0.68	20	3.00	0.99	21
Active Learning Strategy Use 5 items	3.52	0.70	24	3.61	0.87	25	3.72	0.48	20	3.27	0.84	21
Superficial Learning Strategy Use 5 items	1.93	0.52	24	1.62	0.64	25	1.46	0.52	20	1.54	0.75	21

^aPossible range for each scale = 1 to 5.

TABLE 4
Scale Means and Standard Deviations for Philosophy Classroom Assessment Events

Scale ^a	Early <i>Philosophers Test</i>			Philosopher <i>Presentation</i>			Hinduism <i>Presentation</i>			Current Issue <i>Presentation</i>		
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>
Perceived Task Characteristics 5 items	3.38	0.70	12	3.01	0.95	15	2.97	1.09	13	3.95	0.52	11
Perceived Self-Efficacy 5 items	3.62	0.75	12	3.79	0.69	15	3.54	1.04	13	4.20	0.28	11
Amount of Invested Mental Effort 2 items	3.83	0.62	12	3.19	0.72	13	3.23	1.05	13	2.79	0.75	14
Mastery Goal Orientation 4 items	2.98	1.00	12	3.08	0.57	13	2.79	1.30	13	2.54	0.80	14
Performance Goal Orientation 3 items	2.86	0.82	12	3.36	0.48	13	2.95	0.90	13	2.57	0.95	14
Active Learning Strategy Use 5 items	3.07	0.94	12	3.55	0.79	12	3.20	1.04	13	2.01	0.92	14
Superficial Learning Strategy Use 5 items	1.42	0.40	12	1.70	0.60	12	1.83	0.66	13	1.91	1.08	14

^aPossible range for each scale = 1 to 5.

suggested to her which classroom assessment events would be the 12 observed for the study, and assessments were selected after discussion. Because the theoretical framework suggested that different assessment events might operate differently, this field study was conducted with the intention of including a range of assessment types. The teacher–researcher was the person in a position to know the type, format, difficulty, and expected popularity with students of each of his assessments and thus was able to include variety in the selection of assessment events. In addition, because in this particular study there was no comparison with another teacher, he had no “competition” or need to select assessments for which all students did well (in fact, they did not) or for which his intentions were always realized.

A paper-and-pencil test or quiz was selected as the first classroom assessment event observed in each class; this event was the more conventional form of classroom assessment for these students and provided a comparison against which to view more alternative-style assessment events. For subsequent observations, performance assessments, some of which were group projects and some of which involved the students’ writing their own rubrics, were selected. For each classroom

TABLE 5
Scale Means and Standard Deviations for World Cultures Classroom
Assessment Events

Scale	<i>Renaissance Quiz</i>			<i>Hobbes/Locke Conversation</i>			<i>Industrial Age Game</i>			<i>World War II Time Line</i>		
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>
Perceived Task Characteristics 5 items	2.97	0.84	38	2.44	0.86	37	2.81	0.69	29	3.01	0.84	31
Perceived Self-Efficacy 5 items	3.12	0.82	38	3.22	1.07	37	3.57	1.02	29	3.74	0.88	31
Amount of Invested Mental Effort 2 items	3.92	0.83	39	3.61	0.68	37	3.69	0.79	35	3.92	0.64	25
Mastery Goal Orientation 4 items	2.64	0.99	39	2.51	0.91	37	2.95	0.93	35	3.06	0.83	25
Performance Goal Orientation 3 items	2.97	1.09	39	3.01	1.14	37	2.94	1.20	35	3.04	0.95	25
Active Learning Strategy Use 5 items	2.46	0.96	39	3.36	0.75	36	3.49	0.90	35	3.72	0.74	23
Superficial Learning Strategy Use 5 items	2.02	0.80	39	2.02	0.72	36	1.69	0.70	35	1.62	0.83	23

^aPossible range for each scale = 1 to 5.

assessment event observed, presurveys and postsurveys were administered to the entire class, and selected students were interviewed.

The teacher–researcher identified 2 students per class for the researcher to interview, one low achiever and one high achiever (relative to other members of the class); again, because the achievement of interest was classroom based, the teacher was the person in the best position to identify students at various levels. The researcher conducted the interviews: Because the questions required students to talk about assessments that had been assigned to them by their teacher, it was reasoned that they would talk more freely if they knew he would not hear. Because there were two sections of world cultures and U.S. history, there were four interviews per event for these courses and two per event for philosophy. The interview questions were designed to elicit information about the same concepts as those covered by the questionnaires, but the interview format allowed the researcher to ask the students the reasons for their responses. Interviews were coded into the

same categories as for the questionnaires. A subset of five interviews was coded by two people; coder agreement was 87%. The second coder was a university graduate assistant who had not visited the classroom or met the students or the teacher.

