Measuring Reading Comprehension and Mathematics Instruction in Urban Middle Schools: A Pilot Study of the Instructional Quality Assessment

CSE Technical Report 681

Lauren Resnick and Lindsay Clare Matsumura
LRDC/University of Pittsburgh

Brian Junker
Carnegie Mellon University

May 2006

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
GSE&IS Bldg., Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

Project 2.3: Indicators of Classroom Practice and Alignment
Lauren Resnick and Lindsay Clare Matsumura, Project Directors, LRDC/University of Pittsburgh;
Brian Junker, Project Director, Carnegie Mellon University

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R305B960002, as administered by the Institute of Education Sciences (IES), U.S. Department of Education.

MEASURING READING COMPREHENSION AND MATHEMATICS INSTRUCTION IN URBAN MIDDLE SCHOOLS: A PILOT STUDY OF THE INSTRUCTIONAL QUALITY ASSESSMENT

Lindsay Clare Matsumura and Sharon Cadman Slater
LRDC/University of Pittsburgh

Brian Junker
Carnegie Mellon University

Maureen Peterson, Melissa Boston, Michael Steele, and Lauren Resnick
LRDC/University of Pittsburgh

Abstract

The quality of reading comprehension and mathematics instruction was explored in five urban middle schools using the Instructional Quality Assessment (IQA) toolkit (*N* = 34 teachers). The IQA is comprised of protocols for rating observed instruction and the quality of teachers' assignments with student work. The purpose of this research was to investigate the reliability and potential validity of the ratings of these data sources. Commensurate with other research on the quality of middle schools, our results indicated that the quality of instruction varied a great deal within schools and was of a 'basic' quality overall. Results indicated a moderate to high level of reliability. Four assignments with student work yielded a stable estimate of quality in both content areas, and when teachers complied with the requirements of the research as few as two observations yielded a stable estimate of teaching quality in both content areas as well. The quality of teachers' observations and assignments were significantly associated in mathematics, but not in reading comprehension. Because of the small sample size it was not possible to apply multi-level models. The relation between the IQA and student achievement on the SAT-10 was explored using linear regression techniques. Results indicated that after controlling for students' prior achievement, socio-economic status (SES), ethnicity, language, and IEP status, the IQA assignment measure in reading comprehension predicted student achievement on the Total Reading, Reading Comprehension, and Vocabulary subscores of the SAT-10. The observation measure in reading comprehension predicted student outcomes on the Reading Comprehension subscore of the SAT-10 only. In mathematics, the quality of teachers' assignments predicted students' achievement on the Procedures subscore of the SAT-10. The quality of observed instruction in mathematics predicted students' achievement on the Procedures and Total Math subscores. Without accounting for clustering within classrooms and schools as multilevel models do, our linear regression analyses may lead to results that appear stronger than they actually are. Nevertheless our analyses indicate the direction of trend in these relationships and raise important questions regarding which data sources may be best (classroom assignments or observations) for measuring specific aspects of instruction and student outcomes. Additional research with larger samples of teachers is needed to make definitive conclusions about the validity of the IQA ratings and under what conditions one might choose to either observe in classrooms or collect assignments with student work.

Introduction

The quality of instruction is at the forefront of public concern and policy decision-making for good reason: Research has shown that instructional quality is the most important *school* factor influencing student achievement (Sanders & Horn, 1994). At the same time, research shows that many teachers do not have the knowledge and skills needed to achieve the goals for learning and instruction set out in many reform programs (National Commission on Teaching and America's Future, 1996; Spillane & Zeuli, 1999). Even within the same school the quality of the learning environments teachers create for students varies dramatically, resulting in different rates of academic growth for demographically and academically similar students depending on which teachers they receive (Rowan, Correnti, & Miller, 2002).

While few would argue the importance of improving instruction to support student learning, it is notable that states and school districts do not yet have systems in place for directly monitoring the quality of teaching. Tremendous investments are made in professional development and in the development and dissemination of new curricula and assessments. And, due in part to the requirements of the No Child Left Behind act of 2001, students' academic progress is carefully tracked. The link between these efforts and student learning—that is, the interactions between teachers and their students—often remains unmeasured and unknown. The end result is that little information is available to policymakers, school officials, and teachers regarding the ways in which schools and classrooms may (or may not) transform over time in response to reform efforts.

Part of the reason why this is the case is because few measures exist that *directly* measure instructional practice on a large-scale. Survey measures—some of which have been linked to student achievement—are limited by the fact that they rely on teachers' self-report of what they do and how they do it. Certain dimensions of practice, such as what topics are covered over the course of a year, may be reliably obtained from a survey (Porter, 2002). However, other dimensions of practice are less reliably measured in this way because individuals can disagree about what specific words (such as "problem-solving" or whether or not a "discussion" was held) mean in practice (Ball & Rowan, 2004; Spillane & Zeuli, 1999). This can severely limit the ability of surveys to adequately measure the extent to which reform-oriented teaching practices are occurring.[1]

Direct measures of instructional quality clearly are needed that can monitor the influence of reform programs on classroom practice on a large scale. Outside of serving as a monitoring tool, however, measures of instructional quality are important and needed because they have the potential to improve instruction, not just monitor it. Shepard (2000) (and others) have argued that classroom assessments can improve students' learning provided that the content and format of these assessments embody whatever thinking and reasoning abilities are the goals of learning, and that information from these assessments is integrated into an ongoing learning process. We believe that similar principles apply in the development of adults' (or teachers') professional knowledge as well. That is, assessments of teachers' practice can play a key role in furthering professional learning and practice, provided these assessments are aligned with standards for instruction, or what research has identified to be the "best practices" within a discipline.

Researchers have long noted that what is tested takes on heightened importance. At the present time, the assessments that have been linked to instructional behavior are high-stakes student achievement tests (Hamilton & Stecher, 2004) which can result in a tendency on the part of teachers to emphasize "test preparation" over more challenging content and activities that are not necessarily represented in standardized tests (e.g., writing extended responses to literature, solving problems with multiple solutions, etc.). Measures of instructional quality have the potential to serve as a counterbalance to this tendency by encouraging teachers to teach to standards for instruction rather than to the specifics of an achievement test.

Of course, as Shepard (2000) has argued, to fully realize their potential as a learning tool, assessments need to be integrated into an ongoing learning process,

---

[1]Spillane and Zueli (1999), for example, observed and interviewed 25 teachers who reported to engage in instructional practices that were commensurate with the state's mathematics reforms and found that only four of these teachers were teaching mathematics in ways that came close to the reform program's intent.

which for teachers would mean a school's or a district's professional learning system. Current theories of effective professional development emphasize communities of practitioners working together to plan and reflect on the concrete tasks of instruction. Assessments of instructional quality, provided they were closely aligned with teachers' everyday work in classrooms, could support these professional development processes and settings by setting forth a common vision of what "good" instruction looks like within a school or district, and by serving as a heuristic for analyzing and planning instruction that achieves this vision.

