Teachers' Assignments as Indicators of Instructional Quality in Elementary Schools

CSE Technical Report 545

Lindsay Clare, Rosa Valdés, Jenny Pascal, and
Joan Rector Steinberg
CRESST / University of California, Los Angeles

July 2001

Center for the Study of Evaluation
National Center for Research on Evaluation,
Standards, and Student Testing
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90095-1522
(310) 206-1532

Project 1.2 Assessment in Action
Linsday Clare, Project Director, CRESST/UCLA

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R305B60002, as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

TEACHERS' ASSIGNMENTS AS INDICATORS OF INSTRUCTIONAL QUALITY IN ELEMENTARY SCHOOLS

Lindsay Clare, Rosa Valdés, Jenny Pascal, and Joan Rector Steinberg
CRESST/University of California, Los Angeles

Abstract

This report describes ongoing research investigating the use of teachers' assignments as an indicator of classroom practice. The purpose of this work is to develop a measure of students' learning environments that could potentially be used to help monitor the influence of school reform efforts on the quality of the classroom learning environment and support the improvement of instructional practice. Past research indicated good results in regard to the technical quality and feasibility of the method for describing practice in urban school environments. This report describes additional research investigating the reliability and validity of this method for describing the quality of students' learning environments in schools serving higher achieving and middle-class elementary school students as well. Results indicated good reliability between raters, but not within raters over time. Results also indicated that the quality of assignments was statistically significantly associated with the quality of observed instruction and student work. Finally, results indicated that at least three assignments would need to be collected from teachers to determine a stable estimate of quality.

Over the past two decades a number of high-profile reports, such as *A Nation at Risk* (National Commission on Excellence in Education, 1983), have drawn widespread attention to the quality of education in this country. Reports like these have pointed to the need to have higher standards for student learning and an improvement in the quality of instruction. In response to public concern, numerous reform programs have been launched, especially in schools serving primarily poor and minority students. Despite these efforts, however, students who are considered most at risk for academic problems have continued to perform well below state and national norms on standardized tests of achievement.

A broad range of factors influence both student achievement and the quality of reform implementation. One factor that has been identified as a barrier to the success of reform efforts, however, is the difficulty of changing the nature of instruction—the most important school factor influencing student achievement (Cuban, 1990; Tharp & Gallimore, 1988). Even when teachers subscribe to a new instructional practice or seek to implement new standards for instruction, the

quality of the implementation does not always reflect the reform program's intentions. Spillane and Zeuli (1999), for example, studied the ways in which mathematics reform efforts were implemented in 25 elementary and middle school classrooms. They found that, in all but 4 of the classrooms, there was misalignment between the goals of the reform efforts and what teachers and students actually did in their classrooms. Similarly, Cohen and Ball (1994) found that teachers implemented instructional reform policies in different ways, some of which were remarkably different from the intent of the policies. They concluded that teachers too often were left on their own to implement new reforms, and so interpreted these reforms through the "lens" of their own professional experience. The resulting "new" instructional practices thus very often resembled what teachers had been doing previously.

While instructional quality is central to the reform process, limitations in available methodologies have made it difficult to collect information on a broad scale regarding the nature of classroom practice. The quality of students' learning environments has thus been unexplored in many large-scale evaluation designs. Teacher surveys frequently have been used to indirectly assess the quality of students' learning environments, though this method has limitations as far as yielding accurate information about the nature of interactions between teachers and students and interpretations of reform practices (Mayer, 1999). Likewise, analyses of student work have provided very important information about student performance but have not directly assessed or drawn attention to the opportunities students have in the classroom to produce high-quality work. Classroom observations have been the most direct way to measure instructional quality, but these are time-consuming and expensive.

This report describes research focused on developing efficient and effective indicators of classroom practice based on the collection of teachers' assignments and student work. Developing indicators that describe the quality of classroom practice and have the potential to be used in a large-scale study is important for several reasons. First, these indicators can provide important information about the implementation of education reform efforts. This is key to providing high-quality formative feedback that helps diverse reform programs focus their efforts so that they more effectively benefit students. Additionally, indicators of classroom practice can help draw attention to important aspects of students' learning environments, and can help guide teachers' attention toward these areas to improve teaching and

learning (Linn & Baker, 1998). This could be important as well for supporting teachers' reflection and self-evaluation of their practice (Newmann, Lopez, & Byrk, 1998).

The conceptual framework for how the quality of students' learning environments was defined in this study is rooted in a review of research focused on the most effective instructional practices for promoting improved student learning (Newmann et al., 1998; Porter & Brophy, 1988; Slavin & Madden, 1989). This included research based on sociocultural theory (Tharp & Gallimore, 1988), as well as established professional standards for teaching (Danielson, 1996). While the review was not intended to be complete, research indicated that effective teachers set clear instructional goals that were appropriate to student needs and aligned these goals to the content of their lessons and assessment criteria (Aschbacher & Herman, 1991; Danielson, 1996; Duffy, 1981; Porter & Brophy, 1988; Slavin & Madden; 1989). Effective teachers also were knowledgeable about their content area and provided an academically intense curriculum that balanced higher as well as lower level cognitive objectives including a focus on metacognitive strategies (Goldenberg & Gallimore, 1989; Newmann et al., 1998; Porter & Brophy, 1988; Slavin & Madden, 1989). Effective teachers also clearly communicated their expectations to students and provided frequent feedback to students about their progress as well as opportunities to display their knowledge (e.g., through classroom discussions, etc.) (Brophy & Good, 1986; Danielson, 1996; Olson, 1990; Schunk & Swartz, 1993; Slavin & Madden, 1989; Tharp & Gallimore, 1988). Based on this body of research, the quality of classroom practice in this study thus was defined by the degree to which students engaged in complex thinking and used content knowledge in lessons and assignments. Also considered were the clarity of teachers' goals and the alignment of these goals both with assignment and lesson tasks and with the grading criteria teachers used to assess students' work. We also examined students' opportunities to engage in classroom discussions and receive informative instructional feedback.

The research questions for this study were based on findings from the previous years. The research described here was conducted over a 3-year period (see Aschbacher, 1999; Clare, 2000). In previous reports, findings were presented that focused on the validity and reliability of using teachers' assignments as an indicator of classroom practice. Results were promising in terms of the reliability and validity of this method for describing the quality of students' learning environments (Clare, 2000). Results also indicated that our design of collecting four assignments from

teachers yielded a stable estimate of quality, and further analyses estimated that collecting only two assignments from teachers might also yield a stable estimate of practice. A limitation of those findings, however, was that the generalizability of the analyses might have been limited by a restriction in range in our sample in terms of classroom learning environments. All of the schools in the study sample up to that point were in urban areas that served primarily lower achieving, poor and minority students. This restriction in our study sample also potentially limited our ability to fully explore the utility of our method for describing broad variation in students' learning environments. Additionally, we rescored assignments collected at Year 1, so we were not able to investigate intrarater reliability from the first to the second year of the study.

For these reasons, at the third year of the study we scaled up our sample to include schools that served primarily middle-class and higher achieving students in order to investigate whether the pattern of relationships we observed at previous years continued in a much more varied sample of classrooms. We also collected two assignments from teachers (as opposed to four) to explore whether or not collecting fewer assignments would still in fact yield a stable estimate of quality. Additionally, we examined intrarater reliability in order to investigate whether raters were consistent with themselves in terms of how they rated assignments from one year to the next. The primary research questions we addressed, based on this year of data collection, were as follows:

1. How reliable and independent are the classroom assignment rating scales?

2. How many assignments and raters are needed to obtain a consistent estimate of the quality of classroom practice?

3. What is the relation of classroom assignment ratings to other indicators of instructional quality (i.e., classroom observations and students' written work)?

4. How does the quality of assignments and instructional practice differ across third-grade classrooms in lower achieving and higher achieving schools?

Methods

This research took place in a subsample of schools participating in a large-scale school reform initiative. Data sources for this study included teachers' assignments, student work, and observations of language arts lessons.

Sites and Participants

Four third-grade teachers at each eight elementary schools were recruited to participate in the study. Of the teachers who were originally recruited, 4 declined to participate because they were too busy (N = 29 teachers). Four of the sample schools served mostly poor Latino and African American students (n = 13 teachers). The students in these schools were lower achieving on the whole in reading as assessed by the Stanford Achievement Test (SAT-9). On average, 28% of the third-grade students at these schools scored at or above the 50th National Percentile Rank (NPR) on the 1999-2000 SAT-9 in reading. These schools served a high proportion of English language learning students, though all assignments and student work were in English due to recent California legislation that all children be instructed in English unless they received a special exemption.

Four of the sample schools served primarily middle-class students, the majority of whom were White and Asian (n = 16 teachers). The students at these schools were primarily higher achieving in reading as assessed by the SAT-9. On average, 81% of the third-grade students at these schools scored at or above the 50th NPR in reading on the 1999-2000 SAT-9. The average number of years teachers had been teaching was approximately 12.5 years, with a range from 1.0 to 34.0 years. Table 1 presents the demographic characteristics for both groups of elementary schools.

Table 1

Demographics and SAT-9 Scores for Elementary Schools (N = 8)

	Low-achieving schools (n = 4)		High-achieving schools (n = 4)	
	Mean %	% Range	Mean %	% Range
Enrollment by ethnicity				
Asian	7.6	0.3–27.0	22.5	5.5–40.3
African American	12.1	1.4–20.0	4.4	3.0–5.2
Latino	62.9	34.0–92.1	16.1	9.8–24.6
White	14.6	3.8–37.3	55.0	36.9–79.3
Other	2.9	1.0–7.0	2.0	0.8–3.1
English language learner	66.3	50.4–82.2	9.0	4.2–13.1
Free / reduced lunch	89.9	86.7–93.8	11.2	7.4–17.0
1990-00 SAT-9 scores at or about 50th NPR in reading for Grade 3	28.0	14.0–45.0	81.8	77.0–94.0

Procedures

Teachers were asked to submit two language arts assignments. These were a "typical" writing and "typical" reading comprehension assignment. For each assignment, teachers completed a one-page information sheet and submitted four samples of student work that they considered to be of "medium" quality and "high" quality. The teacher assignment materials (notebook, cover sheets, consent forms, etc.) were distributed in fall 1999 and collected in winter and spring 2000 (see Appendix A).

All but three of the teachers who submitted assignments were observed in winter 2000 (N = 26). The three teachers who declined to participate in this part of the study were from schools that served the higher achieving students. These teachers reported that they were too busy to have us visit their classrooms. Observations lasted for one class period and were of a "typical" language arts lesson. Before each observation, we contacted principals and asked them to suggest dates and times when we could visit teachers' classrooms. We then contacted teachers to confirm that these dates and times were convenient for them. We also briefly interviewed the teachers about their lessons before we observed their classrooms. These interviews were approximately 15 minutes long and focused on the goals for the lesson, instructional context, and the specific needs of the students in the classroom (see Appendix B).

Measures

As described in the introduction of this report, our criteria for looking at the quality of classroom assignments were based on research investigating effective teaching practices and were embedded a standards-based approach to curriculum and teaching (Clare, 2000). Based on this research, we used a 4-point scale (1 = *poor* to 4 = *excellent*) to rate the following six dimensions of quality for each assignment.

Cognitive challenge of the task. This dimension describes the level of thinking required of students to complete the task. Specifically this dimension describes the degree to which students have the opportunity to apply higher order reasoning and engage with academic content material. For example, an assignment given a high score for cognitive challenge might require students to synthesize ideas, analyze cause and effect, and/or analyze a problem and pose reasonable solutions using content-area knowledge (e.g., comparing themes from different books, etc.). An

assignment given a low score on this dimension, in contrast, might require students only to recall very basic, factual information.