Analyses

Coded interview data were summarized into themes. Descriptive statistics were calculated for the scaled variables. Means and standard deviations for the PTC, PSE, AIME, MGO, PGO, ACTL, and SUPL scales were organized by assessment event and sorted by assessment type. These descriptions were examined for patterns. In addition, zero-order correlations among all pairs of variables were organized by assessment type (paper-and-pencil vs. performance) and examined for patterns. This procedure allowed identification of simple patterns of relationships. Because there was only minor variation between the two sections in each course, results are reported by course, not individual section, as a way to maximize the sample size in each analysis. Maximum available data were used for each analysis because the main research purpose was to describe particular classroom assessment events. Thus, it was important to use as many data as possible within events. The variation in sample sizes represents the real attendance patterns in the classes.

RESULTS

Description of the Classroom Assessment Events

The first research question asked this: For a variety of classroom assessment events, what are student perceptions of task, self-efficacy, effort, goal orientations, and cognitive strategy use, and what are achievement levels, in high school social studies classes? Two analyses addressed this research question. Descriptive statistics are presented for the quantitative variables that measured these constructs. Thematic analysis is presented for the interview questions that were coded according to these same constructs.

Quantitative description. Tables 3 through 5 present the profiles for each classroom assessment event, organized by class. Comparisons among assessments are made in the next subsection.

Qualitative description. Table 6 illustrates themes that emerged from this review by presenting examples of quotations from the interviews. Each quotation is from a different student's interview and is labeled according to the classroom

TABLE 6
Themes and Illustrative Quotations From Student Interviews

Perceived Task Characteristics

- One common perception of the task for all the assessments was that the task was to internalize the teacher's or the text's material—that is, the material “belonged” to someone other than the students.
 - “I knew I'd have to reread all the material he gave us and the notes we'd taken.” (U.S. History, Revolutionary America Test)
 - “I just looked over the notes he gave us on Friday.” (Philosophy, Early Philosophers Test)
 - “Just read the book.” (Philosophy, Hinduism Presentation)
 - “Not much. I just would copy out of the book.” (World Cultures, World War II Time Line)
- “Hard” work often referred to work that would take a nontrivial amount of time, not necessarily conceptually difficult work.
 - “The first day I thought it was going to be really hard and take a lot of time to do all the reading and draw all the pictures.” (U.S. History, Civil War Comic Book)
 - “Well, it wasn't really too easy—I mean you had to go back to the chapter and reread it over again—but when we started putting it together it just worked.... I spent like three class days on it and I worked a little bit at home on it, too.” (U.S. History, History Game: 1877–1900)

Perceived Self-Efficacy

- Some students expressed more confidence working in group because they were working with others. Other students expressed difficulties with group work.
 - “I was pretty confident because I wasn't working alone and I had some ideas as soon as I heard him give the assignment.” (U.S. History, Civil War Comic Book)
 - “Still confident, yeah, cause we worked as a group.” (U.S. History, History Game: 1877–1900)
 - “...but it's hard to do in a group when everyone is not together. You want to learn as much as you can but it's hard to do in that situation.” (Philosophy, Hinduism Presentation)
 - “Because I knew my teammates were competent people, and I expected to be doing most of the work anyway.” (U.S. History, the Evaluating JFK Project)
- The more fun tasks are, the easier the tasks.
 - “Pretty confident because it seemed like not an easy project but one that would be fun to do. The more fun it is, the easier the hard stuff becomes, because you have fun doing it.” (U.S. History, History Game: 1877–1900)

Mastery and performance goal orientations

- There were at least three articulations of motivations to study (besides getting a good grade, the most commonly offered reason): (a) wanting to learn for its own sake, (b) wanting to show what I learned, and (c) wanting to help others learn or to learn from others. The third reason was expressed only for performance assessments featuring group work.
 - “You go in not knowing what you're getting yourself into, but as you work through the process you get a feel for actually mastering the information and completing the assignment itself.” (U.S. History, Evaluating JFK Project)
 - “I'm not sure if [teacher's name] is going to go over it again. I didn't do a good job and they [the class] didn't actually get as much Kant as I would have hoped.” (Philosophy, Philosopher Presentation)
 - “I wanted to be able to show through the game what I learned, and try and help other people learn what was in the section.” (U.S. History, History Game: 1877–1900)
 - “If we play theirs [another group's game], I might learn more about the topic.” (U.S. History, History Game 1877–1900)

(continued)

TABLE 6 (*Continued*)