Development of the IQA

This report presents research focused on developing an assessment of instructional quality, termed the Instructional Quality Assessment (IQA). A number of researchers are also working to develop new ways to measure classroom practice, including collecting a range of artifacts from classrooms (Borko et al, 2005) and using teacher logs to document instructional activities (Camburn & Barnes, 2004). Our project is one such effort. Funded in part through the national Center for Research on Evaluation, Standards, and Student Testing (CRESST), the central goal of our research is to develop a measure of instructional quality that can monitor the quality of instruction with minimal burden on teachers, and that provides information that can be readily incorporated into the professional development system to improve teaching and learning. Specifically, the potential uses of the IQA include being used by:

- researchers or district administrators to monitor instructional quality across schools,
- instructional coaches and principals to help teachers reflect on their practice and plan learning activities, and
- teachers to self-assess their practice.

The IQA is comprised of protocols for rating observed instruction and the quality of teachers' assignments with student work in reading comprehension and mathematics. While ultimately the IQA is intended to be developed in multiple content areas, we chose to focus initially on reading comprehension and mathematics instruction. Reading is central to success in every content area, and is a major focus of federal, state, and school district reform policies and effort. Student under-achievement in mathematics also has received much attention as a result of high-profile studies showing, for example, the relative inferiority of mathematics instruction and learning in the United States compared with other industrialized countries (National Center for Educational Statistics [NCES], 2004).

Teaching is a very complex behavior, and it is not feasible to measure all of the skills needed to teach effectively within every content area. In developing the IQA we made the decision to focus on a relatively focused set of topics and behaviors in order to minimize task-sampling variability, the primary source of measurement error in performance assessment (Shavelson, Baxter, & Gao, 1993). At the same time we wanted the IQA to be able to provide a reasonably "rich" description of instructional practice. Based on our review of a wide range of research on effective instruction, described in a series of previous reports on the IQA (Boston & Wolf, 2005; Junker et al, 2005; Matsumura et al, 2005) we identified instructional behaviors that would likely be most proximal to student achievement—the quality of classroom discussions, the rigor of learning activities, and expectations communicated to students for the quality of their work.[2]

In brief, the IQA focuses on the degree to which teachers foster students' participation in class discussions, build on and extend students' contributions, press students to support their assertions, and ask challenging questions; as well as the degree to which students build on each other's contributions and give evidence for their statements (e.g., Cobb, Boufi, McClain, & Whitenack, 1997; O'Connor & Michaels, 1996; Tharp & Gallimore, 1988). In reading comprehension instruction, the IQA additionally considers the quality of the texts students read, and the degree to which tasks support students to apply higher-level thinking skills and engage with the deeper meaning of a text (e.g., Beck, McKeown, Hamilton, & Kucan, 1997; Snow, 2002). On the other hand, in mathematics, the potential of a task to support students to form connections between mathematical ideas, procedures, and representations, and the degree to which the level of cognitive demand of a task is maintained when implemented, is assessed (Doyle, 1983; Hiebert, et al., 1997; Stein, Smith, Henningsen, & Silver, 2000). Finally, the IQA considers the specificity (clarity) and rigor of the expectations teachers express to students regarding the quality of their

[2] The research described in these NRC reports has been summarized for practitioners in *The Principles of Learning* (Resnick & Hall, 2001). The specific principles of learning we focus on are Accountable Talk, Academic Rigor in a Thinking Curriculum, and Clear Expectations.

work, as well as how teachers communicate their expectations to students (e.g., Black & William, 1998).

We sought data sources that were closely aligned with teachers' everyday work in classrooms in order to minimize teacher burden and maximize the potential utility of the assessment for providing information that could help teachers improve their practice. This led us to concentrate on classroom observations and teachers' assignment augmented with samples of student work.

Observations clearly are the "gold standard" for assessing the quality of classroom discourse and, theoretically at least, require no special preparation or work on the part of teachers. To facilitate comparison, however, teachers are asked in advance to hold a discussion about a text (for reading comprehension), or to engage students in a problem-solving activity and related discussion (in mathematics). Asking teachers in advance to hold a discussion on the day(s) they are observed potentially biases the findings in the sense that a teacher who never holds a discussion could do so for the data collection only. On the other hand, comparing text-based discussions within and across teachers likely would yield more reliable results than trying to compare the quality of a text discussion to the quality of a lesson on short-story writing. And discussions and problem-solving activities are fairly standard instructional behaviors, however varied in quality they may be in practice. The IQA thus provides a measure of the quality of observed classroom discussion, for example, but does not directly provide information on how representative these discussions (or activities) are of everyday classroom practice for a teacher.

To augment the information we gain from the observation, we also collect samples of teachers' assignments along with student work following the methodology established in past CRESST research (Aschbacher, 1999; Clare, 2000; Matsumura et al., 2001). Specifically, we ask teachers to provide us with four assignments and four representative samples of student work. For each assignment, teachers complete a two-page cover sheet describing the context for the assignment, directions they provided to students, and their criteria for determining the quality of students' work. Assignments cover a significantly longer period of instructional time than a single observation because they encapsulate an instructional cycle where teachers communicate an objective (e.g., a skill or set of skills they want students to master), students practice or enact that skill in their work, and teachers then provide feedback to students on their efforts. Collecting multiple assignments from teachers obviously increases the amount of instructional time covered.

Collecting assignment has other advantages as well, namely providing insight on the opportunity students have to produce high-quality written work on their own. As such, they seemed to us to be a good complement to observed instruction that provides potentially more insight on group or whole-class "meaning-making" and problem-solving. The quality of student work provides a further check on the rigor and implementation of lesson activities, and application of classroom assessments. Moreover, assignments with student work are a potentially excellent source of material for collaborative professional development. That is, they are concrete artifacts of practice that could be used by teachers working together and/or with instructional coaches to reflect and plan for instruction by looking both at the quality of students' work and the opportunity students have to produce high quality work. To consider, for example, what the task required of students, what the directions were to students, the criteria for "good" work that was communicated to students, and how future tasks could be structured to improve students' performance.

Again, however, as with the observations we had to decide what sort of assignments to collect from teachers. To minimize variability within the collection, we decided to ask reading comprehension teachers for samples of response to literature tasks, that is, assignments where students are asked to write in response to a text. In mathematics, we asked teachers to provide us with assignments that require students to engage in a problem-solving activity. These are activities that are represented in most standards documents. Additionally, in the most recent version of the IQA we ask that the four assignments teachers prepare for us represent work that they consider to be "challenging" for their students.[3] Asking for challenging assignments, as opposed to "typical" assignments provides teachers with the opportunity to show their (and their students') best work. Additionally, we wanted to be able to compare teachers fairly to one another. It seemed to us that it was more likely that a teacher would give us their best work and call it "typical" than would

---

[3] In past versions of the IQA we also asked for assignments that teachers consider to be 'typical' of their practice. Analyses indicated, however, that asking for either *all* typical or *all* challenging assignments from teachers would lead to a more reliable outcome (Slater, Matsumura, & Junker, 2005).

give us their “typical” work and call it challenging. We wanted to give teachers an equal chance to put their “best foot forward.”