An assignment that received a high score on this dimension required students to write a new ending to a chapter book they had read, *Mr. Popper's Penguins*, by Richard Atwater. The assignment required students to engage with literary content and to use their creative writing skills by writing an original and developed story whose elements aligned with the facts of the original story in the book. The teacher expected students to use sophisticated vocabulary and sentence structure such as that in the following example: "Nelson and Columbus fought out of the circle of penguins and . . . tobogganed down the stairs to a secret room." An assignment given a low score for this dimension, in contrast, asked students to write a story about what they would find in a pot at the end of a rainbow. Students were required to write no more than one paragraph consisting of about five sentences that were basic in content and structure. The story that the teacher required was in fact merely a listing of the contents of the pot. The low cognitive demand of the task was further evidenced by the teacher's providing students with the first sentence of the paragraph, "My pot at the end of the rainbow has . . . ."

Clarity of the learning goals. This dimension describes how clearly a teacher articulates the specific skills, concepts, or content knowledge students are to gain from completing the assignment. The primary purpose of this dimension is to describe the degree to which an assignment could be considered a purposeful, goal-driven activity focused on student learning. An assignment given a high score on this dimension would have goals that were very clear, detailed, and specific as to what students are to learn from completing the assignment. It would also be possible to assess whether or not students had achieved these goals. For example, the following set of goals received a high score: "We expected the students to continue developing paragraphs that develop a central idea. It was expected that they stay on the topic, give details, and show awareness of audience." The goals of one assignment that received a low score, in contrast, read as follows: "I wanted them to properly express their ideas and answer the prompt correctly."

Clarity of the grading criteria. The purpose for this dimension is to assess the quality of the grading criteria for the assignment in terms of their specificity and potential for helping students improve their performance. How clearly each aspect of the grading criteria is defined is considered in the rating, as well as how much detail is provided for each of the criteria. An assignment given a high score for this

dimension would have grading criteria in which the guidelines for success were clearly detailed and provided a great deal of information to students about what they needed to do to successfully complete the task. Most of the assignments that received a high score on this dimension used writing rubrics with well-differentiated and elaborated score points focusing on a number of critical aspects of writing.

For example, an assignment that received a high score on this dimension included a rubric that consisted of three dimensions measuring different aspects of students' written work: writing strategies, writing applications, and writing conventions. Each dimension was assessed on a 4-point scale (1 = *beginning* to 4 = *exceeds standards*). Each scale point for each dimension of this rubric gave detailed, elaborated information about what was expected from students' writing. On the other hand, a lower score was assigned to the following writing rubric because of its relative lack of specificity. This rubric consisted of a simple 4-point scale where a 4 was designated as *excellent—the writing is fluent and articulate*, and a 1 was designated as *not satisfactory—incomplete and difficult to understand.* The rubric did not include separate dimensions for different aspects of students' written work and did not provide much information to students about what they needed to include in their writing to successfully complete the task. Finally, a very low score was assigned to the following criterion stated for the grading of an assignment: "Teacher judgment."

Alignment of goals and task. This dimension focuses on the degree to which a teacher's stated learning goals are reflected in the design of the assignment tasks students are asked to complete. Specifically, this dimension attempts to capture how well the assignment appears to promote the achievement of the teacher's goals for student learning. An assignment given a high score on this dimension would involve tasks and goals that overlapped completely.

For example, an assignment whose goals and task were well aligned had the goals of improving students' reading comprehension by having them analyze a text and connect it to their own lives. The assignment called for students to reflect on the story, *The Hundred Dresses*, by Eleanor Estes, and then to write responses to questions like "What did we learn?" "With which characters do we identify the most?" "What connection can we make with the story?" Prior to the writing portion of the assignment, the teacher engaged students in discussions around community,

friendship, diversity, fear and prejudice, the author's purpose in writing the novel, and her use of descriptive language and character development.

In contrast, the learning goals for an assignment given a low score on this dimension were that students be able to show their understanding of a text, *Chicken Sunday,* by Patricia Polacco, by connecting what they read to their own experience and that they learn to appreciate the ideas of others. This story is about how three children raise money to buy a hat for their "gramma" to thank her for her chicken dinners on Sundays. The actual assignment, however, required students to write a description of a project for which they needed to raise money by selling decorated eggs. The link to the rich content of the story, which addresses both intercultural and intergenerational issues, was superficial and indirect. Additionally, there was no evidence that the assignment helped promote students' understanding and appreciation of the ideas of others.

Alignment of goals and grading criteria. This dimension is intended to describe the degree to which a teacher's grading criteria support the learning goals, that is, the degree to which a teacher assesses students on the skills and concepts they are intended to learn through the completion of the assignment. Also considered in this rating is whether or not the grading criteria include extraneous dimensions that do not support the learning goals, as well as the appropriateness of the criteria for supporting the learning goals. An assignment given a high score for this dimension would have goals and grading criteria that overlapped completely. An assignment given the lowest score on this dimension, in contrast, would have grading criteria that did not support the learning goals. An example of an assignment that received a high score is one whose goals were to have students improve their writing skills by engaging them in the writing process and by having them practice interviewing their peers and writing up the results of their interviews. The teacher also wanted her students to learn to distinguish good interview questions from bad ones. The grading criteria the teacher used to measure attainment of these goals focused on the extent to which the student's write-up contained "factual information [gathered] using the question-and-answer form," and the extent to which "readers get to know the person interviewed through questions the student has asked."

An assignment that was judged to be lacking in this alignment was one in which the teacher's goals were to improve students' reading comprehension and to teach them how to answer comprehension questions fully and in detail. In

describing her grading criteria for the assignment, however, the teacher wrote only "Each question is worth 20 points. Partial answer 10 points. Doesn't apply 0 points." The teacher's grading criteria did not reference the skills that were directly connected to her goal that the students develop reading comprehension skills (e.g., ability to provide a complete plot summary, identify theme, make a prediction about what might happen next based on previous events, etc.), nor did she provide students with detailed information in her criteria that would help them understand how to complete the assignment successfully.

Overall quality. This dimension is intended to provide a holistic rating of the quality of the assignment based on its level of cognitive challenge, the specificity and focus of the learning goals, the clarity of the grading criteria, the alignment of the learning goals and the assignment task, and the alignment of the learning goals and the grading criteria (see Appendix C).

Each assignment was scored by three independent raters on these dimensions. These were the same raters (N = 3) who scored teachers' assignments in the first and second year of the study (Clare, 2000). For this year of scoring, the raters underwent approximately 1 week of training before scoring the assignments. This training included scoring assignments individually and as a group, and selecting anchor papers by scale point and dimension to calibrate ratings and refine the rubric. Overall exact scale-point agreement between at least two raters for the classroom assignment scales was 87%.

We scored student work from the writing assignments (a final writing project with earlier drafts). Student writing was rated by an experienced bilingual rater using three standards-based scales measuring organization, content, and MUGS (i.e., mechanics, language use, grammar, and spelling). These scales were from the Language Arts Project rubric developed by LAUSD and United Teachers-Los Angeles in partnership with CRESST at UCLA (Higuchi, 1996). Each of these dimensions was rated on a 4-point scale (1 = *poor*, 4 = *excellent*). This rater was highly trained to use these scales (see Appendix D).

Observations were conducted by research staff who were experienced in qualitative methodology (N = 3). Researchers wrote detailed field notes describing the classroom, lesson activities, and the interactions between the teacher and the students. The length of each activity (measured in number of minutes) and the number of students involved in each of the observed lesson activities also were

recorded and categorized according to the social organization, behavior of the teacher and students, resources in use, and language arts content. A 4-point scale (1 = *poor*, 4 = *excellent*) was used to describe the overall quality of the observed lesson for the following eight dimensions.

1. Cognitive challenge of lesson activities. This dimension describes the level of thinking required of students to participate in the observed lesson activities; for example, the degree to which students had the opportunity to think critically, to predict, analyze, and synthesize information, and to engage with substantive content material.

2. Quality of classroom discussions. The quality of the classroom discussion, or instructional conversation, also is considered as a critical dimension of classroom practice. This dimension captures the extent to which the teacher provided students with the opportunity to learn through and engage as partners in meaningful classroom discussions. This includes the nature of a teacher's questions, the degree to which student contributions are extended and built on, and the amount of time spent in discussion.

3. Level of student participation in classroom discussions. This scale is intended to describe the percentage of students who engaged in classroom discussions. Fewer than a quarter of the students participating in classroom discussions were rated a 1, and more than three quarters of students engaging in the discussion were rated a 4.

4. Quality of instructional feedback. This dimension describes students' opportunity to receive information about their performance and progress toward learning goals and the degree to which this feedback appears to support learning. The accuracy, substance, specificity, and helpfulness of the teacher's feedback are considered in the ratings, as well as the amount of feedback the teacher provided to students during the observed lesson.

5. Level of student engagement in the lesson. This scale is intended to capture the level of student engagement in the observed lesson activities. As with the scale describing student participation in classroom discussions, this scale describes the percentage of students who appeared to be on task and participating in the lesson activities. Fewer than a quarter of the students engaging in the lesson activities were rated a 1, and more than three quarters of the students appearing to be actively engaged in the lesson activities were rated a 4.

6. Lesson implementation/classroom management. This dimension is intended to describe the degree to which a teacher effectively carries out the lesson activities. This scale focuses on a teacher's classroom management skills, including the amount of time spent on transitions from one activity to another or on procedural tasks, and how disruptive or distracting student behavior was handled.

7. Clarity of the learning goals. This dimension is intended to capture the degree to which a teacher is able to articulate the specific skills, concepts, or content knowledge students are to gain from participating in the lesson or lesson activities. This information is obtained from interviewing the teacher prior to observing the lesson. The primary purpose of this dimension is to capture the degree to which lessons could be considered purposeful, goal-driven activities focused on student learning versus "activity for activity's sake." In other words, this dimension attempts to differentiate between teachers who plan their lessons with clear and specific learning goals in mind versus those who plan activities with no clearly defined learning objective. The quality of the activities themselves—which the goals may or may not describe—is not considered in this rating.

8. Alignment of goals and lesson activities. This dimension attempts to capture the degree to which a teacher's stated goals for the lesson are reflected in the design of the learning activities. Specifically this dimension attempts to capture how well the learning activities promote the achievement of the teacher's goals for student learning (see Appendix B).

Training for observers was conducted during the previous year of data collection and included coding videotapes as a group and observing in pairs in non-sample third- and seventh-grade classrooms at two different points during the year prior to observing in our sample schools. Reliability was assessed by comparing the scores for each possible pair of observers. Overall exact scale-point agreement was 77.5%.

Analyses

Descriptive statistics were used to describe the quality of teacher assignments, classroom observations, and student work. Content analyses of ethnographic data based on field notes from classroom observations, teacher assignments, and student work also were conducted to further investigate the range of observed quality and relationship to ratings. Cohen's kappa coefficients were calculated to investigate the proportion of agreement between raters after chance agreement was removed.

Cronbach's alpha coefficients were calculated to estimate the internal consistency of the ratings (Abedi, 1996). Factor analyses were conducted to explore the interrelationship of scale items and dimensions of quality practice for teacher assignment and classroom observation ratings, and to reduce the data. Correlation coefficients were computed to examine the relation of teacher assignment ratings to ratings of classroom observations and student work, and to investigate the interrelationship of the rating scales. Regression analyses also were conducted to investigate the relationship of assignment quality to the quality of student work. Finally, a generalizability study was conducted to investigate the consistency of our classroom assignment ratings, and a decision study was conducted to explore options for future research designs.

Results

In this section, results are organized around the different research questions based on statistical analyses of teachers' assignments, student work, and classroom observations. Ethnographic "portraits" of two classrooms—one from a lower achieving school and one from a higher achieving school—also are presented to further explore the findings from the statistical analyses and to illustrate the range of learning environments for students.

How Reliable and Independent Are the Classroom Assignment Rating Scales?

Overall, results indicated an acceptable level of interrater agreement for both the reading comprehension assignments and the writing assignments. As shown in Table 2, kappa coefficients for the reading comprehension assignments were significant at the $p < .001$ level. Alpha coefficients were acceptable and ranged from .76 for the scale measuring the alignment of goal and task to .94 for the scale measuring the overall quality of the assignment. The percent of exact scale-point agreement between three raters averaged 62%, and ranged from 48% for the scale measuring the alignment of the goals and grading criteria to 69% for the scales measuring the challenge of the assignment, the clarity of the learning goals, and the clarity of the grading criteria.