-
- When the content of the assessment connected with a conscious, expressed interest, students articulated that they wanted to learn more.
 - “I don’t think about this kind of stuff very often. It doesn’t interest me much.” (U.S. History, Civil War Comic Book)
 - “It was easy for me. I’m used to it. It helps when you’re interested in the subject, though. . . . I know about him (JFK) now. I’d hope to [learn more], if the chance ever came up.” (U.S. History, Evaluating JFK Project)
 - “After we started talking about it [gun control], I found out it was really controversial. So I’d like to find out more about what the government does about it and stuff like that.” (Philosophy, Current Issue Presentation)
 - “And besides I find the concept of wars between the U.S. and other countries fascinating and it seems like every decade has a war, and I think we should learn about military tactics.” (World Cultures, World War II Time Line)
 - Students did not care what others thought about their grades, except for significant others (friends). Students did care about what their teacher thought about their work, some because he was the one who assigned the grade and some because he was a significant person for them.
 - “No. Because their opinion doesn’t mean anything to me. I don’t care what they think about me.” (U.S. History, Revolutionary America Test)
 - “No. Because I know I did good and I ain’t worrying if they think I did good or not.” (World Cultures, Industrial Age Game)
 - “I care that my friends are concerned about my grades, as I am theirs. If I don’t know the person, it doesn’t concern me if they’re not worried about my grades.” (World Cultures, Renaissance Quiz)
 - “To make sure I could get a good grade and since I’m passing already, to show [teacher’s name] that I do do my work and put a lot of effort into it.” (World Cultures, Industrial Age Game)
 - “I wanted to get a good grade out of it and I wanted to show [teacher’s name] that I could accomplish the task and make it worthwhile.” (U.S. History, Evaluating JFK Project)
-

assessment event that was the focus of the interview. General themes were identified if several students made similar comments. Idiosyncratic comments are not reported.

Students’ perceptions of the assessment task were characterized by two themes. One common perception of the task for all the assessments was that the task was to internalize the teacher’s or the text’s information—that is, the material “belonged” to someone other than the students. A second theme about characteristics of the task was that “hard” work often referred to work that would take a nontrivial amount of time, not necessarily conceptually difficult work.

Students’ comments about their perceived self-efficacy to accomplish their classroom assessment tasks had to do with their judgments about group work and their enjoyment of their work. Some students expressed more confidence working in groups because they were working with others. Other students expressed

difficulties with group work. Tasks that were described as more fun were perceived as the easier tasks.

Regarding goal orientations, students expressed at least three motivations to study besides getting a good grade, which was the most commonly offered reason: (a) wanting to learn for its own sake, (b) wanting to show what I learned, and (c) wanting to help others learn or to learn from others. The third reason was expressed only for performance assessments featuring group work. When the content of an assessment connected with a conscious, expressed interest, students stated that they wanted to learn more. Students did not care what other students thought about their grades and made many global "It's none of their business" statements. Some students made exceptions to this rule for people who were significant to them: friends, study partners, and their teacher.

Variation by Student Level or Assessment Type

The second research question asked whether there were discernible patterns by student level or assessment type. Two descriptive methods were applied: inspections of patterns in variable means for different assessment events and inspection of patterns in zero-order correlations between pairs of variables. Tests of statistical significance were not used because of varying sample sizes and the presence of unmeasured variables, most notably student ability.

Inspection of individual variables. Task perceptions would be expected to vary across tasks, as they did. Mean PTC and PSE scores varied approximately 1 *SD* among different assessment events within the same class. It is especially interesting that there is variability across assessments even of goal orientations and learning strategy use, which are often studied as general student traits. Goal orientations and learning strategy use did not vary as much as task perceptions within any one class, but they did vary. MGO and PGO, and ACTL and SUPL means for each assessment varied within each class approximately .5 *SD*. This was the case for U.S. history (Table 3), philosophy (Table 4), and world cultures (Table 5).

For the U.S. history class (Table 3), the Revolutionary America Test, a conventional test covering a unit of material, demonstrated the highest mean AIME but the lowest mean MGO among the four assessments studied. PSE for this test was low. Thus, the Revolutionary America Test was an assessment for which the students studied actively and hard, but for which they had low expectations of success.

The Civil War Comic Book, a group project for which students wrote and illustrated comic books about the events leading up to the Civil War, demonstrated fairly high PSE, high PGO, fairly high ACTL, and fairly low AIME. Students wanted their comic books to look good to others, perceived that they could do so,

and worked actively, but they did not think that the assignment required as much thoughtful work as the test did.

The History Game, in which groups constructed board games about the events in American history from 1877 to 1900, demonstrated the lowest mean PTC but the highest mean ACTL and MGO. Students did not think that this assessment was very important, but they wanted to learn the material and worked actively.

The Evaluating JFK Project, in which groups of students had to write a report card or another evaluation of John F. Kennedy's presidency in prescribed categories such as domestic and foreign policy, demonstrated high PTC and PSE but low ACTL. Students thought that this task was important and that they could do it, but they did not report as much active involvement in their learning as that for the other assessments.