In summary, the IQA in its current form focuses on reading comprehension and mathematics instruction, and includes protocols for observing in classrooms and rating teachers’ assignments with student work. Teachers are asked in advance to hold a class discussion when they are observed, and to provide four ‘challenging’ assignments (response to literature or problem-solving tasks) with representative samples of student work. Observed lessons are assessed with respect to the quality of the discussion, the rigor of the lesson activities and the quality of the expectations teachers communicate to students. The assignments are assessed with respect to their level of academic rigor, and the nature of the expectations teachers hold for the quality of students’ written work.

The Technical Quality of the IQA

A major challenge to developing a measure of instruction is sampling—deciding what should be looked at and the frequency with which these instructional behaviors should be documented (Ball & Rowan, 2004). The following sections describe some of our findings from our most recent pilot study in middle schools with regard to the technical quality of the IQA—specifically, the reliability of the rubrics; how many observations and assignments are needed to yield a stable teacher-level measure of quality; and the relationship of the observations and assessment measures to each other and, potentially, to student achievement. The quality of instruction we observed in middle schools is also described, with samples of classroom transcripts, assignments, and student work to investigate the range of instructional quality captured by the IQA ratings.

As a precursor to the study described below, an earlier pilot study of the IQA was conducted in elementary schools in two similar-sized districts on the east coast. A major focus of that research was to investigate whether the IQA was sensitive to differences in instructional quality between two demographically similar districts—one of which had been involved in a specific reform program while the other had not. A second purpose of that research was to design and evaluate a rater-training program and see if novice raters could be trained to reliably use the IQA.

Some key findings from that study were as follows:

- Raters had only a poor to moderate level of agreement with each other (Junker et al., 2005; Boston & Wolf, 2005; Matsumura et al., 2005).

- Collecting four assignments from teachers yielded a stable estimate of instructional quality at the teacher level in mathematics but not in English language arts (Boston & Wolf, 2005; Matsumura et al., 2005).

- When the data were aggregated across teachers the observation rubrics yielded a stable estimate of quality at the district level and were sensitive to differences in the reform environment of the two districts (Junker et al., 2005). Because teachers were only observed once, it was not possible to determine how many observations might be needed to obtain a stable teacher-level estimate of instructional quality.

During the 2004-05 academic year we continued to investigate the reliability, stability, and feasibility of the IQA, this time in urban middle schools. Following from our previous study, our goals for this year included improving inter-rater agreement (by improving the rater training program and using more experienced raters) and investigating the potential of our observation and assignment quality rubrics for yielding a stable estimate of instructional quality at the teacher level (as opposed to the school or district level). Additionally, we expanded the scope of our work to include exploratory analyses focused on the validity of the IQA ratings. Specifically, we investigated the following questions:

A. How reliable are the classroom observation and assignment rating scales?

B. How many observations or assignments are needed to yield a stable estimate of instructional quality at the teacher level?

C. What is the relation of observed instruction and the quality of classroom assignments?

D. To what extent are ratings of classroom observation and assignments potentially related to student achievement?

E. What do the ratings of the observation and classroom assignments reveal about the quality of students' opportunity to learn in urban middle schools?

Methods

Participants. Sixth- and seventh-grade teachers (*N* = 34) from five middle schools in an urban school district on the east coast participated in the study. Twenty-one teachers taught English language arts and 13 taught mathematics. Their students (*N* = 1334, 53% female) were primarily from low-income families: 71.4% of the students qualified for free lunch and 8.8% were eligible for reduced-price lunch. Nearly half the students were Hispanic (46.4%). The remaining students were 23.2% African American, 18.7% white, 10.1% Asian, and 1.6% Native American. Almost none of the students were classified as Limited English Proficient (0.6%). Of the students in the ELA classes observed, 46.2% of them were categorized as "Basic" on total reading performance, 24.6% were categorized as "Below Basic," 9.7% were categorized as "Proficient" and 1.1% were categorized as "Advanced." For the students in the mathematics classes observed, 43.2% were categorized as "Below Basic" on total mathematics performance, 24.4% were categorized as "Basic," 12.2% were categorized as "Proficient" and 2.6% were categorized as "Advanced." (The percentages do not add up to 100% in either case due to cases labeled as "Not Applicable" in the dataset.) Tables 1 and 2 below contain means and standard deviations of student achievement in ELA and mathematics as measured by the state standardized test score results.

Table 1

Means and Standard Deviations of Student Achievement Scores for English Language Arts

	Mean	SD
Total Reading Score 2005	650.42	36.99
Total Reading Score 2004	653.38	34.20
Reading Comprehension Score 2005	648.12	37.87
Reading Comprehension Score 2004	653.26	35.48
Vocabulary Score 2005	657.57	44.46
Vocabulary Score Prior 2004	654.74	40.69

Table 2

Means and Standard Deviations of Student Achievement Scores for Mathematics

	Mean	SD
Total Math Score 2005	654.04*	40.83
Total Math Score 2004	649.39*	47.68
Problem Solving Score 2005	654.04*	46.91
Problem Solving Score 2004	648.05*	46.35
Procedures Score 2005	655.17	47.59
Procedures Score 2004	652.01	57.58

*$p < 0.01$ for mean difference between current and prior year scores

Procedures and measures. A member of the IQA research team contacted the principals of each of the nine middle schools in the district. Members of the research team visited each of the five schools that agreed to participate to discuss the study with interested teachers, to schedule the observations, and to distribute the assignment collection materials.

Classroom observation data were collected over a two-week time period. Each of the 34 teachers was observed on two consecutive days for the same class period by the same rater resulting in 38 English language arts observations and 26 mathematics observations. Because of scheduling conflicts, four of the English language arts teachers (out of the 21) were only observed once. Observations were conducted by members of the IQA research team and a graduate student recruited and trained for the data collection. Inter-rater agreement was assessed in non-sample classrooms prior to the study with each possible rater pair for each content area observing two consecutive lessons.[4]

Twenty-five of the 34 teachers (16 English language arts, 9 mathematics) also submitted four assignments they considered to be challenging for their students for a total of 99 assignments (one teacher only submitted three assignments). For each assignment teachers filled out a two-page cover sheet describing the instructional context for the assignment task, the directions they provided to students, their assessment criteria for grading student work, and how they shared these criteria with students. Teachers also submitted four samples of student work for each task—two samples of work they considered to be of high and medium quality respectively. Teachers were given $100 gift certificates as a token of appreciation for completing the assignment materials.

Four-point scales (1 = poor, 4 = exemplary) were used to assess the quality of observed instruction and assignments. The quality of the classroom discussions (i.e., evidence of Accountable Talk) and the expectations for learning teachers communicated to their students (i.e., evidence of Clear Expectations) were assessed on similar dimensions in both content areas (reading comprehension and mathematics). These are summarized below.

[4] This decision was made so that we could focus our resources on conducting multiple observations of teachers in our study sample using one rater, rather than sending two raters in to each classroom for a single observation as we had done in previous years.

Class Discussions (Accountable Talk)

- **Student participation in the discussion:** The percent of students who participate in the class discussion.