For the writing assignments, kappa coefficients also were significant at the $p < .001$ level, with the exception of the scale measuring the alignment of goal and task, which was significant at the $p < .01$ level (see Table 3). Alpha coefficients for this assignment also were acceptable and ranged from .70 for the scale measuring

Table 2

Reliability of Rating Scales for the Reading Comprehension Assignments (N = 29 Teachers)

Scale	Kappa	Alpha	% Agreement 3 raters	% Agreement 2 raters
Cognitive challenge of the task	.63***	.91	69	83
Clarity of the learning goals	.61***	.92	69	86
Clarity of grading criteria	.66***	.92	69	92
Alignment of goals and task	.41***	.76	59	76
Alignment of goals and grading criteria	.43***	.80	48	91
Overall quality	.80***	.94	59	89

***p < .001.

Table 3

Reliability of Rating Scales for the Writing Assignments (N = 29 Teachers)

Scale	Kappa	Alpha	% Agreement 3 raters	% Agreement 2 raters
Cognitive challenge of the task	.47***	.87	48	91
Clarity of the learning goals	.50***	.89	59	89
Clarity of grading criteria	.76***	.96	76	87
Alignment of goals and task	.27**	.70	41	82
Alignment of goals and grading criteria	.65***	.91	72	83
Overall quality	.55***	.86	41	94

**p < .01. ***p < .001.

the alignment of goals and task to .96 for the scale measuring the clarity of the grading criteria. The percent of agreement between three raters averaged 56%, somewhat lower than for the reading comprehension assignments, and ranged from 41% for the alignment of goals and task to 76% for the clarity of the assessment criteria. The percent of agreement between two raters averaged 88% and ranged from 82% for the alignment of goals and task, to 94% for the overall quality of the assignment. For both assignment types the percent of agreement between raters within one scale point was 100%.

To investigate intrarater agreement, the same raters (N = 3) rescored reading comprehension and writing assignments chosen at random from the previous year of data collection (N = 40). The scores individual raters gave to the same assignments at different years were compared, and the percent of agreement from

one year to the next was calculated. This analysis revealed that raters were not very consistent with themselves over time. Intrarater exact scale-point agreement averaged 63% from one year to the next and ranged from 58% to 73%. This pattern was consistent across all scales. Intrarater agreement within one scale point was 100%.

Intrarater inconsistency over time could be explained by various reasons. First, the researchers who had rated the assignments had collected assignments and observed lessons in very different classrooms during the third year of the study (i.e., classrooms in higher achieving schools) and so may have changed their understanding of the scoring dimensions based on this new information. Additionally, the passage of time in light of our continued discussion about the scales may also have been responsible for changes in the raters' scores.

Whatever the reasons, we concluded from our findings that caution must be advised if classroom assignment scores are to be used to gauge *change over time* in the quality of classroom practice. In our own evaluation work we have circumvented this problem by re-rating assignments in order to minimize potential error introduced by rater inconsistency over time. Pending future development and standardization of this measure, taking such precautions is critical, especially in settings where different raters are used to score assignments at different years.

Most of the assignment scales were statistically significantly associated with each other at the $p < .001$ level. The exceptions to this general pattern were the scales measuring the quality of the grading criteria teachers used to assess student work and the alignment of teachers' learning goals for the assignment and their grading criteria (see Table 4). These scales were significantly associated with each other ($r = .68$, $p < .001$). The quality of the assessment criteria teachers used to score student work, however, was not associated with any of the other scales. The scale measuring the alignment of the learning goals for the assignment and the grading criteria only was associated with the scale measuring the clarity of the learning goals for the assignment ($r = .40$, $p < .05$).

A factor analysis using orthogonal rotation with these scales confirmed this same general pattern (see Table 5). This analysis revealed two underlying dimensions for the assignment scales. The first and most robust factor (accounting for 53.2% of the variance explained) included the cognitive challenge of the task,

Table 4

Intercorrelation of Classroom Assignment Scoring Dimensions (N = 29 Teachers)

	Cognitive challenge of the task	Clarity of the learning goals	Clarity of grading criteria	Alignment of goals and task	Alignment of goals and grading criteria	Overall quality
Cognitive challenge of the task	1.00					
Clarity of the learning goals	.66***	1.00				
Clarity of grading criteria	-.16	.09	1.00			
Alignment of goals and task	.75***	.59***	.12	1.00		
Alignment of goals and grading criteria	.11	.16	.68***	.32	1.00	
Overall quality	.83***	.59***	.10	.82***	.40*	1.00

*p < .05. **p < .01. ***p < .001.

Table 5

Exploratory Factor Analysis of Classroom Assignment Rating Scales (N = 29 Teachers)

Variables	Assignment Factor 1	Assignment Factor 2
Cognitive challenge of the task	**.89**	-.17
Clarity of the learning goals	**.73**	.14
Alignment of goals and task	**.87**	.28
Overall quality	**.91**	.13
Clarity of grading criteria	-.07	**.91**
Alignment of goals and grading criteria	.28	**.87**
Eigenvalue	3.19	1.56
Percent of variance explained	53.2	25.9

clarity of the learning goals, alignment of goals and task, and the overall quality of the task. The second factor (accounting for only 25.9% of the variance explained) included the clarity of the grading criteria and the alignment of the learning goals and the grading criteria (see Table 5).

These results differed from our previous analyses where the scoring dimensions were mostly all significantly associated with each other and formed one

factor (Clare, 2000). Why would the interrelationship of scales be different across years? As noted earlier, the lower achieving schools in our sample had been part of a 3-year reform effort that focused on improving students' literacy skills. Part of this reform effort included creating and distributing rubrics for scoring student work in language arts. During the third year of data collection (reported here) we found that these rubrics had made their way into many of our sample classrooms serving low-achieving students, and teachers had included them with the assignments and student work they submitted to us. A likely reason for why these scoring dimensions stood apart from the other dimensions of quality is because these rubrics were not generated by our sample teachers and so would not necessarily be linked to other aspects of their practice. We suspected, therefore, that the classroom assignment factor measuring the clarity of teachers' grading criteria and the alignment of their learning goals and their grading criteria were not serving as accurate measures of the quality of instruction in these classrooms. We investigated this issue further by looking at the relation of these classroom assignment factors with other indicators of instructional quality. The results of these analyses are presented later in this report.

How Many Assignments and Raters Are Needed to Obtain a Consistent Estimate of the Quality of Classroom Practice?

In previous years of the study, our design based on four assignments and three raters revealed G-coefficients of .91 and .87 for elementary and middle schools respectively (Clare, 2000). Based on those results, and the results of a decision study conducted with those data, we decided to collect only two assignments from teachers during this year of the study. The purpose for this was to investigate whether collecting those few assignments would yield a consistent estimate of quality.

Contrary to what we expected based on prior analyses, our design based on two assignments and three raters revealed a G-coefficient of only .64—an unacceptable level of stability. Similar to our previous year's investigation, the estimated variance components of the teacher assignment ratings continued to show (as we had hoped) that most of the variation in the ratings was accounted for by differences across teachers and not by differences across raters and assignment type (see Table 6). Teachers differed quite a bit, however, in the quality of the assignments they submitted to us, which points to the need to collect more assignments from teachers in the future to obtain a stable estimate of quality.

Table 6

Estimated Variance Components and Percent of Variance Explained by Teacher, Assignment Type, and Rater (N = 29 Teachers)

	Variance components	% of variance explained
Teacher	.0817	36.4
Assignment type	.0170	7.6
Rater	.0011	0
Teacher by assignment type	.0760	33.9
Teacher by rater	.0000	0
Assignment type by rater	.0000	0
Teacher by assignment type by rater	.0487	21.7

We also conducted a decision study in order to estimate generalizability coefficients for varying numbers of assignments and raters. The results of the decision study indicated that increasing the number of assignments (and not raters) would significantly improve the design (see Table 7). Based on our current data, the results of this decision study indicated that the minimum number of assignments needed to yield a consistent measure of quality was four, and the minimum number of raters was two (G-coefficient = .77). We concluded from this that it would likely be necessary to collect as many as four assignments from teachers to obtain a consistent estimate of quality.

Table 7

Estimated G-Coefficients Based on the Number of Assignments and Raters (N = 29 Teachers)

Number of assignments	Number of raters	Estimated G-coefficient
3	1	.66
3	2	.70
3	3	.73
4	1	.72
4	2	.77
4	3	.78

What Is the Relation of Classroom Assignment Ratings With Other Indicators of Instructional Quality (Classroom Observations and Students' Written Work)?

To investigate the validity of using teachers' assignments as an indicator of classroom practice we correlated the two classroom assignment factors described earlier with ratings of observed classroom instruction and student work. As a first step we conducted an exploratory factor analysis with the classroom observation variables to reduce the data. This analysis revealed two underlying dimensions.[1] The first factor measured student engagement in the lesson activities, the alignment of the teacher's goals for the activity, the quality of classroom management, the clarity and focus of the teacher's goals on student learning, and the challenge of the lesson activities. The second factor measured the quality of the observed instructional conversation, student participation in that discussion, and the quality of the teacher's instructional feedback (see Table 8).

We then investigated the relation of the classroom assignment factors with the classroom observation factors and measures of the quality of student work. This analysis revealed that the first classroom assignment factor (measuring the cognitive challenge of the assignment, the clarity of the learning goals, the alignment of the goals and task, and the overall quality of the assignment) was statistically significantly associated with the quality of observed practice (Observation Factor 1:

Table 8

Exploratory Factor Analysis of Classroom Observation Variables (N = 26 Teachers)

Variables	Observation Factor 1	Observation Factor 2
Level of student engagement in the lesson	**.79**	.12
Alignment of goals and lesson activities	**.75**	.11
Lesson implementation / classroom management	**.74**	.38
Clarity of the learning goals	**.73**	.26
Cognitive challenge of the lesson activities	**.66**	.59
Quality of classroom discussions	.17	**.94**
Level of student participation in classroom discussions	.32	**.87**
Quality of instructional feedback	.15	**.71**
Eigenvalue	4.36	1.20
Percent of variance explained	54.5	15.0

[1] See Clare, 2000, for a description of the factor structures for the observation variables at previous years.

r = .53, p < .001; Observation Factor 2: r = .43, p < .01). This pattern of relationship was similar to what we saw in our previous year of data collection. In contrast, the second classroom assignment factor (measuring the quality of the grading criteria and the alignment of the goals with the grading criteria) was associated negatively with the first observation factor (r = -.34, p < .05), and was not associated with the second observation factor.

The relation between the quality of writing assignments and student work showed a similar pattern of results. These findings indicated that the first classroom assignment factor (measuring the cognitive challenge of the assignment, the clarity of the learning goals, the alignment of the goals and task, and the overall quality of the assignment) was associated with the quality of the content of student work (r = .40, p < 0.05). This classroom assignment factor also was associated with the quality of writing mechanics (also referred to as MUGS, for mechanics, usage, grammar, and spelling; r = .49, p < 0.01). The second classroom assignment factor (measuring the quality of the teachers' grading criteria and the alignment of the goals with the grading criteria), in contrast, was not associated with the quality of student work (see Table 9).

In summary, teachers' classroom assignments were associated with the quality of their observed practice—specifically with regard to the cognitive challenge of the assignments, the quality of the learning goals for the assignments, the alignment of goals and task, and the overall quality of the assignments. This pattern of relationships based on both higher and lower achieving schools was similar to what was found during the previous year of data collection. The quality of teachers' classroom assignments also was associated with the quality of student work.

Table 9

Relationship of Classroom Writing Assignments and Student Work (N = 29 Teachers)

	Student writing scales		
	Content	Organization	MUGS
Classroom Assignment 1: (Cognitive challenge of the task, Clarity of the learning goals, Alignment of goals and task, Overall quality)	.40*	.33	.49**
Classroom Assignment 2: (Clarity of grading criteria, Alignments of goals and grading criteria)	-.08	.24	-.09

Note. MUGS = Mechanics, Usage, Grammar, and Spelling.