In the philosophy class (Table 4), the Early Philosophers Test ranked high in AIME, relatively high in PTC, and in the middle of the other assessments with regard to goal orientations. Superficial Learning Strategy Use was low. As for the U.S. history class, philosophy students reported studying actively and hard for this test. Unlike the U.S. history students, the philosophy class did think that they could be successful on the test.

The Philosopher Presentation and Hinduism Presentation had similar profiles except for goal orientations. Both the MGO and the PGO were lower for the Hinduism Presentation than for the Philosopher Presentation. The Current Issue Presentation, for which the students chose an issue such as gun control on which to report, had a different profile even though the presentation format was the same as that for the other two presentations. This assessment ranked highest in this class for PTC, PSE, and SUPL but lowest on AIME, MGO, PGO, and ACTL. Apparently, even though the students perceived this assessment as important, they did not think it was difficult and did not expend much effort on it.

For the world cultures class (Table 5), the Renaissance Quiz showed fairly high PTC and AIME but low PSE and particularly low ACTL. Like the more able students in the U.S. history and philosophy classes, the world cultures students thought that the paper-and-pencil assessment was important and perceived that they tried hard. For these students, this "trying" was not associated with the use of active learning strategies, as it was for the more successful students.

The Hobbes/Locke Conversation was an assignment to write a hypothetical conversation that the philosophers Hobbes and Locke might have had if they had discussed the merits of the American Revolution. This assignment was abstract, and classroom observation suggested that many of the students did not understand it. As students questioned the teacher, he gave more and more directions and left less and less for the original analysis required of students. Mean PTC, PSE, AIME, and MGO scores were low for this assessment. SUPL was $.5 SD$ higher for this assessment than for the other performance assessments in this class, and the

same as for the Renaissance Quiz. These two assessments were more difficult assessments for these students than the following two were.

The Industrial Age Game was an assignment to create a board game about events during the Industrial Revolution. World cultures students perceived this task to be more important than the Hobbes/Locke writing assignment. They perceived themselves as more able to do the game than to do either the quiz or the writing assignment. However, the performance assessment that ranked highest in this class's perceptions of task, self-efficacy, effort, goal orientations, and active learning, and lowest on superficial learning, was the World War II Time Line, a very concrete and clear task. Students transcribed events from the chapter on World War II in their textbook onto charts constructed according to the teacher's directions. Dates were plotted along a horizontal line through the middle of the paper. European theater events were written above the line, and Pacific theater events below it. The information on each individual's time line was the same; individual variation was apparent only in lettering, graphic design, and completeness.

Inspection of correlations. Tables 7 through 9 present zero-order correlations among variables for each classroom assessment event. Some relationships were similar across all classroom assessment events, some relationships were mixed across events, and some relationships were essentially zero. Observations from these correlations include the following.

For all these classroom assessment events, correlations between MGO and PTC scores were positive and moderate in strength. The more important and interesting that students perceived a task to be, the more likely they were to report wanting to learn about its topic. For all these classroom assessment events, correlations between ACTL and AIME scores were positive and moderate in strength. For all but two of these classroom assessment events, correlations between ACTL and MGO scores were positive and moderate in strength.

Several patterns of relationships were different for tests than they were for performance assessments. In general, correlations of AIME scores with both MGO and PGO scores was mixed for tests but positive and moderately strong for performance assessments. Thus, for performance assessments, perceptions of trying hard were associated with perceptions of wanting to learn both for the sake of learning and for the sake of external judgments on performance.

For one of the tests and all but one of the performance assessments, correlations between SUPL and AIME scores were negative and moderate in strength. For one of the tests and all the performance assessments, correlations between SUPL and ACTL scores were negative and low to moderate in strength. Thus, for performance assessments, students who reported more of either of the active concentration variables—mental effort or active learning—reported less use of superficial strategies such as just doing something to get it done.

TABLE 7
Correlations Among Study Variables for U.S. History Classroom Assessment Events