- **Teacher links student contributions to each other:** The degree to which a teacher makes a consistent effort to ensure that all participants understand the ideas and positions shared during the whole-group discussion, and links students' contributions to one another so that the discussion builds ideas within the learning community.

- **Students link to each other's contributions:** The degree to which students make consistent efforts to ensure that all participants understand the ideas and positions shared during the whole-group discussion, and make efforts to link contributions to one another so that the discussion builds ideas within the learning community.

- **Teacher presses for accurate knowledge and rigorous thinking:** The degree to which teachers press students to provide specific and accurate knowledge as evidence to back up their contributions and explain their reasoning.

- **Students provide accurate knowledge and rigorous thinking***:* The degree to which students provide specific and accurate knowledge as evidence to back up their contributions and explain their reasoning.

Expectations for Learning (Clear Expectations)

- **Clarity and detail of the expectations for student learning***:* The specificity and elaborateness of the teacher's expectations for the quality of students' work or participation.

- **Rigor of the expectations for student learning**: The degree to which the teacher's expectations focus on students applying higher-level thinking skills during the lesson activity and discussion.

- **Student access to expectations***:* The extent to which the teacher ensures that all students have access to these expectations.

The dimensions used to assess the level of academic rigor demonstrated in a lesson or in an assignment, in contrast, were different for each content area and were based in research describing "best practices" for instruction within. These dimensions for reading comprehension and mathematics are summarized below: (See Appendix A for the complete set of rubrics in each content area).

Rigor of Reading Comprehension Lessons (Academic Rigor)

- **Rigor of the text:**[5] The degree to which a text that is discussed during a reading comprehension lesson contains literary or informational content that is complex and engaging enough to warrant extended discussion. Additionally, this dimension measures the richness and variety of the language (vocabulary and sentence structures) in the text.

- **Analyzing and interpreting the text through discussion:** The degree to which the whole-group discussion assists students in deepening their comprehension of a text, as opposed to recalling, describing, or identifying basic information.

- **Analyzing and interpreting the text through lesson activities:** The degree to which lesson activities support students in deepening their comprehension of a text. Additionally, this dimension measures the extent to which students were given the opportunity to develop and elaborate their ideas, including providing evidence for their positions.

Rigor of Mathematics Lessons (Academic Rigor)

- **Potential of the task**: The potential of a task given to students to build deep conceptual understanding in mathematics (i.e., "doing mathematics" or developing "procedures with connections"), as opposed to recalling memorized facts or practicing application of procedures.

- **Active use of knowledge: Implementation of the task:** The extent to which the high-level demands of a task were maintained and achieved as the task was implemented during the lesson.

- **Academic rigor of student discussion following task:** The extent to which students show their work and explain their thinking about the important mathematical content during the discussion.

- **Academic rigor in the teacher's expectations:** The degree to which the teacher's expectations for quality work focus on content and skills central to developing conceptual understanding in mathematics, as opposed to expectations that focus on memorization, application of procedures, or non-mathematical criteria such as neatness.

- **Analyses***:* Descriptive statistics were used to characterize observed instruction and assignment quality. Reliability of the instruments was assessed using Cronbach's alpha and intraclass correlations (ICCs).

[5] This dimension was assessed using a three-point scale.

ANOVA techniques were used to investigate the relative variation in instructional quality between and across schools, and correlations were computed to investigate the relation of the data sources (assignments and observed lessons). Generalizability studies were conducted to investigate whether our design (observing twice and collecting four assignments from teachers) yielded a stable estimate of quality, and decision studies were conducted to explore options for future research design. The relationship between IQA observation and assignment scores and student achievement was investigated using multiple regression techniques

Findings

Reliability of the IQA Rubrics

Results indicated that the IQA measure has good reliability overall in both reading comprehension and mathematics. The overall exact scale-point agreement for the observation rubrics was 86.4% in English language arts and 81.8% in mathematics. Exact point rater agreement for the individual rubrics ranged from moderate (around 70%) to excellent (100%). The exception to this was the rubric measuring the clarity of a teacher's expectations. This rubric had poor interrater agreement in both English language arts (50.0%) and mathematics (42.9%) (see Tables 3 and 4).

The level of inter-rater agreement was somewhat lower for the assignment measure. Exact scale-point agreement averaged across pairs of raters was moderate, 71.3% in English language arts and 76.3% in mathematics (see Tables 5 and 6).

Cronbach's alphas were calculated to investigate the consistency of the rating scales. Results yielded an alpha of .92 for the IQA observation measure in mathematics (an excellent level of consistency) and .76 for the ELA observation measure. Cronbach's alphas were acceptable at .77 and .76 for the ELA and mathematics assignments respectively.

Table 3

Inter-rater reliability of observation ratings for the English language arts lessons (*n* = 3 observers, 4 observations)

	% Agreement	ICC
Overall	86.4	0.96
Accountable Talk:		
Student participation in the discussion	75.0	0.80
Teacher links student contributions to each other	75.0	0.84
Students link to each other's contributions	100.0	1.0
Teacher presses for evidence or for students to explain thinking	100.0	1.0
Students give evidence or explain their thinking	75.0	0.57
Clear Expectations:		
Clarity and detail of expectations	50.0	0.89
Rigor of expectations	75.0	0.80
Student access to expectations	100.0	1.0
Academic Rigor:		
Quality of the text read by students	100.0	1.0
Opportunity for students to analyze and interpret a text in the discussion	100.0	1.0
Opportunity for students to analyze and interpret a text in the lesson activity	100.0	1.0

Table 4

Inter-rater reliability of observation ratings for the mathematics lessons (n = 4 raters, 7 observations)

	% Agreement	ICC
Overall	81.8	.98
Accountable Talk:		
Student participation in the discussion	85.7	0.99
Teacher links student contributions to each other	100.0	1.0
Students link to each other's contributions	100.0	1.0
Teacher presses for evidence or for students to explain thinking	85.7	0.99
Students give evidence or explain their thinking	71.4	0.99
Clear Expectations:		
Clarity and detail of expectations	42.9	0.11
Rigor of expectations	71.4	0.98
Student access to expectations	100.0	1.0
Potential of the task	71.4	0.75
Implementation of the task	100.0	1.0
Student discussion of mathematical concepts following the task	71.4	0.99

Table 5

Inter-rater reliability of assignment ratings for English language arts (n = 3 Raters, 64 Assignments)

	% Agreement	ICC
Overall	71.3	0.93
Academic Rigor:		
Rigor of the text read by students	93.0	0.93
Rigor of the task	72.9	0.78
Rigor in teacher's expectations	60.9	0.80
Clear Expectations:		
Clarity and detail of expectations	64.1	0.79
Student access to expectations	74.5	0.81

Table 6

Inter-rater reliability of assignment ratings for mathematics (n = 3 Raters, 35 Assignments)

	% Agreement	ICC
Overall	76.3	0.88
Academic Rigor:		
Rigor of the task	72.9	0.51
Overall engagement in the task	79.0	0.72
Rigor in students' responses	62.9	0.67
Rigor in teacher's expectations	75.2	0.63
Clear Expectations:		
Clarity and detail of expectations	82.9	0.77
Student access to expectations	84.8	0.90

Number of Observations and Assignments Needed to Yield a Stable Teacher-Level Measure of Instructional Quality

Generalizability and decision studies were conducted to explore the number of observations and assignments needed for a reliable measure of instructional quality. Our results indicated that as few as two observations yielded a stable estimate of quality, *when teachers complied with the requirements of the data collection* ($\hat{\phi}$ = .80 and .86 for reading comprehension and math respectively)[6] (see Tables 7 and 8 below). As described earlier, teachers agreed in advance to hold class discussions on each of the two days we visited. Four teachers (two in each content area), however, did not comply with the data collection requirements on one of these two days. One reading comprehension teacher engaged students in a writer's workshop the entire class period, the other teacher had students work independently throughout the class period (e.g., listening to books on tape and following along with the text, etc.).