*p < .05. **p < .01.

As we suspected, the clarity of teachers' grading criteria and the alignment of their goals and grading criteria were not associated with the quality of student work, however, and were even negatively associated with some variables measuring the quality of observed instruction. As we described earlier, we believe that this finding could be explained by the fact that many teachers in the lower achieving schools were using rubrics developed by outside sources (e.g., districts). Teachers in these schools tended to have poorer quality lesson activities overall in comparison to the teachers in the higher achieving schools. This could explain the negative relationship (or trend) we found between the quality of observed practice and the variables associated with the assessment criteria teachers used to grade student work.

How Does the Quality of Assignments and Instructional Practice Differ Across Third-Grade Classrooms in Lower Achieving and Higher Achieving Schools?

In addition to investigating the technical quality of the classroom assignment ratings across a broad range of classrooms, we also were concerned with looking at the variation in students' learning environments as revealed by the quality of their assignments. We conducted additional exploratory analyses as well to look at the relation of classroom assignments to the quality of student work in schools serving lower achieving versus higher achieving students.

The results of *t*-test analyses indicated that the classroom assignments from the higher achieving schools were rated significantly higher than the assignments from lower achieving schools with respect to variables that made up the first assignment factor (see Table 10). These were the level of cognitive challenge ($p < .001$), clarity of the learning goals for the assignment ($p < .01$), alignment of these learning goals with the assignment task ($p < .05$), and the overall quality of the assignment ($p < .001$). As suggested by our previous analyses, the quality of the grading criteria teachers used to assess student work was rated higher in the lower achieving schools, though this difference was not statistically significant. These schools were rated slightly higher with respect to the alignment of the teachers' learning goals for students and their assessment criteria, though again this difference was not statistically significant. As described earlier, our sample of schools serving lower achieving students had been part of a reform effort that emphasized (in part) the development and dissemination of rubrics for scoring student work. These rubrics were new in many of our sample classrooms this year, which likely explains why this dimension was rated higher than other dimensions of quality in these schools (e.g., cognitive challenge).

Table 10

Quality of Assignments in Classrooms Serving Traditionally Lower and Higher Achieving Students (N = 29 Teachers)

	Lower achieving (n = 13)		Higher achieving (n = 16)		
	M	(*SD*)	*M*	(*SD*)	*p* value
Cognitive challenge of the task	1.64	(.44)	2.23	(.61)	.000
Clarity of learning goals	1.92	(.50)	2.32	(.56)	.007
Clarity of grading criteria	2.37	(1.01)	1.94	(.66)	.07
Alignment of goals and task	1.83	(.49)	2.17	(.48)	.013
Alignment of goals and grading criteria	1.81	(.59)	1.71	(.55)	.52
Overall quality	1.71	(.43)	2.21	(.48)	.000

Note. Items were scored on a 4-point scale (1 = *poor*, 4 = *excellent*).

We also observed teachers who returned assignments to us (with the exception of a few teachers who asked not to be observed). Analysis of these observations yielded the same general pattern of results as our analysis of teachers' assignments. Overall, teachers in the higher achieving schools had better quality lessons as assessed across all of the scoring dimensions, most notably with regard to the level of cognitive challenge, the clarity of learning goals for the lesson, and the quality of classroom discussions (see Table 11).

Table 11

Quality of Observed Language Arts Lessons (N = 26 Teachers)

	Lower achieving (n = 13)			Higher achieving (n = 13)		
	M	(*SD*)	Range	*M*	(*SD*)	Range
Cognitive challenge of lesson activities	1.93	(.27)	1.00–2.00	2.92	(.79)	2.00–4.00
Quality of classroom discussions	1.07	(.27)	1.00–2.00	1.75	(.87)	1.00–3.00
Level of student participation in classroom discussions	1.00	(.00)	1.00–1.00	1.92	(1.00)	1.00–3.00
Quality of instructional feedback	2.00	(.55)	1.00–3.00	2.50	(1.00)	1.00–4.00
Level of student engagement in the lesson	2.93	(.27)	2.00–3.00	3.83	(.39)	3.00–4.00
Lesson implementation / classroom management	2.57	(.65)	1.00–3.00	3.42	(.79)	2.00–4.00
Clarity of the learning goals	2.21	(.58)	1.00–3.00	2.92	(.67)	2.00–4.00
Alignment of goals and lesson activities	2.29	(.47)	2.00–3.00	2.67	(.65)	2.00–4.00

Note. Items were scored on a 4-point scale (1 = *poor*, 4 = *excellent*).

Case Studies of Two Classrooms

We looked at assignments, student work, and field notes from classroom observations for individual teachers in our sample to confirm our findings from the statistical analyses, as well as to gain a more qualitative understanding of the utility of this method for describing variation in students' learning environments. In this section, we describe our findings for two third-grade classrooms—one from a lower achieving school (Ms. Lincoln) and one from a higher achieving school (Ms. Roosevelt). While the quality of classroom practice tended to be higher in the higher achieving schools than in the lower achieving schools, the quality of individual teachers' instruction clearly ranged within these settings. We chose to describe these two classrooms because these teachers illustrate what we considered in our ratings to be "moderate" and "excellent" quality practice.

Classroom Assignments and Student Work in Two Classrooms

The following section describes the writing and reading comprehension assignments and student work from our two case study teachers. Ms. Lincoln was a teacher at a lower achieving school, and Ms. Roosevelt was a teacher at a higher achieving school. The quality of their assignments is described according to each scale dimension.

Cognitive challenge of the assignments. Both of the assignments submitted by Ms. Lincoln received moderate ratings (score point 2 on a 4-point scale) on their level of cognitive challenge. The writing assignment asked students to write a description of a time that they brought something to school to share in class. Students were to write three paragraphs over two successive drafts, in between which they received feedback from the teacher. The paragraphs followed a chronological order in the description of events, with some description about how students felt about sharing their possession in class. In their written work students mainly described what happened in relatively simple sentences, such as "The whole class was sitting on the floor. We were all sitting in my theird [*sic*] grade classroom." Ms. Lincoln's reading comprehension assignment asked students to read a book independently then complete a worksheet in which they wrote five sentences (one for each of the following story elements) describing the setting, characters, problem, plot, and conclusion of the book.

In contrast, the assignments provided to us by Ms. Roosevelt, the teacher in a higher achieving school, were both given a high score for level of cognitive

challenge. Before the writing assignment, the class visited a local desert preserve as part of a larger unit focused on learning about their community. For this trip the teacher hired a naturalist to teach the students about the native plants and animals. The class had a discussion on important events and facts about their trip, and students wrote a report summarizing what they learned. Students were expected to produce well-organized paragraphs written in conventionally correct sentences. They included in their reports such details as "One particularly interesting part of our trip that I liked was seeing my mineral. It is called a Dalmatian Granite. The black spots are called hornblende." The reading comprehension assignment provided to us by this teacher asked students to read and discuss in literature circles *The Big Wave*, by Pearl Buck, and *The Island of the Blue Dolphins*, by Scott O'Dell, and then write an essay comparing and contrasting the two novels on key story elements. With assistance from the teacher, students were expected to write about sophisticated ideas in the stories, as exemplified by the following excerpt from a student's essay: "Jiya and Karana, characters from the books 'The Big Wave' and 'Island of the Blue Dolphins,' have many differences, however, they both needed to face their losses of their family."

Clarity of the learning goals. The goals of both assignments from Ms. Lincoln's classroom also received moderate ratings (score point 2 on our 4-point scale). The goals for each assignment were relatively clear in terms of what students were expected to learn, but they tended to be broadly stated and were not very elaborated. For example, for the writing assignment Ms. Lincoln's goals included that the students stay on topic, address all aspects of the writing prompt, and write four paragraphs including an introduction and conclusion. For the reading comprehension assignment her goals were that the students identify the various story elements independently.

Ms. Roosevelt's goals for both of her assignments, in contrast, were far clearer, more specific, and more elaborated in terms of what students were to learn. Both sets of goals are described here to demonstrate their high quality. As Ms. Roosevelt wrote for her writing assignment:

> My main overall goal was that students learn to take experiential information and transfer it in an organized way into main points and detail. Specifically in writing, we have worked all year on topic sentences, main points, evidence, description, and conclusion. I continually give students opportunities to format information in this way. One other important goal for the individual was to document their experience to keep in

> their [community journal book], from which this assignment was taken. It is therefore a part of a much larger project.

Her goals for the reading comprehension assignment in which students compared two books were:

> The broad learning goal for this assignment was to give students experience in organizing their thoughts about the story element of conflict for a main character. The goal was analysis of character traits and synthesis of story elements. They could show me their skill by comparing two characters from two books, as opposed to simply answering questions about a character. A specific skill we worked on just prior to the writing assignment the second day was how to write a comparison sentence. I did a mini-lesson on the use of the word "however" in a sentence. We discussed forms that comparison sentences could take, such as using "but," "however," "or," "also," "both," etc.

Clarity of grading criteria. For the writing assignment, Ms. Lincoln referred to a very detailed and elaborated district-generated rubric which she included with her assignments, but she did not specify how she used the rubric. Ms. Roosevelt, in contrast, referred to a rubric and described in general how she used it but did not attach it. As such, it is difficult to make any sort of meaningful comparisons between these classrooms in terms of this dimension for this assignment.

For the reading comprehension assignment, Ms. Lincoln provided the following information about her grading criteria, "I was looking for papers with 1) complete sentences, 2) accurate information, 3) detailed description." These criteria, while somewhat specific, only outlined what the teacher was looking for in students' work. Ms. Roosevelt's grading criteria were more detailed and elaborated. These criteria focused on the extent to which students compared ideas from the two texts and used examples from the texts as well as on the following aspects of writing: content (e.g., use of comparative sentences, evidence of opinion, topic sentences) and mechanics (e.g., sentence structure, punctuation, etc.).

Alignment of goals and task. Both assignments from Ms. Lincoln's class were given an average rating in terms of how their learning goals aligned with the actual task. The goal of one assignment was to teach students how to identify the setting, characters, problem, and conclusion of a story selected by the students. The actual assignment asked students to complete a basic worksheet in which they wrote approximately one sentence per story element. We considered the alignment between learning goals and design of the task to be of moderate quality because there was a basic overlap, but the task did not extend students' understanding of the

elements of narrative to a very great extent. The assignments from Ms. Roosevelt's classroom, on the other hand, had excellent alignment between the learning goals and the design of the task. For example, the goals of one assignment were very specific about how to teach students to analyze text. The grading criteria were also specific in their focus on the learning goals. That is, the teacher looked in students' work for evidence of the learning she had expected to take place, for example, if they used examples from the text in their writing, if they used complex sentences that compared two ideas, etc.

Alignment of goals and grading criteria. The match between Ms. Lincoln's goals and grading criteria was moderate. The primary reason for this was that both her learning goals and her grading criteria were somewhat vague. As such, these two dimensions were aligned at only a general level. For example, the goal of the reading comprehension assignment was to teach students to identify story elements in a book, and the grading criteria were that students use complete sentences, accurate information, and detailed descriptions. There was not complete overlap. There was much more of an overlap, however, between the learning goals and grading criteria of Ms. Roosevelt's assignments. For example, her learning goals for one assignment were essentially to teach students how to organize their thoughts about the story element of conflict for a main character, and specifically, to practice the skill of writing a comparison sentence. The criteria that the teacher used to evaluate these goals focused on the extent to which a student organized (using complex topic sentences and proper paragraphing) a composition, illustrated his or her ideas through comparisons, opinions, and examples from the books, and used comparative sentences.

Visiting Ms. Lincoln's and Ms. Roosevelt's Classrooms

The lesson that was observed in Ms. Lincoln's classroom, like her assignments, tended to be of moderate quality in terms of the level of cognitive challenge. The lesson lasted about 45 minutes. During the lesson, Ms. Lincoln held a short discussion in which she reviewed concepts about conservation and explained an assignment that students were to work on individually. Then students completed the worksheets on their own as she helped individual students. The following excerpt is from a brief (15-minute) review session at the beginning of the lesson.