<i>Correlation Between</i>	<i>Revolutionary America Test</i>	<i>Civil War Comic Book</i>	<i>History Game</i>	<i>Evaluating JFK Project</i>
PSE, PTC	.37	.19	.39	-.10
AIME, PTC	-.17	.06	.24	-.03
AIME, PSE	-.09	.07	.43	-.41
MGO, PTC	.62	.40	.27	.32
MGO, PSE	.11	-.15	.47	-.43
MGO, AIME	-.15	.25	.25	.69
PGO, PTC	-.26	.06	.27	.32
PGO, PSE	-.11	.00	.13	-.18
PGO, AIME	.34	.53	.32	.50
PGO, MGO	-.02	.46	.18	.46
ACTL, PTC	.23	.27	.08	.20
ACTL, PSE	.34	-.08	.09	-.13
ACTL, AIME	.35	.57	.52	.66
ACTL, MGO	.25	.49	.35	.80
ACTL, PGO	.10	.34	.07	.42
SUPL, PTC	.35	-.16	.40	.07
SUPL, PSE	.03	-.36	.23	-.04
SUPL, AIME	.05	-.66	-.31	-.38
SUPL, MGO	-.12	-.20	.09	.05
SUPL, PGO	-.06	-.37	-.03	.17
SUPL, ACTL	.16	-.30	-.37	-.26
ACH, PTC	.16	-.14	.38	-.30
ACH, PSE	.34	.29	.51	.36
ACH, AIME	-.05	-.05	.04	.15
ACH, MGO	.22	-.29	.29	.20
ACH, PGO	-.02	-.28	.09	.26
ACH, ACTL	-.24	-.21	-.08	.45
ACH, SUPL	-.15	-.07	.02	-.10

Note. Correlations were computed with maximum available sample size; n ranged from 18 to 28. PSE = Perceived Self-Efficacy; PTC = Perceived Task Characteristics; AIME = Amount of Invested Mental Effort; MGO = Mastery Goal Orientation; PGO = Performance Goal Orientation; ACTL = Active Learning (Strategy Use); SUPL = Superficial Learning (Strategy Use); ACH = Achievement.

For the tests, correlations between Achievement (ACH) and PSE scores were positive and moderate in strength. For the performance assessments, these correlations were mixed. Thus, for tests, students were more accurate judges of their own achievement. This accuracy may be because students were more accustomed to judging themselves as test takers than as performers in school, or it may be because of low variability in performance assessment scores.

For the tests in the philosophy and U.S. history honors-level classes, PGO and MGO scores were not correlated or were slightly negatively correlated. For

TABLE 8
Correlations Among Study Variables for Philosophy Classroom Assessment Events

<i>Correlation Between</i>	<i>Early Philosophers Test</i>	<i>Philosopher Presentation</i>	<i>Hinduism Presentation</i>	<i>Current Issue Presentation</i>
PSE, PTC	-.22	.29	.42	-.30
AIME, PTC	.50	-.27	.56	.35
AIME, PSE	.30	.47	.60	-.34
MGO, PTC	.62	.26	.83	.46
MGO, PSE	.28	-.17	.53	-.44
MGO, AIME	.69	.29	.75	.73
PGO, PTC	.26	.01	.49	-.13
PGO, PSE	-.16	-.01	.00	.00
PGO, AIME	-.20	.26	.50	.60
PGO, MGO	-.17	.37	.39	.76
ACTL, PTC	.45	-.43	.71	.29
ACTL, PSE	-.19	.09	.64	-.54
ACTL, AIME	.18	.12	.67	.58
ACTL, MGO	.55	-.12	.75	.70
ACTL, PGO	.03	.42	.63	.52
SUPL, PTC	-.27	-.30	-.20	.09
SUPL, PSE	.22	-.55	-.31	.48
SUPL, AIME	-.54	-.31	-.43	.24
SUPL, MGO	-.39	-.25	-.35	-.12
SUPL, PGO	.06	.08	-.02	.18
SUPL, ACTL	-.42	-.20	-.32	-.10
ACH, PTC	-.06	.10	.18	.24
ACH, PSE	.72	-.15	.19	-.43
ACH, AIME	.50	-.02	-.10	.42
ACH, MGO	.45	.56	-.07	.46
ACH, PGO	-.36	.51	.24	.42
ACH, ACTL	-.07	.53	.30	.55
ACH, SUPL	-.36	-.35	-.14	-.15

Note. Correlations were computed with maximum available sample size; *n* ranged from 11 to 15. PSE = Perceived Self-Efficacy; PTC = Perceived Task Characteristics; AIME = Amount of Invested Mental Effort; MGO = Mastery Goal Orientation; PGO = Performance Goal Orientation; ACTL = Active Learning (Strategy Use); SUPL = Superficial Learning (Strategy Use); ACH = Achievement.

all performance assessments, correlations between PGO and MGO scores were positive and moderate in strength. This pattern is interesting. With regard to tests, among the more able students, wanting to learn for its own sake was not related or was negatively related to wanting to learn to garner favorable external judgments and rewards. This is the pattern that educators typically want to encourage: Students should want to learn for its own sake and not worry about being perceived as intelligent by others. With regard to performance assessments, wanting to learn for its own sake is related to wanting to perform well for

TABLE 9
Correlations Among Study Variables for World Cultures Classroom Assessment Events