[6] Data were analyzed using the mGENOVA software program (Brennan, 2001a, 2001b). For the analysis of observation data, a random-effects Teacher x Observation design was used, raters never overlapped in their observations so rater was not included as a facet in this design. For the analysis of assignment data, a random-effects Teacher x Rater x Assignment analysis of variance was performed, and variance components were estimated. Variance component estimates and dependability coefficients are presented for each analysis. Dependability coefficients for *absolute* decisions are reported throughout the paper (i.e., phi-coefficients), which describe absolute level of performance, rather than generalizability coefficients for relative decisions intended for rank ordering and norm-referenced comparisons. To aid in interpretation of the variance component estimates, percent of total variability accounted for by each variance component is presented.

Similarly, one math teacher had students take a test for one entire class period; another math teacher had students give presentations. These four teachers received zero scores (meaning behavior not observed) on most of the IQA rubrics (e.g., quality of the classroom conversation, rigor of the lesson activity, etc.) on those days. When their scores were included in the analyses the number of observations needed to obtain a stable estimate of quality increased considerably.

Table 7

Decision study for the reading comprehension observation eata [*t xo* Design]: Two noncompliant teachers removed from dataset ($n = 15$)

Source of Variation	G Study	Alternative D Studies		
n_o =	2	3	4	5
Teacher (*t*)	41.2 (67%)	41.2 (86%)	41.2 (89%)	41.2 (91%)
Observation (*o*)*	0	0	0	0
Residual (*to,e*)	19.9 (33%)	6.66 (14%)	5.0 (11%)	4.0 (9%)
Absolute Error SD	3.16	2.58	2.24	2.00
Dependability Coefficient	0.80	0.86	0.89	0.91

*Negative variance components set to zero.

Table 8

D-Study for the Mathematics Observation Data [*t xo* Design]: Two Noncompliant Teachers Removed from Dataset ($n_t = 11$)

Source of Variation	G Study	Alternative D Studies		
n_o =	2	3	4	5
Teacher (t)	63.2 (75%)	63.2 (90%)	63.2 (92%)	63.2 (94%)
Observation (o)*	0	0	0	0
Residual (to,e)	21.4 (25%)	7.2 (10%)	5.4 (8%)	4.3 (6%)
Absolute Error SD	3.27	2.67	2.31	2.07
Dependability Coefficient	0.86	0.90	0.92	0.94

*Negative variance components set to zero.

Collecting four assignments from teachers yielded a stable estimate of quality in both content areas. Specifically, results for reading comprehension indicated a dependability coefficient of 0.82 for four assignments per teacher rated by three raters. Decision studies estimated that reducing the number of raters to two did not effect the dependability coefficient much ($\hat{\phi}$ = .81) and that a design of collecting only three assignments per teacher is likely to be sufficient, with dependability being just below 0.80 ($\hat{\phi}$ = .78) (see Table 9). Results for mathematics assignments were similar. This analysis yielded a dependability coefficient of 0.80 for four assignments per teacher rated by three raters. Decision studies estimated that reducing the number of raters to two only minimally reduced the dependability coefficient ($\hat{\phi}$ = .77). Collecting only three assignments per teacher may not be sufficient, however ($\hat{\phi}$ = .74) (see Table 10).

Table 9

D-Study for the English Language Arts Assignment Data [*t x r x a* Design] (N_t = 16)

Source of Variation	G Study	Alternative D Studies			
n_r =	3	2	2	2	3
n_a =	4	4	3	2	3
Teacher (t)	.385 (50.6%)	.385 (81.2%)	.385 (77%)	.385 (70%)	.385 (78.6%)
Rater (r)*	0	0	0	0	0
Assignment (a)	0	0	0	0	0
tr	.021 (2.8%)	.011 (2.3%)	.011 (2.2%)	.011 (2%)	.007 (1.5%)
ta	.273 (35.9%)	.068 (14.4%)	.091 (18.2%)	.136 (24.7%)	.091(18.5%)
ra	.0002 (<.01%)	.0002 (.04%)	0	0	0
Residual (tra,e)	.081 (10.7%)	.010 (2.1%)	.014 (2.8%)	.020 (3.7%)	.009 (1.8%)
Absolute Error SD	.28	.30	.34	.41	.33
Generalizability Coefficient	.82	.81	.77	.70	.78

*Negative variance components set to zero.

Table 10

D-Study for the Mathematics Assignment Data [*t x r x a* Design] ($N_t = 9$)

Source of Variation	G Study	Alternative D Studies			
n_r =	3	2	2	2	3
n_a =	4	4	3	2	3
Teacher (t)	.095 (36.1%)	.095 (76.4%)	.095 (70.7%)	.095 (61.8%)	.095 (74.1%)
Rater (r)*	.001 (0.4%)	.0004 (0.3%)	.0004 (0.3%)	.0004 (0.3%)	.0003 (0.2%)
Assignment (a)	.002 (0.8%)	.0005 (0.4%)	.0006 (0.5%)	.0009 (0.6%)	.0006 (0.5%)
tr	0	0	0	0	0
ta	.064 (24.3%)	.016 (12.9%)	.021 (15.9%)	.032 (20.9%)	.021 (16.6%)
ra	0	0	0	0	0
Residual (tra,e)	.102 (38.8%)	.013 (10.5%)	.017 (12.6%)	.025 (16.6%)	.011 (8.8%)
Absolute Error SD	.16	.17	.20	.24	.18
Generalizability Coefficient	.80	.77	.71	.61	.74

*Negative variance components set to zero.

Relationship of the IQA Ratings of Observed Instruction and Assignment Quality

In mathematics, the overall average quality of the assignments teachers provided and the overall average quality of their observed lessons was significantly associated ($r = .66$, $p < .001$).[7] More specifically, the set of variables measuring the rigor of the assignment tasks was associated with the set of variables measuring the quality of the class discussion in the observed lessons ($r = .68$, $p < .001$). The rigor of the assignment tasks also was significantly, but less strongly associated with the rigor of the observed lessons ($r = .38$, $p < .001$). The clarity and communication of the expectations teachers held for students as measured in the assignment collection was associated with the quality of the class discussion ($r = .63$, $p < .001$) and the clarity of the expectations teachers communicated to students in the observed lessons ($r = .45$, $p < .001$).