T: If I say that it is very important to conserve trees, what am I saying about them?

S: To take care of them.

T: Right, to take care of them. What else am I saying?

S: Don't waste them.

T: Yes, I'm saying I don't want to waste them.

S: Recycle stuff.

T: Okay. So when we say that we need to conserve trees, we mean we need to take care of them. We mean we want not to waste them.

S: They help us breathe.

T: Yeah, they help us breathe, that's true. Yesterday we talked a lot about why people need trees. What are some reasons people need trees?

S: To make paper and pencils.

T: Yes. Why else? Piper gave a good reason. What was it, Celeste?

S: They help us breathe.

T: Right. What else?

S: For shade?

T: Good. So we know why people need trees. Besides being good for people, animals also need trees. What are some reasons that animals need trees?

At the end of this review, Ms. Lincoln gave a brief explanation of the worksheet that students needed to complete. Then the class spent about half an hour completing the assignment. The assignment was to complete a simple worksheet on which students labeled parts of a drawing using vocabulary terms like owl, branch, nut, and squirrel, and then wrote six sentences explaining why animals and people need trees.

The lesson we observed in Ms. Roosevelt's classroom, in contrast, received the highest score on the rubric in terms of the level of challenge it provided to students. The observed activities also lasted about 45 minutes, but in this class the entire time was spent in an in-depth discussion about a book that the class had been reading. The class had been studying various themes in fairy tales, comparing elements across stories. During this observation, the teacher read another fairy tale out loud to the class—"Princess Furball," by Charlotte Huck. As she read, she paused now and then, either to hear a student's volunteered observation and encourage the class to discuss it, or to point out her own observations. As the teacher read, each student wrote on a teacher-prepared sheet notes about what the story made them feel, see, touch, taste, and smell, as well as metaphors or similes. As the teacher read, she also

paused to let the students write down their thoughts. For example, she paused after she read descriptive phrases, such as “A coat that is soft and downy and warm” and “Silvery as the moon, as glittery as a star.” She paused every few pages to ask students questions about the themes and events in the story, and brief discussions would emerge. This is illustrated in the following exchanges between the teacher and the students.

T: Are you looking for a trickster in this story?

S: I think there are two tricksters.

T: Oh, Tyler thinks that there are two tricksters. Who are the tricksters?

S: The king and the princess.

T: Oh, and what are the tricks?

[No response.]

T: Winston, what’s the trick?

S: The princess wants to trick the king by making him make her all these things, and the king wants to give the princess away.

T: Okay, good. That’s a good observation.

. . .

T: Are things getting worse for Furball? Are there complications?

S: Yeah, dying.

T: Okay, dying. What was the first complication of dying? Who died first?

S: Her mom.

T: Okay, then who?

S: Her nurse.

T: Okay, now what’s another complication?

S: She has to marry an ogre.

T: Okay, then what?

S: She had to be a servant.

T: Okay, she had to be a servant to servants. Could things get any worse?

S: Yeah, she has to marry the ogre.

T: Okay, we already said that. Let’s read on to see if we reach the climax.

Throughout the story, the students freely chimed in with predictions and connections to other stories. All the students listened attentively and wrote on their sheets. Many participated verbally. A few times, the teacher held up a star and said, “Co-Ca” aiming it at students for making contributions. The letters stood for

Connection Captain. Holding up the star was an "award" in response to a student's making a connection to another story the class had read or to his or her life experience. After reading the story, the teacher called on students and asked them to read from the notes they had been writing during the reading of the story. The following is an excerpt from that portion of the lesson:

T: Okay, someone tell me what "sight" words were in Furball that you recognized. Oh, let's do sticks because I know everyone has some sight examples. Okay, John, first sight.

S: Snowy, dark, and silent.

T: Okay, so that was partly sight and partly hearing. [Pulls out the name of another student.] Another one for sight.

S: Puffy hair, good food, and Princess Furball is clever.

T: Okay, these are sight, so which of those can you see?

S: Puffy hair.

T: Okay, one more. [Pulls the name of another student.]

S: I put "hair as good as gold," but that's a simile.

T: Okay, that's good because often similes are used to describe how things look.

The Influence of Assignment Quality on Student Work in Classrooms Serving Lower Achieving Versus Higher Achieving Students

Research has indicated that the quality of classroom practice may be more critical to the success of students who do not have access to the same opportunities to develop their literacy skills outside of schools that more privileged students have (see for example, Black & Wiliam, 1998). In addition to investigating differences in the quality of students' learning environments, therefore, we also explored the relation of assignment quality to student work in classrooms serving lower achieving students (n = 13 teachers) versus higher achieving students (n = 16 teachers). These analyses were extremely limited by the small sample size of both groups, however, and so these findings should be considered exploratory at best. Commensurate with other research, our results indicated that the quality of classroom assignments (i.e., the first classroom assignment factor) significantly predicted the quality of the content (b = .72, R^2 = .33, $p < 0.05$) and organization (b = .89, R^2 = .32, $p < 0.05$) of student work in those schools serving lower-achieving students. Assignment quality, in contrast, did not predict the quality of student work in those schools serving higher achieving students. Further research with larger numbers of classrooms would be needed to draw more definitive conclusions

about the quality of assignments and student work in classrooms serving differing populations of students.

Summary

As with our previous findings (Clare, 2000), the results from this year's data collection indicated an acceptable level of interrater agreement for both the reading comprehension and the writing assignments. Although interrater agreement was acceptable overall, intrarater exact scale-point agreement averaged only 63%. This lack of within-rater consistency could have been caused in part by the fact that assignments were collected and lessons were observed in a much broader range of classrooms during this year of the study, which may have influenced the raters' understanding of the scoring dimensions. Additionally, a full year had passed since the raters scored assignments. This also could have contributed to the lack of intrarater agreement. Results of correlation and factor analyses indicated that the scales measuring the cognitive challenge of the task, clarity of the learning goals, alignment of goals and task, and the overall quality of the assignments were associated with each other and created one factor. The scales measuring the quality of the grading criteria and the alignment of goals and grading criteria formed another factor. This was contrary to our previous investigation where we found that the scale items formed a single factor and were highly interrelated (Clare, 2000).

Our results based on this year of data collection revealed that two assignments were not enough to obtain a stable estimate of quality. Similar to our previous year's investigation, the estimated variance components based on the teacher assignment ratings continued to show that most of the variation in rating was accounted for by differences across teachers, and not by differences across raters or assignment type. Teachers varied, however, in the quality of the assignments they submitted to us, which is likely why we did not obtain a stable estimate of quality based on only these two assignments. It appears that it would be necessary to collect more than two assignments—as many as four—and have them scored by at least two raters to obtain a stable estimate of quality.

To explore the validity of using teachers' assignments as an indicator of classroom practice, we investigated the relationship between the classroom assignment scales and the quality of the observed language arts lessons. We found as before that the scales measuring the cognitive challenge of the task, clarity of the learning goals, alignment of goals and task, and the overall quality of the

assignment were statistically significantly associated with the quality of the language arts lessons we observed. These scales also were associated with the quality of the student work. Exploratory analyses indicated as well that the quality of classroom assignments might be better predictors of the quality of student work in classrooms serving lower achieving versus higher achieving students. Additional analyses would need to be undertaken with larger samples of classrooms to confirm this finding, however. The classroom assignment scales measuring the quality of the grading criteria and the alignment of goals and grading criteria, in contrast, were not associated with the quality of student work, and were negatively associated with some aspects of observed instruction.

The classroom assignments we collected from the higher achieving classrooms were rated significantly higher with respect to the level of cognitive challenge, clarity of the teacher's learning goals for the assignment, alignment of these learning goals with the assignment task, and the overall quality of the assignment. The quality of the observed practice followed this same general pattern. Teachers in the lower achieving schools were rated higher overall with respect to the quality of their grading criteria, though this difference was not statistically significant. Many of the teachers in the lower achieving schools were using rubrics designed by outside sources (e.g., their district), which likely accounted for this difference.

The presence of these rubrics likely also accounted for the contrary relationship we observed between the classroom assignment variables measuring the quality of teachers' grading criteria and other indicators of instructional quality. Other research has indicated that improving classroom assessment tools improves the quality of the classroom learning environment by raising teachers' standards (see Black & Wiliam, 1998, for a review of this research). While it is very possible that these rubrics might ultimately contribute to raising the level of instructional quality, it appeared in this year's study that the standards the rubrics imply for student success were not yet reflected in other aspects of classroom practice (e.g., the level of challenge of the assignments and lessons, etc.). For these assessment tools to serve a formative purpose, the results they yield about the quality of student work would need to be used to adjust teaching and learning. The rubrics were new in most of these classrooms this year, however, and it would likely take teachers time and participation in guided, ongoing professional development sessions to make the necessary adjustments.

Recommendations and Future Research

We concluded from this year's investigation that although some of the classroom assignment scales were problematic, on the whole, our method of looking at teachers' assignments served as a valid and reliable indicator of classroom practice across diverse educational settings. Based on what we learned over the course of this year's investigation, we make the following recommendations.

- Because of the conflicting pattern of results we obtained with regard to the relationship of the teachers' grading criteria and other indicators of instructional quality, we recommend that it be noted when the rubrics teachers are using have been generated by outside sources. This is important so that this information can be considered appropriately in the analyses.
- To facilitate the scoring process in a large-scale evaluation setting, it might be necessary to reduce the number of dimensions that are rated. In this case, our analyses indicated that the level of challenge of the assignments, and the overall quality of the assignments (a holistic rating of all of the other dimensions) showed good interrater reliability and were associated with the quality of observed instruction and student work. We recommend, therefore, that at least these scales be retained. Analyses would need to be conducted to determine whether the use of only two scales provides a good estimate of quality. If the intention of using this method is to provide diagnostic information to teachers (i.e., to support teacher reflection and self-evaluation), however, it would likely be important to retain all, or at least most, of the scales.
- We recommend that extreme caution be exercised if the goal of implementing this method is to use the assignment rating scores to track progress over time. As described before, in our evaluation work we circumvented the problem of intrarater inconsistency by rescoring assignments collected at previous years. This was time-consuming, however, and would not be feasible in a large-scale evaluation effort. Establishing anchor assignments to exemplify the different scale points would most likely facilitate rater consistency over time, but this would need to be investigated in future research efforts.

- We also recommend that additional people (including teachers) be trained to score assignments. This would yield important information about the reliability of scores, as well as the feasibility of training a relatively large number of raters (e.g., how much time it takes, how difficult it is, etc.).
- Our results indicated that our design of two assignments and three raters did not yield a stable estimate of quality. Based on our results, it appears that, at least with a small sample size (around 30 teachers), it would be necessary to collect as many as four assignments and score them using two raters. The number of assignments that need to be collected is influenced by the overall sample size, however, so with a larger effort it might be necessary to collect fewer assignments. This would need to be investigated in future studies.

  In determining how many assignments to collect from teachers, however, it is important to consider numerous other issues in addition to the potential stability of the research design. One consideration is the amount of effort collecting assignments entails for teachers and researchers. Teachers, especially in urban schools, are being asked to participate in numerous reform activities. Completing the assignment cover sheets and photocopying samples of student work could be considered burdensome by many teachers. In our work we ameliorated this problem by offering assistance to teachers (for example, we offered to photocopy their student work for them). We also paid teachers a small stipend as a token of appreciation for their time and effort. Despite this, collecting the assignments from teachers was a time-consuming task that required a great deal of effort on our part.

  It also is important to remember that teachers need an appropriate length of classroom time to generate the requisite number of assignments. Specifically, teachers need enough time to teach the lessons and collect and grade the student work as well as time to fill out the cover sheets (we estimate that this takes about a half hour to 45 minutes per assignment) and photocopy student work. During the first year of the study (1997-98) we distributed the materials and collected the assignments in the spring. Teachers protested that they had not been given enough time, however, and also that we were asking too much from them at the end of the year when they were preparing for student testing. Based on this feedback, we revised

our procedures for the next two years of data collection so that we distributed the assignment cover sheets and other materials to teachers in the fall (generally October) and collected the completed materials in February and March. This avoided spring testing and provided teachers with a long enough period of time to generate assignments.