<i>Correlation Between</i>	<i>Renaissance Quiz</i>	<i>Hobbes/Locke Conversation</i>	<i>Industrial Age Game</i>	<i>World War II Time Line</i>
PSE, PTC	.51	.22	.30	.44
AIME, PTC	.22	.15	.02	-.18
AIME, PSE	.31	.39	.27	-.17
MGO, PTC	.69	.46	.44	.44
MGO, PSE	.42	.35	.27	.23
MGO, AIME	.07	.36	.41	.18
PGO, PTC	.22	.41	.34	.52
PGO, PSE	.48	.40	.39	.23
PGO, AIME	.23	.26	.42	-.18
PGO, MGO	.29	.57	.74	.48
ACTL, PTC	.38	.39	-.01	.09
ACTL, PSE	.51	.24	.12	.47
ACTL, AIME	.51	.17	.54	.45
ACTL, MGO	.60	.44	.30	.03
ACTL, PGO	.53	.63	.61	.02
SUPL, PTC	.15	-.33	.10	-.09
SUPL, PSE	.09	-.35	-.33	-.45
SUPL, AIME	.09	-.34	-.37	-.13
SUPL, MGO	.15	-.06	.17	.13
SUPL, PGO	.15	-.04	.11	.20
SUPL, ACTL	.14	-.12	-.19	-.51
ACH, PTC	.37	.00	-.32	.12
ACH, PSE	.37	.16	.14	.51
ACH, AIME	.10	.13	.05	.22
ACH, MGO	.09	.02	-.15	-.09
ACH, PGO	-.17	.18	.08	-.22
ACH, ACTL	.07	.21	.23	.30
ACH, SUPL	-.18	-.27	-.01	-.59

Note. Correlations were computed with maximum available sample size; n ranged from 21 to 39. PSE = Perceived Self-Efficacy; PTC = Perceived Task Characteristics; AIME = Amount of Invested Mental Effort; MGO = Mastery Goal Orientation; PGO = Performance Goal Orientation; ACTL = Active Learning (Strategy Use); SUPL = Superficial Learning (Strategy Use); ACH = Achievement.

others, and this finding makes sense given the “performance” nature of performance assessments. This pattern held for the presentations—assessments that involved public performances—as well as for the games and written performance assessments. It also held for all three classes. Even in the world cultures class, where PGO and MGO scores were slightly correlated for the quiz (Table 9; $r = .29$), these variables were much more related for the performance assessments than for the quiz.

DISCUSSION

The Classroom Assessment Event

The results from this study, even with its limitations, do allow for one conclusion. Differences from one assessment to the next among student perceptions of the assessment task, their perceived self-efficacy to meet the task, their motivations for doing so, and the effort they reported expending were interpretable in light of the nature of the classroom assessment events and the type of class.

Regarding classroom assessment theory, this interpretability of descriptive results implies that studying the classroom assessment event (Brookhart, 1997b; Stodolsky, 1988) is a productive direction in which to continue research. The classroom assessment environment (Haydel et al., 1999; Stiggins & Conklin, 1992) is a good way to understand the classroom assessment context. However, the environment level (e.g., a class with one teacher) is not a fine-grained-enough analysis for understanding the effects of individual assessments. The importance of specificity that Pajares (1996) described for self-efficacy constructs seems to apply to other constructs related to classroom assessment tasks as well. Studying the effects of classroom assessments on individual students would be yet another level of interest. Although this study's research questions asked about classroom assessment events and not individual students, the variability within both survey and interview questions was a reminder that not all students have the same perceptions, even in the same classroom (Ames, 1992).

Current suggestions about measuring learning, based in validity theory, emphasize matching assessment content and format with information needs. Such matching is of utmost importance. The fact that students do meet different assessments with different perceptions, orientations, and effort suggests, in addition, that knowledge about effects of various types of assessments, various levels of students, and various classroom assessment environments may become informative for validity arguments. This study's results began to show what kinds of assessments are associated with what kinds of motivation and effort perceptions. The interview results hinted at ways that the design of the classroom assessment, including the processes required for completing the task and how much time it would take, affected the constructs actually tapped from the students' view point.

Performance Assessments and Tests

There were some differences by student level and assessment type. The most striking of these were the differences between paper-and-pencil tests and performance assessments. These differences were not apparent in the mean levels of student perceptions and orientations for each event. However, they were apparent in the correlations among these perceptions and orientations and in some of the student interviews.

Student motivation and preparation. Performance assessments evidenced positive correlations among PGO, MGO, and AIME scores, which suggests that performance assessments tapped both internal and external sources of motivation (Ames, 1992; Covington, 1992). Students who reported trying harder, for performance assessments but not for tests, more likely reported both wanting to learn for its own sake and wanting others to see what they could do. Performance assessments evidenced negative correlations among SUPL and both AIME and ACTL. For performance assessments, the students who were more likely to report using active learning strategies and concentrating on their work were less likely to report using superficial learning strategies, whereas this was not the case for tests.