In reading comprehension the results were quite different. The overall quality of the assignments teachers gave to students and the overall quality of their

[7] To help mitigate the probability of making a Type 1 error—only effects significant at the .01 level or less are considered significant.

observed lessons was not associated (r = .09). Looking more closely at the results, it appears that the set of variables measuring the clarity and communication of the expectations for student learning in the assignment collection was weakly associated with the same set of variables in the observed lessons (r = .17, p < .01). The variables measuring the rigor of the assignment tasks was also weakly associated with the clarity and communication of the expectations for student learning in the observed lessons (r = .27, p < .001). The rigor of the assignments we collected was not associated, however, with the quality of the classroom discussions or the rigor of the lesson activities we observed.

Relationship of the IQA Ratings of Observed Instruction and Assignment Quality to Student Achievement

The bottom line for teacher practice is its effect on student achievement. Unfortunately, our sample size was too small (21 reading comprehension teachers and 13 mathematics teachers) to apply multi-level models effectively to investigate the potential relationship between IQA ratings and student achievement (as measured on the SAT-10). Instead, we explored this relationship by using linear regression techniques. Specifically, we investigated several ordinary least squares multiple regression models using the following subscale scores of the SAT-10 as the dependent variables: Total Reading, Reading Comprehension, and Vocabulary (in English language arts); and Total Math, Procedures, and Problem-Solving (in mathematics). Ratings of instructional quality (observed instruction or assignment quality), student demographic variables, and prior year achievement were included as the independent variables in the models.[8]

Without accounting for clustering within classrooms and schools as multilevel models do, our linear regression analyses may lead to results that appear stronger than they actually are. Nevertheless our analyses indicate the direction of trend in these relationships, and statistically insignificant results in our analysis would be insignificant in multilevel analyses as well.

[8] Scatterplots of the variables suggested that linear models would be appropriate. Examination of correlations among all interval level predictors provided assurance that variables did not exhibit collinearity. As expected, current and prior year test scores are highly correlated, as well as the different subscores. However, none of the independent variables *within* a given model correlate greater than .70. Further, all of the variance inflation factors (VIFs) for predictors in the various models were less than 2, further evidence against violation of the collinearity assumption.

Observation ratings and student achievement. After adjusting for students' background and prior achievement, and eliminating the teachers from the analyses who had not complied with the data collection requirements, the IQA classroom observation measure of reading comprehension was a significant predictor of students' achievement on the Reading Comprehension subscale of the SAT-10 (standardized β =.088, p = .046) (see Table 11).

Table 11

Regression results for SAT-10 reading comprehension subscale score (IQA Observation Measure) (R^2 = .52)

Variables	Regression Coefficients		Significance
	Beta	Standardized Beta	
Comprehension Score Previous	.694	.631	.000
Sex	4.87	.069	.129
Free / Reduced-Price Lunch	-12.63	-.141	.002
IEP Status	-1.61	-.013	.780
Ethnicity	-8.43	-.088	.052
IQA Observation Score	4.49	.088	.046

In mathematics, the IQA classroom observation ratings significantly predicted student achievement on the Total Math subscale (standardized β = .163, p = .000), and the Procedures subscale (standardized β = .322, p = .000) (see Tables 14 and 15).

Table 12

Regression results for SAT-10 total math subscale score (IQA Observation Measure) ($R^2 = .77$)

Variables	Regression Coefficients		Significance
	Beta	Standardized Beta	
Total Math Score Previous	.69	.83	.000
Sex	7.55	.092	.009
Free / Reduced-Price Lunch	-6.27	-.065	.119
IEP Status	3.06	.020	.571
Ethnicity	-6.23	-.065	.094
IQA Observation Score	9.81	.163	.000

Table 13

Regression results for SAT-10 procedures subscale score (IQA Observation Measure) ($R^2 = .42$)

Variables	Regression Coefficients		Significance
	Beta	Standardized Beta	
Procedures Score Previous	.476	.584	.000
Sex	3.59	.037	.497
Free / Reduced-Price Lunch	-3.21	-.028	.659
IEP Status	6.44	.036	.515
Ethnicity	-8.87	-.079	.193
IQA Observation Score	22.75	.322	.000

Assignments and student achievement. After adjusting for students' background and prior achievement, the IQA rating of assignments was a positive and significant predictor for all of the reading comprehension outcome scores: Total Reading (standardized $\beta = .107$, $p = .008$), Reading Comprehension (standardized $\beta = .104$, $p = .028$) and Vocabulary (standardized $\beta = .185$, $p = .000$) (see Tables 14 through 16)

Table 14

Regression results for SAT-10 total reading subscale score (IQA Assignment Measure) (R^2 = .65)

Variables	Regression Coefficients		Significance
	Beta	Standardized Beta	
Total Reading Score Previous	.794	.713	.000
Sex	1.78	.026	.493
Free / Reduced-Price Lunch	-5.914	-.068	.086
IEP Status	-4.55	-.038	.336
Ethnicity	-5.254	-.057	.143
IQA Assignment Score	4.907	.107	.008

Table 15

Regression results for SAT-10 total reading comprehension subscale score (IQA Assignment Measure) (R^2 = .53)

Variables	Regression Coefficients		Significance
	Beta	Standardized Beta	
Total Reading Comprehension Score Previous	.659	.599	.000
Sex	5.52	.078	.083
Free / Reduced-Price Lunch	-12.29	-.138	.003
IEP Status	-2.75	-.022	.635
Ethnicity	-8.29	-.087	.055
IQA Assignment Score	4.906	.104	.028

Table 16

Regression Results for SAT-10 Vocabulary Subscale Score (IQA Assignment Measure) (R^2 = .54)

Variables	Regression Coefficients		Significance
	Beta	Standardized Beta	
Total Vocabulary Score Previous	.676	.623	.000
Sex	-2.67	.033	.448
Free / Reduced-Price Lunch	-1.41	-.014	.764
IEP Status	-14.75	-.104	.019
Ethnicity	-3.81	-.035	.434
IQA Assignment Score	9.924	.185	.000

In mathematics, IQA assignment score was a significant predictor of student achievement on the Procedures subscale only (standardized β = .13, p = .034) (see Table 17).

Table 17

Regression Results for SAT-10 Procedures Subscale Score and IQA Assignment Measure (R^2 = .52)

Variables	Regression Coefficients		Significance
	Beta	Standardized Beta	
Procedures Score Previous	.448	.62	.000
Sex	2.72	.029	.622
Free / Reduced-Price Lunch	-1.94	-.018	.809
IEP Status	-5.307	-.029	.628
Ethnicity	-10.29	-.099	.144
IQA Observation Score	16.03	.130	.034

In short, both the observation measure and the assignment measure showed some association with student learning gains. The reading comprehension assignment measure seemed to be more sensitive to instructional behaviors associated with a wider range of student outcomes. In mathematics, the opposite appeared to be true.