- This year we narrowed our investigation to focus on a more varied sample of third-grade classrooms. We recommend that future research efforts focus on investigating the reliability and validity of our method in higher and lower achieving middle and high schools as well. Additionally, while our preliminary analyses suggested that assignment quality is more predictive of the quality of student work in classrooms serving lower achieving students than in classrooms serving higher achieving students, more research with larger samples would be necessary to draw definitive conclusions about this relationship. It also would be important to look more carefully at the specific role assignment quality might play in the achievement of students who have differing home literacy environments and opportunities.
- Finally, future research should also focus on the use of this method by teachers to self-evaluate and reflect on their practice. This year we met informally with teacher-coaches who provided us with feedback about how to improve our method so that it could be more useful to teachers in professional development settings. These teachers were overall very positive about using the assignment scale dimensions as a framework for helping teachers reflect on the quality of their work. About half the teachers reported, however, that the framework should be used for reflective and not for evaluative purposes. Teachers also suggested that anchor assignments should be included with the different scale dimensions and the language in the rubric should be tied to the standards being used by a particular district or school. We suggest that future research explore *how* this framework is used by teachers in professional development settings, and the utility of the framework for helping teachers improve the quality of the assignments they give to students.

References

Abedi, J. (1996). Interrater/test reliability system (ITRS). *Multivariate Behavioral Research, 31,* 409-417.

Aschbacher, P. R. (1999). *Developing indicators of classroom practice to monitor and support school reform* (CSE Tech. Rep. No. 513). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Aschbacher, P. R., & Herman, J. (1991). *Final report of the Humanitas program evaluation, 1990-91*. Los Angeles: University of California, Center for the Study of Evaluation.

Atwater, R., & Atwater, F. (1938). *Mr. Popper's penguins* (Robert Lawson, Illustr.). Boston: Little, Brown.

Black, P., & Wiliam, D. (October, 1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80,* 139-148.

Brophy, J., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 328-375). New York: Macmillan.

Buck, P. S. (1948). *The big wave* (Hiroshige & Hokusai, Illustrs.). New York: J. Day.

Clare, L. (2000). *Using teachers' assignments as an indicator of classroom practice* (CSE Tech. Rep. No. 532). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Cohen, D. K., & Ball, D. L. (1994). Relations between policy and practice: A commentary. *Educational Evaluation and Policy Analysis, 12,* 249-256.

Cuban, L. (1990). Reforming again, again, and again. *Educational Researcher, 19*(1), 3-13.

Danielson, C. (1996). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.

Duffy, G. (1981). Teacher effectiveness research: Implications for the reading profession. In M. Kamil (Ed.), *Directions in reading: Research and instruction. 30th yearbook of the Nation Reading Conference* (pp. 113-136). Washington, DC: National Reading Conference.

Estes, E. (1944). *The hundred dresses* (Louis Slobodkin, Illustr.). New York: Harcourt, Brace.

Goldenberg, C., & Gallimore, R. (1989). Teaching California's diverse student population: The common ground between educational and cultural research. *California Public Schools Forum, 3,* 41-56.

Higuchi, C. (1996). *Improving student learning: High standards, standards-based curriculum, and standards-based assessment models.* Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Huck, C. S. (1989). *Princess Furball.* (Retold by Charlotte Huck; Anita Lobel, Illustr.; 1st ed.). New York: Greenwillow Books.

Linn, R. L., & Baker, E. L. (1998, Fall). School quality: Some missing pieces. *CRESST Line,* pp. 1-3, 7. Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Mayer, D. P. (1999). Measuring instructional practice: Can policymakers trust survey data? *Educational Evaluation and Policy Analysis, 21,* 29-45.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform. A report to the nation and the Secretary of Education*. Washington, DC: U.S. Government Printing Office.

Newmann, F. M., Lopez, G., & Bryk, A. S. (1998). *The quality of intellectual work in Chicago schools: A baseline report.* Chicago: Consortium on Chicago School Research.

O'Dell, S. (1960). *Island of the blue dolphins.* Boston: Houghton Mifflin.

Olson, V. P. B. (1990). The revising processes of sixth-grade writers with and without peer feedback. *Journal of Educational Research, 84,* 22-29.

Polacco, P. (1992). *Chicken Sunday*. New York: Philomel Books.

Porter, A. C., & Brophy, J. (1988). Synthesis of research on good teaching: Insights from the work of the Institute for Research on Teaching. *Educational Leadership, 45*(8), 74-85.

Schunk, D. H., & Swartz, C. W. (1993). Goals and progress feedback: Effects on self-efficacy and writing achievement. *Contemporary Educational Psychology, 18,* 337-354.

Slavin, R., & Madden, N. (1989). What works for students at risk: A research synthesis. *Educational Leadership, 46*(5), 4-13.

Spillane, J. P., & Zeuli, J. S. (1999). Reform and teaching: Exploring patterns of practice in the context of national and state mathematics reforms. *Educational Evaluation and Policy Analysis 21*, 1-27.

Tharp, R. G., & Gallimore, R. (1988). *Rousing minds to life: Teaching, learning, and schooling in social context.* Cambridge, England: Cambridge University Press.

APPENDIX A

TEACHER ASSIGNMENT MATERIALS

Directions for Collecting Assignments and Student Work
Step-By-Step Process

Due: February 2000

Please collect 2 assignments with 4 samples of student work for each assignment. You will be asked to fill out a cover sheet for each assignment. Detailed instructions are given below.

We want to describe the nature of the language arts tasks that students do, what is expected of them, what feedback they are given, and how grades are assigned. Our descriptions depend on what you tell us, so please be explicit and detailed so we can be as accurate as possible. **If you have any questions about any of the following instructions, please call Joan Steinberg at 310-206-1532 x71262.** Thank you.

1. COLLECT THE FOLLOWING 2 ASSIGNMENTS.

Between now and February, collect **2** of the assignments you give your students, with selected examples of student work. Use assignments which ask students to do some individual written work. Do not create new assignments specifically for this study. Please collect the following types of assignments:

- 1 reading comprehension or reading response assignment
- 1 writing assignment that includes a rough draft and final draft, with any written feedback given by peers or teachers (writers' workshop activities are fine)

(over)

2. FOR EACH OF THE 2 ASSIGNMENTS COPY 4 SAMPLES OF STUDENT WORK.

- **Choose two middle-quality and two high-quality pieces of student work from the same class.**
  It is fine to choose different students' papers for the different assignments. We just need two middle and two high for each assignment. If there were no students who did high-quality work on an assignment, attach a note explaining why you are not including any "High" pieces of student work. In that case, please just give us 4 pieces of medium-quality work.
- Copy the four pieces of student work for each assignment.
- Place an ID sticker over each student's name. (We prefer to receive student work without their names so as to protect their privacy.) Please do not cover up any part of the student's work, your feedback, or grade. It is important for us to see the feedback comments or grades. If there is no clear area for the label, put it on the back of the work and cross out / white out the student's name.
  - Note: The student ID labels for Assignment #1 are stapled to the pocket for Assignment #1, and so forth.
- Place an **M (Middle) or H (High)** sticker on each student paper accordingly. These stickers are stapled to each pocket.

We would prefer to see student work from 3rd and 7th graders where possible. However, if you have students of another grade in your class and cannot find high- or middle-level work from your 3rd or 7th graders, record the grade level at the top of the student work if it is from a student who is not in 3rd or 7th grade (e.g. write "grade 2" at the top if you cannot find high-level work from a 3rd grader and instead submit a high-level work sample from a second grader in your 2 / 3 combination class). Call us if you are unsure.

3. FILL OUT A COVER SHEET FOR EACH OF THE 2 ASSIGNMENTS.

Fill out the enclosed Cover Sheets for Teacher Assignments in the pockets in this binder.

- **Please attach whatever will help us understand** the assignment and accompanying student work, such as the following:
  - copy of the directions given to students (**please be as explicit as possible**),
  - grading rubric or guidelines, and
  - outline of the unit.
- Place the cover sheet with attached papers and the 4 pieces of student work in the appropriate pockets in this binder.

General Information Form
3rd Grade Teachers

Please answer the following questions for the 1999-2000 academic year.

1. What grade(s) do you teach? _____

2. Including this year, for how long:

   - have you been a teacher? _____ years
   - have you been a teacher at your current grade level? _____ years
   - have you worked at this school? _____ years

For the following questions, please answer for the class in which you gave the two assignments and collected the student work:

3. What grade level(s) are your students? Check all that apply. [ ] 2nd [ ] 3rd [ ] 4th

4. How many years have you been teaching? _____ years

   a. How many years have you taught 3rd grade? _____ years

5. How many students are enrolled in your class? _____

6. Approximately what percentage of your students have been in your class since the beginning of the school year? _____ %

7. Please circle any of the following which describe your class:

   a. full bilingual b. modified bilingual c. SDAIE or sheltered English
   d. English only e. other (explain)_____

8. Approximately what percent of your students are LEP (Limited English Proficient)? _____ %

   a. In what language(s) do your LEP students receive language arts instruction? (Circle as many as apply.)

      English Spanish Other _____

   b. Approximately what percent of your students have recently (within the past six months) been redesignated as Fluent English Proficient (RFEP)? _____ %

9. a. What is the range in reading level among your students? _____ grade to _____ grade

   b. At what grade level are <u>most</u> of your students currently reading? _____ grade

10. How similar is the language arts curriculum and instruction in your class to that of other teachers at your grade level in your school? (Circle your answer.)

not at all similar		somewhat similar		very similar
1	2	3	4	5

(over)

11. Is there anything else about your language arts class we should know when looking at the assignments and student work?

Thanks so much!

Date assigned: _____

Cover Sheet for Typical Reading Comprehension Assignment

If you need more room to answer the questions, please use the back of this form or attach sheets as necessary.

1. **Describe the assignment below in detail or attach a copy of the assignment directions to this sheet. Be sure to tell us exactly what directions were given to students.**
   Specify the title and type (e.g., poem, novel, textbook, etc.) and grade level of the reading material. If students are working in reading groups, specify which group was given this assignment.

2. **What were your learning goals for the students for <u>this assignment</u>? I.e., what skills, concepts, or facts did you want students to learn as a result of completing this assignment?**

3. **In preparing students for this assignment, how did you accommodate for the range of student needs/skill levels in your classroom?**

4. **How does the assignment fit in with your unit or what you are teaching in your language arts class this month or this year?** Is this an end-of-unit assessment? [ ] Yes [ ] No

5. **How long did students take to complete the assignment?** _____

6. **What type of help, if any, did students receive to complete the assignment?** (Check all that apply.)
   Students received help or formative feedback from [ ] teacher [ ] teacher's aide [ ] other students [ ] parents (e.g., help = substantive revision feedback from teacher or peers). Please explain:

7. How was this assignment assessed? If there is a rubric, student reflection, etc., please attach it. If you are not attaching a rubric, please explain your criteria for grading the work (if graded). Did you share these criteria with students? [ ] Yes [ ] No

8. What criteria did you use to decide which papers are "M" middle papers and which are "H" high? (especially if work was not graded originally or if different from #7 above)

9. Approximately what percent of students performed at the following levels on this assignment:

____% = good - excellent ____% = adequate ____% = not yet adequate

Date assigned: _____

Cover Sheet for Typical Writing Assignment: Final and Rough Drafts

If you need more room to answer the questions, please use the back of this form or attach sheets as necessary.

1. **Describe the assignment below in detail or attach a copy of the assignment directions to this sheet. Be sure to tell us exactly what directions were given to students.**
   If applicable, specify the title and type (e.g., poem, novel, textbook, etc.) and grade level of the reading material.