These patterns of relationships support a strong recommendation for further research. It is worth pursuing whether in other classroom assessment environments, performance assessments, but not tests, would be associated with a positive relationship between student effort and both internal and external motivators. It is worth pursuing whether in other classroom assessment environments, performance assessments, but not tests, would be associated with high-effort students using more active learning strategies and fewer superficial strategies. Either or both of these findings, if replicated, would have strong theoretical and practical implications for understanding and using performance assessment within the context of classroom instruction. Validity arguments about performance assessments point to the fidelity between (a) the task and the performance it elicits (b) and the intended construct (Linn, Baker, & Dunbar, 1991). This argument would be considerably expanded if there were evidence that the instructional use of performance assessments in the classroom affected student motivation and preparation as well as performance—that is, if there were evidence that performance assessments included intended and appropriate study and preparation as part of the indicated construct.

What may be happening is that performance assessments stimulate desires for competence so strongly that competence in the absolute sense (mastery goals) and in the normative sense (performance goals) are both engaged (Elliot & Thrash, 2001). Mastery goal orientations are already acknowledged to be associated with deep learning (Ames & Archer, 1988; Covington, 1992). Hidi and Harackiewicz (2000) reviewed studies whose results suggest that simultaneous adoption of both mastery goals and performance goals is associated with more self-regulated learning and higher grades in school. These researchers encouraged testing the hypothesis that mastery and performance goals, together, might interact to predict effort and performance. As recommended subsequently, the results of this study suggest that a productive way to stimulate mastery and performance goals at the same time would be to use interesting, relevant classroom performance assessments.

Student ability and choice. For tests, the correlation between achievement and perceived self-efficacy suggests that the students who perceive themselves to

be more able also scored higher, whereas this was sometimes but not always the case for performance assessments. It is possible that performance assessments allowed students of varying levels of self-efficacy to achieve. The findings of Meyer et al. (1997), who reported higher self-efficacy for challenge seekers than for challenge avoiders in a project-based mathematics classroom, suggest a possible competing hypothesis. It is possible that students did not perceive the performance assessments in this study to be particularly challenging.

Among performance assessments, the distinction between those in which the students had written their own rubrics and those in which the teacher gave students the rubric he would use did not result in any discernible patterns in the data. There may be any one of several reasons for this. Perhaps it really does not make a difference, even though the literature suggests that giving students responsibility for their own learning is important (Covington, 1992). Or, perhaps the rubrics that the students wrote were so patterned after those that ones the teacher usually used that it did not make a clear difference to the students that they had written the rubrics. It is also possible that if grading is based on rubrics, who writes them is less important than whether they are followed and used to guide performance.

There were no discernible patterns in the data to distinguish between group and individual performance assessments in the 10th grade world cultures class, which was the only class with both individual and group performance assessments. Mean levels of the motivational variables for the two individual assessments suggested that students found the Hobbes/Locke Conversation to be the most difficult performance assessment and the World War II Time Line to be the easiest. The group performance assessment, the Industrial Age Game, occupied a middle ground between these two individual assessments, which suggests that for these students in the non-honors track, the level of difficulty was more salient than the group or individual nature of the assignment. Inspection of the correlations suggested that this study's findings of interest that mastery and performance goals were related for performance assessments held for both the individual and the group performance assessments in the world cultures class.

CONCLUSIONS

This study's design examined individual differences within classroom assessment events for the case of one classroom assessment environment. The event is the theoretical "wrinkle" in the unit of analysis necessary for understanding classroom assessment. This examination is closely grained enough to identify nuances and variations sometimes lost in more generic studies (Pajares, 1996; Stodolsky, 1988). This study's results suggest the methodological conclusion that the classroom assessment event is a productive unit of study.

This study's most intriguing substantive result is the suggestion that performance assessments in the classroom may be connected with productive student goal orientations and learning strategies. From a theoretical viewpoint, understanding how classroom assessments are related to motivational and effort variables furthers research agendas in both educational psychology and measurement. Particularly intriguing is the possibility that students' perceptions of the assessments and their goal orientations may be directly related to the nature of the comparison used in the measurement (Elliot & Thrash, 2001) and to the use of performance assessment in the classroom.

The results of this study offer tentative empirical support for current pedagogical interest in performance assessments (e.g., Wiggins, 1998). Performance assessments were associated with higher student self-efficacy than that associated with tests. Performance assessments were associated with a positive relationship between mastery and performance goal orientations, thus potentially tapping several sources of motivation at once. In interviews about group performance assessments, students talked about wanting to help others or to learn from others. Group performance assessments may be more able to elicit the social aspects of motivation (Ames, 1992; Covington, 1992) than the more "inside-the-head" activities of studying for and taking classroom tests are.