To further explore the validity of our ratings, we returned to our field notes and the artifacts we collected (assignments and the student work) from teachers to take a closer look at the degree to which the IQA rating scales appeared to capture meaningful differences in students' learning environments. To conserve space in the body of this report examples of higher- and lower-quality lessons are described in Appendix B. In summary, however, the analyses of the field notes and artifacts support the statistical results. Lessons and assignments that received differential ratings demonstrated clear differences in students' learning opportunities.

Instruction in Urban Middle Schools

As shown in Tables 18 and 19, the IQA ratings of observed instruction demonstrated a wide variation in instructional quality.[9] On average, however, the quality of the quality of the discussions we observed in most classrooms was of basic quality overall. Teachers frequently did not build on, extend student contributions, or press students to explain their reasoning or give evidence for their assertions, and this was reflected in brief, surface-level student responses. The expectations for learning that teachers communicated to students also did not focus on high-level demands in the majority of classrooms. For example, in two-thirds of the math classes we observed the teachers' expectations did not focus on complex thinking and/or exploring and understanding important mathematical concepts and relationships. Even fewer reading comprehension teachers held expectations that focused on students using high-level skills, such as interpreting a text, inferring major themes, analyzing character motives and relationships, or comparing themes and characters across texts.

Similarly in mathematics, only a few teachers engaged students with tasks that had the true potential to engage students in exploring and understanding the nature of mathematical concepts, procedures, and/or relationships such as using complex and non-algorithmic thinking or applying a broad general procedure and connecting that to underlying mathematical concepts. The implementation of mathematics tasks in most classrooms--even when the task was of a higher quality—also tended to be poor.

While most of the reading comprehension teachers assigned students to read were high quality texts, far fewer of the lessons engaged students in rich discussions

[9] Interestingly, the variation in instructional quality was far greater within schools than between schools. In fact, we found only one school where the quality of the assignments teachers submitted was significantly higher than the other schools, further supporting our focus on developing a stable and reliable teacher-level measure of quality.

of those texts. Teachers mostly asked questions that guided students to recall surface-level information about what they had read, and students mostly gave short, perfunctory answers to these questions. Very few lesson activities supported students to engage with some underlying meanings or nuances of a text, even with limited use of examples.

Finally, in nearly half of the lessons we observed (40.6% reading comprehension and 42.3% mathematics) students were provided with procedural directions only for engaging in the lesson task. In other words, students were given basic information about what to do (e.g., "Write a paragraph in your journal answering the following question; complete problems #3 on page 27 with your group"), but were not provided with information about what they would need to do, or include in their work, to be successful on the task. There was little to no discussion of what high-quality student work would look like, or how students would know if they were doing a good job.

Table 18

Quality of observed instruction for English language arts (*N* = 21 teachers, 38 observations)

	Mean	(SD)	Range	0	1	2	3	4
Accountable Talk:								
Student participation in the discussion	2.00	1.47	0-4	26.3	7.9	23.7	23.7	18.4
Teacher links student contributions to each other	2.0	1.47	0-4	28.9	18.4	28.9	13.2	10.5
Students link to each other's contributions	1.11	1.16	0-4	28.9	52.6	7.9	0	10.5
Teacher presses for evidence or for students to explain thinking	1.55	1.29	0-4	28.9	18.4	28.9	15.8	7.9
Students give evidence or explain their thinking	1.42	1.24	0-4	28.9	26.3	26.3	10.5	7.9
Clear Expectations:								
Clarity and detail of expectations	1.88	1.16	0-4	6.3	40.6	25.0	15.6	12.5
Rigor of expectations	1.35	1.11	0-4	19.4	45.2	25.8	0	9.7
Student access to expectations	2.03	1.64	0-4	18.8	34.4	9.4	0	37.5
Academic Rigor:								
Quality of the text read by students	2.28	0.92	0-3	3.4	20.7	20.7	55.2	N/A
Opportunity for students to analyze and interpret a text in the discussion	1.41	1.26	0-4	37.8	5.4	40.5	10.8	5.4
Opportunity for students to analyze and interpret a text in the lesson activity	1.45	0.91	0-3	15.2	36.4	36.4	12.1	0

Table 19

Quality of observed instruction in mathematics (*N* = 13 teachers, 26 observations)

	Mean	(SD)	Range	0	1	2	3	4
Accountable Talk:								
Student participation in the discussion	2.57	1.04	0-4	8.7	0	30.4	47.8	13.0
Teacher links student contributions to each other	1.54	1.10	0-4	19.2	26.9	42.3	3.8	7.7
Students link to each other's contributions	1.08	.85	0-4	19.2	61.5	15.4	0	3.8
Teacher presses for evidence or for students to explain thinking	1.81	1.27	0-4	19.2	23.1	23.1	26.9	7.7
Students give evidence or explain their thinking	1.73	1.25	0-4	19.2	26.9	23.1	23.1	7.7
Clear Expectations:								
Clarity and detail of expectations	1.81	1.27	0-4	11.5	42.3	11.5	23.1	11.5
Rigor of expectations	2.08	1.16	0-4	7.7	23.1	38.5	15.4	15.4
Student access to expectations	2.65	1.57	0-4	11.5	23.1	3.8	11.5	50.0
Academic Rigor:								
Potential of the task	2.46	.91	0-4	3.8	0	57.7	23.1	15.4
Implementation of the task	2.28	.74	0-4	4.0	0	64.0	28.0	4.0
Student discussion of mathematical concepts following the task	1.65	1.38	0-4	23.1	30.8	19.2	11.5	15.4

Classroom assignments and student work. As shown in Tables 20 and 21, the quality of the mathematics assignments we collected from teachers appeared to be higher than the English language arts (reading comprehension) assignments. It is possible that this might be because teachers were using more prescribed curricula in mathematics, though with such a small sample size, it is not possible to draw conclusions about potential differences between these content areas in the quality of instruction for these schools or for the district as a whole.

Similar to what was observed in the reading comprehension lessons, while the majority of the texts assigned to students to read received the highest rating (a '3' for this dimension), students' opportunity to write meaningfully about these texts was much more constrained. More than a third of the assignments (37.5%) engaged students in basic recall of isolated facts from the text or writing about the text in a surface-level way (e.g., writing about whether or not they liked a book without providing evidence from the text, or applying any type of external criteria). Nearly another third of the assignment tasks engaged students only in a surface-level summary of a story—for example, retelling the basic events in a myth with little detail or any level of analysis (e.g., discussing what message the myth is intended to convey, how it compares to a myth from another culture, etc.). The grading criteria also similarly focused on low-level skills—in many cases the mechanics of students' writing rather than the content.

In math, as described earlier, the quality was higher. Nearly 80% of the tasks were of a high quality (rated a 3 or a 4), which is not surprising given the fact that the district is using an NCTM curriculum. The implementation of the tasks was somewhat lower, however. The expectations communicated to students (in the form of rubrics, etc.) for the quality of their written work also received lower scores overall. For example, in slightly more than one third of the assignments the criteria teachers used to assess students' work focused on the correctness of a single problem-solving strategy, rather than on students' ability to make a meaningful link between representations, etc..