2. **What were your learning goals for the students for <u>this assignment</u>? I.e., what skills, concepts, or facts did you want students to learn as a result of completing this assignment?**

3. **In preparing students for this assignment, how did you accommodate for the range of student needs/skill levels in your classroom?**

4. **How does the assignment fit in with your unit or what you are teaching in your language arts class this month or this year?** Is this an end-of-unit assessment? [ ] Yes [ ] No

5. **How long did students take to complete the assignment?** _____

6. **What type of help, if any, did students receive to complete the assignment?** (Check all that apply.)
   Students received help or formative feedback from [ ] teacher [ ] teacher's aide [ ] other students [ ] parents (e.g., help = substantive revision feedback from teacher or peers). Please explain:

7. How was this assignment assessed? If there is a rubric, student reflection, etc., please attach it. If you are not attaching a rubric, please explain your criteria for grading the work (if graded). Did you share these criteria with students? [ ] Yes [ ] No

8. What criteria did you use to decide which papers are "M" middle papers and which are "H" high? (especially if work was not graded originally or if different from #7 above)

9. Approximately what percent of students performed at the following levels on this assignment:

____% = good - excellent ____% = adequate ____% = not yet adequate

APPENDIX B

CLASSROOM OBSERVATION PRE-INTERVIEW AND OBSERVATION PROTOCOL

LAAMP Classroom Pre-Observation Interview

Researcher _____ Teacher First Name Teacher Last Name _____ Grade _____

Date _____ School _____ Lang Arts Focus ○ Reading ○ Writing ○ Both

NOTE: If time is limited, go directly to question #3, then ask remaining questions at another time.

1. Briefly describe the students in your class, including those with special needs.

   Are there any LEP students in this class? ○ Yes ○ No

   If so, approximately what percentage of the students are LEP? _____

   What languages other than English do students speak? _____

2. What should I expect to see during the observation?

3. What are your goals or objectives for the lesson I will be observing? What skills, concepts, or facts do you want the students to learn as a result of this lesson? (e.g. 3rd grade: learning the different parts of a story, learning how to peer edit, developing reading strategies such as predicting; 7th grade: learning the structure of a five paragraph essay; increasing vocabulary skills, etc.)

4. How do these goals relate to (or build on) students' prior knowledge of this subject?

5. In this lesson, how will you accommodate for the range of student needs/skill levels in your classroom?

6. Are these goals based on a specific set of standards? ○ Yes ○ No

   If so, which standards (e.g. district standards, state Challenge standards, etc.)?

   How do your goals for this lesson support these standards?

7. Do you have a formal way of assessing what students have learned in this lesson? (e.g. rubrics, etc.)

8. Did you plan the lesson we will be observing with other teachers? ○ Yes ○ No

Notes

Classroom Information

Researcher _____ Teacher First Name Teacher Last Name _____ Grade _____

Date _____ School _____ Lang Arts Focus ○ Reading ○ Writing ○ Both

Total no. of minutes observed _____

TEACHER AND STUDENT INFORMATION

No. of students observed _____ No. of Boys _____ No. of Girls _____

Teacher Ethnicity	**2nd Teacher (in a team-taught classroom)** Ethnicity	**TA** Ethnicity	**2nd TA** Ethnicity
Sex ○ Male ○ Female	Sex ○ Male ○ Female	Sex ○ Male ○ Female	Sex ○ Male ○ Female

Please indicate the number of students who belong to the following ethnic groups. If there is no way to tell, please write "missing data" for this section.

African-Amer. _____ Asian _____ Latino/a _____ White _____ Other _____

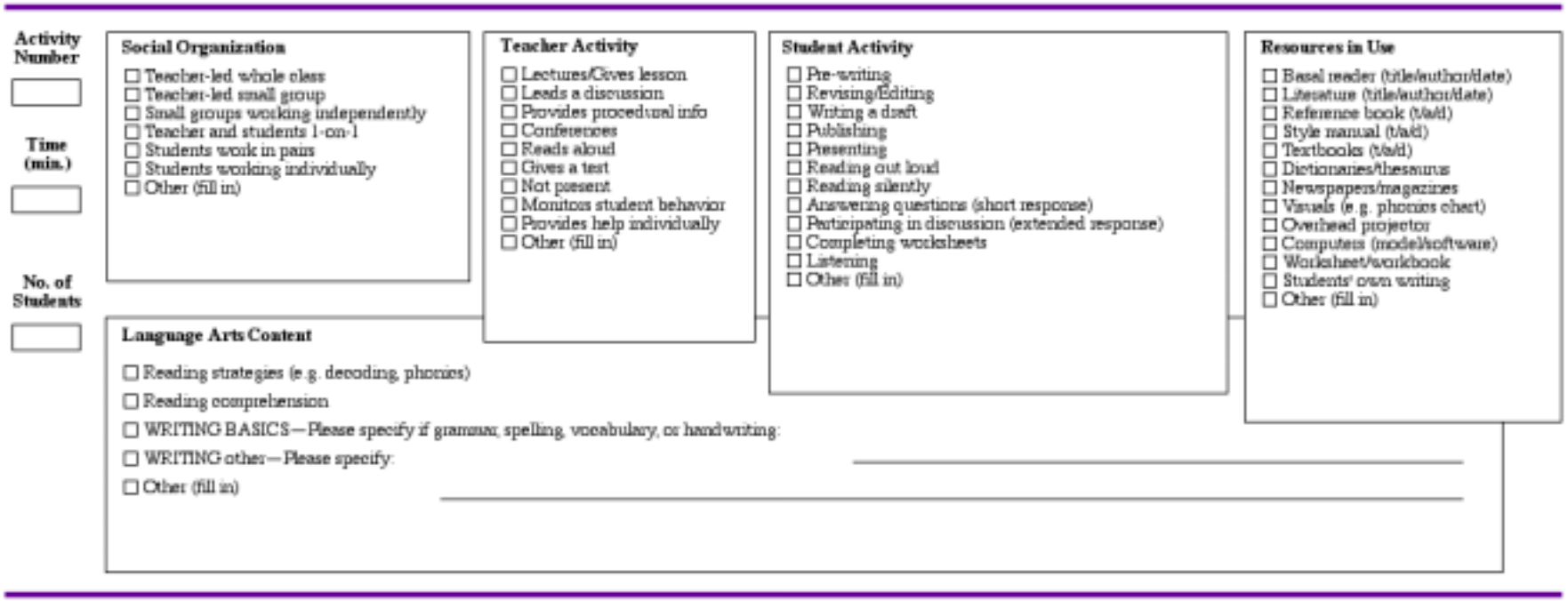
BILINGUAL CLASSROOM INFORMATION (You can mark more than one.)

1. Number of students given instruction in a second language: _____
2. Language(s) used by the *teacher* during the observation: ☐ English ☐ Spanish ☐ Other...
2a. Percentage of time teacher used English: _____
3. Language(s) used by the *TA* during the observation: ☐ English ☐ Spanish ☐ Other...
3a. Percentage of time TA used English: _____
4. Language(s) used by the majority of the *students* during the obs.: ☐ English ☐ Spanish ☐ Other...
4a. Percentage of time students used English: _____

Activity Boxes

Description of Lesson Activities That Occur During the Observation Period

Please code each activity observed and record the number of minutes for each activity. For classrooms in which simultaneous activities are occurring, code the group with the teacher, and the group which has the largest number of students. The remaining activities should be coded together in a third activity box. Label these activities 1a, 1b, 1c; 2a, 2b, 2c, etc. Code teachers giving procedural information as a separate activity if it exceeds 1 minute, and transition periods as separate activities if they exceed 3 minutes.



Lesson Description

Please describe the lesson and sequence of learning activities. Remember to include enough details so that a person who was not present during the lesson could get a clear picture of what the lesson entailed and how it unfolded. Please describe: the different types of groupings (e.g., teacher-fronted, small-group with aide, etc.), the use of different types of instructional materials in these groups (e.g., computers, books, worksheets), and what students were doing in these activities and what for (e.g., participating in a discussion, writing in a workbook, etc.). <u>If computer technology was used please describe:</u> the make/model of computers, how many there were in the classroom, how many students were using them, what the students were doing (e.g., word processing, graphics programs), and purpose for using the computer (e.g., publishing school newspaper, typing out writing assignment, etc.).

Classroom Ratings

1. Stated Goals

In this section rate the clarity and specificity of the teacher's stated goals for the lesson. Do not rate the learning activities.

Clarity and Focus of the Teacher's Goals on Student Learning	○ **1** Goals are not clear in terms of what students are to learn from the assignment. OR all goals may be stated as activities with no definable objective ("activity for activity's sake").	○ **2** Goals are somewhat clear and explicit in terms of what students are to learn from the lesson. OR goals may be a combination of goals and activities with no definable objective.	○ **3** Most goals are clear and explicit in terms of what students are to learn from the lesson. OR some goals may be stated as activities with no definable objective.	○ **4** All the goals are very clear and explicit in terms of what students are to learn from the assignment AND all goals are elaborated and framed in terms of student learning.

Please write a paragraph providing evidence for your rating of the clarity and specificity of the teacher's goals for this lesson. Be as specific as possible and remember to include concrete examples.

2. Learning Activities

In this section rate the degree to which the teacher's stated goals for the lesson were reflected in the design of the learning activities. Specifically rate how well the activity supported achievement of the teacher's goals. Also rate the overall challenge of the learning activities. Additionally, rate the degree to which the lesson was effectively implemented, and the degree of student engagement in the activities.

Alignment Between Goals and Learning Activities (e.g. how well the activity promoted achievement of the teacher's goals)	**O 1** There is very little or no alignment between the teacher's stated learning goals and what students are required to do in the lesson activities. Learning activities do not appear to support the instructional goals.	**O 2** There is some alignment between the teacher's stated goal for the lesson and the lesson activities. The lesson activities appear to somewhat support the instructional goals. OR the goals may be so broadly stated that the task and goals are aligned, but only at a very general level.	**O 3** There is good alignment between the teacher's stated goals and the lesson activities, and the activities appear to support the attainment of the instructional goals.	**O 4** There is exact alignment between the teacher's stated goals and the lesson activities, AND the activities fully support the instructional goals.

Challenge of the Lesson Activities	**O 1** Learning activities involve students in tasks that do not require any degree of complex thinking and/or do not engage students with substantive content material.	**O 2** Learning activities involve students in tasks that require moderately complex thinking. Students' engagement with the material does not promote learning in a rigorous way.	**O 3** At least some of the learning activities require strongly complex thinking as a major focus of the lesson. This level requires the use of higher order cognitive functions, taking students beyond recall, recognition and reproduction of information to evaluation; analysis; synthesis; and production of arguments, ideas, and performances. Students may be asked to synthesize ideas, analyze cause and effect, identify a problem and pose reasonable solutions, hypothesize, speculate giving details or justification, defend opinions or argue a position with evidence, evaluate, analyze (distinguish important or relevant from unimportant or irrelevant information), or determine bias, values, intent.	**O 4** Much or all of the learning activities require strongly complex thinking as a major focus of the lesson. Students also engage in substantive content material. Students may be asked to analyze cause and effect, identify a problem and pose reasonable solutions, speculate giving details or justification, defend opinions or argue a position with evidence to a great extent.

Implementation of the Learning Activities (include classroom management)	○ **1** The learning activity is not effectively implemented (e.g. the class may be disorganized, the teacher may lack control).	○ **2** The learning activity is somewhat effectively implemented.	○ **3** The learning activity is effectively implemented (e.g. transitions are smooth, teacher has control of class).	○ **4** The learning activity is exceptionally well implemented (e.g. transitions are seamless, almost no class time is wasted).

Proportion of Students "On-Task"	○ **1** Less than half of the students appear to be on task.	○ **2** Approximately half the students appear to be on-task.	○ **3** Approximately 85% of the students appear to be on-task.	○ **4** All students are engaged in the activities. Students may also initiate or adapt activities and projects to enhance understanding.

Please provide evidence for your rating of the coherence and perceived level of challenge of the learning activities/lesson.

Describe the rigor and grade-level appropriateness of the activities and resources used (e.g. did the activities or lesson support students' development of HOT, or meaningful content area knowledge).

Comment on reasons for why some students and not others may have appeared to be on task.

Comment as well on the degree to which students appeared to be interested and engaged in the lesson/activities.

Include concrete examples to support your ratings.