Another practical suggestion is implied by the fact that PTC and MGO scores were related for every assessment. This finding supports the recommendation of presenting assessment tasks that are interesting and then helping students to see the value in them. Interest can be developed (Brophy, 1999). Tasks with high relevance are good, but only if they are perceived as real and not too easy to do (Ames, 1992; Bergin, 1999). Suggested strategies include designing assessments that allow students to easily feel ownership of the task, designing assessment tasks that are clearly important and worth working for, and designing assessment tasks that students clearly see they can accomplish with reasonable effort.

REFERENCES

- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84*, 261–271.
- Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology, 80*, 260–267.
- Bergin, D. A. (1999). Influences on classroom interest. *Educational Psychologist, 34*, 87–98.
- Brookhart, S. M. (1997a, March). *The relationship of classroom assessment to student effort and achievement in the college classroom: Pilot study technical report*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Brookhart, S. M. (1997b). A theoretical framework for the role of classroom assessment in motivating student effort and achievement. *Applied Measurement in Education, 10*, 161–180.
- Brookhart, S. M., & DeVoge, J. G. (1999). Testing a theory about the role of classroom assessment in student motivation and achievement. *Applied Measurement in Education, 12*, 409–425.

- Brophy, J. (1999). Toward a model of the value aspects of motivation in education: Developing appreciation for particular learning domains and activities. *Educational Psychologist*, 34, 75–85.
- Covington, M. V. (1992). *Making the grade: A self-worth perspective on motivation and school reform*. Cambridge, England: Cambridge University Press.
- Elliot, A. J., & Covington, M. V. (2001). Approach and avoidance motivation. *Educational Psychology Review*, 13, 73–92.
- Elliot, A. J., & Thrash, T. M. (2001). Achievement goals and the hierarchical model of achievement motivation. *Educational Psychology Review*, 13, 139–156.
- Gipps, C. (1994, April). *Quality assurance in teachers' assessment*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED372086)
- Haydel, J. B., Oescher, J., & Kirby, P. C. (1999, March). *Relationships between evaluative culture of classrooms, teacher efficacy, and student efficacy*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada.
- Hidi, S., & Harackiewicz, J. M. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. *Review of Educational Research*, 70, 151–179.
- Jackson, D. N. (1994, August). *The exploration of a selection of conative constructs relevant to learning and performance* (CRESST/Stanford Project 2.3 Final Rep.). (ERIC Document Reproduction Service No. ED379300)
- Lepper, M. R. (1988). Motivational considerations in the study of instruction. *Cognition and Instruction*, 5, 289–309.
- Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 20(8), 15–21.
- Meece, J., Blumenfeld, P. C., & Hoyle, R. (1988). Students' goal orientations and cognitive engagement in classroom activities. *Journal of Educational Psychology*, 80, 514–523.
- Meece, J. L., & Miller, S. D. (2001). Longitudinal analysis of elementary school students' achievement goals in literacy activities. *Contemporary Educational Psychology*, 26, 454–480.
- Meyer, D. K., Turner, J. C., & Spencer, C. A. (1997). Challenge in a mathematics classroom: Students' motivation and strategies in project-based learning. *Elementary School Journal*, 97, 501–521.
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, 66, 543–578.
- Pintrich, P. R., & Schrauben, B. (1992). Students' motivational beliefs and their cognitive engagement in classroom academic tasks. In D. H. Schunk & J. L. Meece (Eds.), *Student perceptions in the classroom* (pp. 149–183). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement*, 53, 801–813.
- Salomon, G. (1983). The differential investment of mental effort in learning from different sources. *Educational Psychologist*, 18, 42–50.
- Salomon, G. (1984). Television is “easy” and print is “tough”: The differential investment of mental effort as a function of perceptions and attributions. *Journal of Educational Psychology*, 76, 647–658.
- Schunk, D. H. (1994). Self-regulation of self-efficacy and attributions in academic settings. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulation of learning and performance: Issues and educational applications* (pp. 75–99). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Stiggins, R. J., & Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment*. Albany: State University of New York (SUNY) Press.
- Stodolsky, S. S. (1988). *The subject matters: Classroom activity in math and social studies*. Chicago: University of Chicago Press.

- Weiner, B. (1979). A theory of motivation for some classroom experiences. *Journal of Educational Psychology*, 71, 3–25.
- Wiggins, G. (1998). *Educative assessment*. San Francisco: Jossey-Bass.
- Wolf, L. F., Smith, J. K., & Birnbaum, M. E. (1995). Consequence of performance, test motivation, and mentally taxing items. *Applied Measurement in Education*, 8, 341–351.
- Yin, R. K. (1993). *Applications of case study research*. Newbury Park, CA: Sage.