Provide the titles and authors of the texts used in the classroom.

3. Classroom Discussion

In this section rate students' opportunity to learn through and engage as partners in meaningful classroom discussions. This includes both the nature of the teacher's questions, as well as the amount of time spent in discussion. This also includes the level of student participation.

Opportunity to Participate in Instructional Conversation	○ **1** A discussion does not take place at all (e.g., the students work individually revising drafts while the teacher fills out paperwork at the front of the class) or interaction between teacher and students is predominantly recitation style, with teacher mediating all questions and answers. The teacher's questions are close-ended, known-answer questions.	○ **2** Teacher makes some attempt to engage students in true discussion, with uneven results. Some of the teacher's questions are open-ended (e.g., "What was remarkable about the story?" What did you like about that character?"). There may be some attempt to have students respond to invitations to comment about a book. Teacher's attempt at engaging students may fail (e.g., she may ask open-ended questions but not wait long enough for a student to answer).	○ **3** Most of teacher's questions are of high quality. Adequate time is available for students to respond and teacher activity solicits student input (e.g., "Tell me why you think that." "Can you tell me a little more about that?"). Teacher builds on student contributions.	○ **4** Classroom interaction represents true discussion. Students initiate topics and make unsolicited, on-topic contributions. Students formulate many questions. Teacher's questions are uniformly high quality, with adequate time for students to respond. Teacher builds on students' contributions, and students build on each other's contributions.

Student Participation in the Classroom Interactions	○ **1** Teacher involves only a few students in the discussion. OR no discussion takes place.	○ **2** Teacher attempts to involve all students in the discussion, but with only limited success.	○ **3** Teacher involves many of the students in the discussion.	○ **4** Teacher involves nearly all students in the discussion.

Please provide justification for your ratings. Remember to include all of the following:

Examples of teacher's questions, student responses, and techniques the teacher used to include students in the discussion.

Whether certain groups of students were participating or not in the classroom discourse (e.g., were these LEP students, etc.).

The degree to which complex language and vocabulary were used by the teacher and modeled for/presented to the students AND the extent to which the teacher facilitated student use of more complex language (e.g. did the teacher explain unfamiliar words, did the teacher rephrase students' questions and statements using more complex language, etc.).

The context and extent of discussion with attention to why discussion may not have been happening (e.g., students were reading silently or reading aloud).

4. Instructional Feedback

In this section rate the opportunity students have to receive information about their performance and progress toward learning goals, and the degree to which this feedback supports learning. Focus on the following components of feedback quality: accuracy, substance, specificity, and helpfulness.

Quality of Instructional Feedback	○ **1** Feedback is either not provided or is of uniformly poor quality. Feedback may be inappropriate (e.g. humiliating, punitive). Feedback does not support instructional goals.	○ **2** Feedback is inconsistent in quality: Some elements of high quality may be present during a small portion of the observation or minimally informative feedback that only supports the instructional goals may be given throughout the observation.	○ **3** Feedback is mostly high quality (e.g. expectations are made explicit to students; students are shown examples of good work). Feedback mostly supports the instructional goals. It is provided either consistently throughout the observation period or in a focused way during a portion of the period.	○ **4** Feedback is uniformly high quality. Provision is made for students to use feedback in their learning. Feedback fully supports the attainment of the instructional goals.

Please provide justification for your rating. Remember to provide concrete examples of the type of feedback the teacher gave to students.

Describe how this feedback did or did not support the instructional goals and the assessment criteria (i.e. how the criteria the teacher had in mind for knowing whether or not the instructional goals had been met by the students).

APPENDIX C

TEACHER ASSIGNMENT SCORING RUBRIC

Language Arts Assignment Rubric

COGNITIVE CHALLENGE **(In judging this dimension, refer to assignment cover sheet item #1, assignment directions that teacher attaches, and student work.)**	**4** Task requires strongly complex thinking as an extensive, major focus of task. Student also engages with substantive content material. E.g., student may be asked to synthesize ideas; analyze cause and effect; identify a problem and pose reasonable solutions; hypothesize; speculate with details or justification; defend opinions or argue a position with evidence; evaluate; analyze (distinguishing important or relevant from unimportant or irrelevant); determine bias, values, intent.	**3** Task requires complex thinking. Student may also engage with substantive content material. E.g., student may be asked to synthesize ideas; analyze cause and effect; identify a problem and pose reasonable solutions; hypothesize; speculate with details or justification; defend opinions or argue a position with evidence; evaluate; analyze (distinguishing important or relevant from unimportant or irrelevant); determine bias, values, intent.	**2** Task requires some moderately complex thinking. Some substantive content area material may be covered. E.g., student may be asked to summarize straightforward information, infer simple main idea, or simply apply the appropriate format for a given genre.	**1** Task does not require any degree of complex thinking and/or does not engage students with substantive content material. E.g., student may be required to recall basic information, or recall definitions. Or student may be asked to answer simple reading comprehension questions or write on a topic with little focus or structure.
FOCUS OF THE GOALS ON STUDENT LEARNING **(Refer to assignment cover sheet item 2 and assignment directions.)**	**4** Goals are very focused on student learning. Goals are very clear and explicit in terms of what students are to learn from the assignment. Additionally, all the goals are elaborated.	**3** Goals are mostly focused on student learning. Goals are mostly clear and explicit in terms of what students are to learn from the assignment.	**2** Goals are somewhat focused on student learning. Goals are somewhat clear and explicit in terms of what students are to learn from the assignment. Goals may be very broadly stated (e.g. reading comprehension). Or there may be a combination of learning goals and activities.	**1** Goals are not focused on student learning; goals are not clear and explicit in terms of what students are to learn from the assignment; OR all goals may be stated as activities with no definable objective (activity for activity s sake).
CLARITY OF THE GRADING CRITERIA **(Refer to assignment cover sheet items 5 & 6 and rubric that teacher attaches.)**	**4** Teacher s grading criteria are very clear, explicit, and elaborated. E.g., teachers rubric or guidelines are detailed and elaborated. Additionally a model of good work may be provided to the students.	**3** Teacher s grading criteria are mostly clear and explicit with regard to what is expected with little or no question. E.g., teacher may use a rubric or a very elaborated and specific list of dimensions.	**2** Teacher s grading criteria are somewhat clear and explicit. Teacher provides some general directions or a rudimentary rubric. E.g., a list of dimensions such as style, creativity, and organization, but some dimensions are undefined or vague.	**1** Teacher does not specify grading criteria, OR it is not possible to determine the grading criteria from the teacher s documents.

ALIGNMENT OF LEARNING GOALS AND TASK **(Refer to assignment cover sheet items 1 & 2 and attached assignment directions.)**	**4** There is exact alignment between teacher s stated learning goals for students on that assignment and what the task asks students to do, AND task fully supports instructional goals. E.g., goal is being able to summarize several points and activity entails summarizing; tasks and goals overlap completely neither one calls for something not included in the other. *Note: This dimension cannot be rated a 4 if the goals are unclear, broadly stated, or stated as activities.*	**3** There is good alignment between teacher s stated learning goals and what the task asks students to do —AND the task supports instructional goals.	**2** There is only some alignment between teacher s stated goals and what the task asks students to do. The task only somewhat supports the instructional goals. E.g., goal is to be able to write an essay, but task calls for completing a concept map and making an outline for an essay (but NOT actually writing an essay). —OR the goal may be so broadly stated that the task and goal are aligned at a very general level.	**1** There is very little or no alignment between teacher s stated goals and what the task asks students to do. The task does not support the instructional goals. E.g., goal calls for writing an essay, but task calls for giving an oral report.
ALIGNMENT OF LEARNING GOALS AND GRADING CRITERIA **(Refer to assignment cover sheet items 5 & 6 and rubric that teacher attaches.)**	**4** There is exact alignment between teacher s stated learning goals for students on that assignment and teacher s stated grading criteria. E.g., goal is to write a persuasive essay, and criteria include appropriate dimensions such as stating a point of view and providing relevant supporting evidence and do not include dimensions not mentioned in goals (e.g., creativity). *Note: This dimension cannot be rated a 4 if the goals are unclear, broadly stated, or stated as activities.*	**3** There is good alignment between teacher s stated learning goals and the stated criteria for grading. E.g., goal is to write a persuasive essay, and criteria include appropriate dimensions but also extraneous ones. Or, fail to include critical dimension (e.g., support for assertions or point of view).	**2** There is only some alignment between teacher s stated learning goals and the stated grading criteria. E.g., goal is to write a business letter, but criteria include <u>mostly</u> extraneous dimensions; e.g., participation in class discussion is given more weight than letter format. Or, criteria given are not very appropriate, e.g., slang is acceptable in a business letter.	**1** There is very little or no alignment between teacher s stated learning goals and the stated grading criteria.
OVERALL TASK QUALITY **(Consider all previous dimensions.)**	**4** Excellent quality in terms of level of cognitive challenge, clarity and application of learning goals, and grading criteria.	**3** Good quality in terms of level of cognitive challenge, clarity and application of learning goals, and grading criteria.	**2** Limited quality in terms of level of cognitive challenge, clarity and application of learning goals, and grading criteria.	**1** Poor quality in terms of level of cognitive challenge, clarity and application of learning goals, and grading criteria.

APPENDIX D

STUDENT WORK RUBRIC

Language Arts Project

Generic Rubric: Reading, Writing, and Literature Analysis

Four-Point

LEVEL	Content	Organization	Style	MUGS
4	A "4" paper fully achieves the purpose of the assignment, clearly expresses the ideas to an intended audience, and conveys a distinct point of view. It *fully* and *elaborately* develops and integrates appropriate ideas with supporting details from the text.	A "4" paper follows the form required. It adheres to the topic and makes logical and explicit connections; is organized in paragraphs; and has a *clear* sense of beginning, middle and end.	A "4" paper contains a quality of uniqueness that enriches its meaning and readability. It exhibits various techniques such as vivid images, descriptive and expressive phrases, variation in sentence patterns, and appropriate tone.	A "4" paper demonstrates superior command of mechanics, usage, grammar, and spelling (MUGS). It is free of errors that interfere with the writer's meaning.
3	A "3" paper develops purpose, audience, and point of view. It *sufficiently* provides as much information *as called for* and develops and integrates appropriate ideas with supporting details from the text. It may contain *a* minor inaccuracy.	A "3" paper follows the form required. It adheres to a topic; makes logical connections among most of the ideas; and has a sense of beginning, middle, and end.	A "3" paper may exhibit techniques such as vivid images, descriptive and expressive phrases, variation in sentence patterns, and appropriate tone.	A "3" paper may contain a number of minor errors, but demonstrates a considerable command of most of the elements of MUGS. It may have an error such as *an* unclear sentence that somewhat interferes with the writer's meaning.
2	A "2" paper contains limited evidence of purpose, audience, and point of view. It may make limited use of the text and may show limited development of that information. It may have obvious factual errors and omissions.	A "2" paper has serious organizational problems: it may not adhere to a topic; may have unclear passages; may make limited connections between ideas; and has a limited sense of beginning, middle, and end. Digressions may significantly interfere with the writer's meaning.	A "2" paper has limited command of the elements of style. It may be mechanical and almost robotic. It typically shows less control of the use of language. There is limited evidence of various techniques such as vivid images, descriptive and expressive phrases, variation in sentence patterns, and appropriate tone.	A "2" paper contains frequent errors that may or may not distract or interfere with the writer's meaning.
1	A "1" paper may contain little or no evidence of purpose, audience, point of view, or a relevant topic. It may contain few or *no* details from the text and show little or *no* development of that information. It may consist mainly of sentences copied from a text. It may have *serious* factual errors and omissions.	A "1" paper has little or no order. It may be a rambling collection of thoughts. It has severe organizational problems: little or no connection among ideas; no sense of beginning, middle, and end; and many digressions.	A "1" paper has little or no command of the elements of style. There is no consistency. It may *only* consist of a string of words conveying little or no meaning.	A "1" paper demonstrates little or no command of MUGS. Errors appear in all, or nearly all, sentences and interfere with the writer's meaning.