Table 20

Quality of classroom assignments for English language arts (*N* = 16 teachers, 64 assignments)

	M	(SD)	Range	0	1	2	3	4
Academic Rigor:								
Quality of the text read by students*	2.53	.85	0-3	2.8	13.9	11.1	72.2	N/A
Rigor of the task	1.80	.91	0-4	4.7	37.5	32.8	23.4	1.6
Academic rigor of the teacher's expectations	1.90	1.29	0-4	17.5	20.6	30.2	17.5	14.3
Clear Expectations:								
Clarity and detail of expectations	2.29	1.02	0-4	3.2	15.9	46.0	19.0	15.9
Communication of expectations	2.42	1.22	0-4	3.8	26.4	17.0	30.2	22.6

* This dimension is assessed on a 0 to 3 point scale.

Table 21

Quality of classroom assignments in mathematics (*N* = 9 teachers; 35 assignments*)

	Mean	(SD)	Range	0	1	2	3	4
Academic Rigor:								
Potential of the task	3.06	.78	1-4	0	2.9	17.6	50.0	29.4
Implementation of the task	2.79	.77	1-4	0	2.9	32.4	47.1	17.6
Rigor in students' response to the task	3.03	.86	1-4	0	3.1	25.0	37.5	34.4
Rigor of teacher's expectations	3.03	.78	1-4	0	5.9	11.8	55.9	26.5
Clear Expectations:								
Clarity and detail of expectations	2.69	.86	1-4	0	6.3	37.5	37.5	18.8
Communication of expectations	2.61	.84	1-4	0	6.5	41.9	35.5	16.1

* One of the nine teachers submitted only 3 assignments.

Discussion

In summary, the results of this most recent pilot study in middle schools indicate that the reliability of the IQA rubrics is good overall, especially for assessing mathematics instruction, and showed improvement from our previous pilot study in elementary schools. In part, this was accomplished by improving the rater-training program and using more experienced raters—that is, raters who had more expertise in each content area.

Our results also indicated that when teachers comply with the data collection requirements, as few as two observations may be needed to yield a stable estimate of quality at the teacher-level. Of course, it may not be realistic to expect perfect compliance in real world research settings as we found in our data collection efforts. Even though all of the teachers agreed in advance to hold a discussion or engage students in a problem-solving activity on the days we visited, a few teachers decided to engage in other instructional activities instead on one of these days (such as having their students give oral presentations, etc.) and this understandably negatively influenced the stability of the rating scales. If one were using the observation rubrics as a stand-alone indicator of instructional quality then, it would be important to plan for a certain number of teachers not complying with the data collection requirements, and to be prepared to return to the classroom on another day.

Four assignments with student work yielded a stable estimate of quality at the teacher level in both English language arts and mathematics. This replicates our results in mathematics from our previous study in elementary schools, and is different from our previous results for English language arts assignments (Boston & Wolf, 2005; Matsumura et al., 2005). As described earlier, for our current study we asked for four of the same kind of assignments (challenging) from teachers, as opposed to two recent assignments and two challenging assignments. This improved the stability of the ratings for obtaining a teacher-level estimate of quality and is commensurate with other studies that indicate that four assignments can yield a stable estimate of quality (Clare & Aschbacher, 2001). Whether or not to choose to collect either "challenging" or "typical" assignments is likely best determined by the goals of the data collection effort. For the reasons described earlier, however, we chose to ask teachers for challenging assignments for this data collection at least.

Based on these results, and with all the necessary caveats about sample size, it appears that the individual data sources—assignments or observations--could be used to infer a stable estimate of quality at the teacher level. At the same time, however, these results raise important questions about sampling with regard to measuring such a complex phenomenon as instruction. We made the decision to focus on a relatively narrow set of topics and artifacts/behaviors in order to minimize task-sampling variability. For this end we appeared to have had success. Obtaining a stable estimate of practice with the fewest tasks or observations possible is important for reducing the cost and burden on teachers posed by this type of data collection. The drawback, of course, is that constraining what is assessed constrains the generalizability of the findings.

With regard to validity, the IQA rubrics show promise for measuring meaningful differences in students' learning opportunities. The findings from this study suggested that the assignment measure might predict student achievement on all of the relevant subscores of the SAT-10 for ELA (Total Reading, Reading Comprehension, and Vocabulary), as well as the Procedures subscale in mathematics. The observation ratings predicted student achievement on the reading comprehension subscores in ELA, and the Total Math and Procedures subscores in mathematics. Analysis of the field notes and artifacts we collected suggested as well that the IQA ratings are sensitive to differences in the quality of the classroom learning environment.

While our results are inconclusive regarding the predictive validity of the IQA, they raise intriguing questions about the instructional behaviors that might be most associated with student learning and assessment methods that might be most sensitive to those behaviors. This is especially the case given the different relationship between assignment quality and observed instruction in the content areas. Multilevel analyses would need to be conducted with larger samples of teachers to draw definitive conclusions about the relationship of the IQA ratings to student achievement and the relative value of observing in classrooms versus collecting assignments with student work.

Finally, research is needed that focuses on the use of the IQA as a tool for professional learning. Our ratings of observed instruction and assignment quality parallel the findings from other research showing that the quality of academic work in middle schools does not keep pace with the increasingly sophisticated reasoning abilities and cognitive processing that one would expect from adolescents (Juvonen et al, 2004; Midgely & Edelin, 1998; Snow, 2002). Commensurate with what others have found regarding the state of reading comprehension instruction (reviewed in Snow, 2002), we observed few opportunities for students to infer meaning beyond what was represented on the page, to link ideas in texts to larger ideas (or to other texts), or to develop any degree of analysis and interpretation skills. In mathematics, only a very few teachers engaged students in exploring and understanding the nature of mathematical concepts, procedures, and/or relationships such as using complex and non-algorithmic thinking or applying a broad general procedure and connecting that to underlying mathematical concepts. Research is needed that focuses on the utility of the IQA for improving instructional practice in these areas.

References

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, Center for Research on Evaluation, Standards, and Student Testing.

Ball, D. L., & Rowan, B. (2004). Measuring instruction. *The Elementary School Journal, 105*(1), 3-10.

Beck, I. L., McKeown, M. G., Hamilton, R. L., & Kucan, L. (1997). *Questioning the Author: An approach for enhancing student engagement with text*. Delaware: International Reading Association.

Black, P., & William, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan 80*(2), 139–148.

Borko, H., Stecher, B. M., Alonzo, A., Moncure, S., & McClam, S. (2005). Artifact packages for characterizing classroom practice: A pilot study. *Educational Assessment, 10*(2), 73-104.

Boston, M., & Wolf, M. (2005). *Assessing Academic Rigor in Mathematics Instruction: The Development of the Instructional Quality Assessment Toolkit*. (CSE Tech. Rep. No. 672). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Brennan, R. L. (2001a). *Manual for mGENOVA*. Iowa City, IA: Iowa Testing Programs, University of Iowa.

Brennan, R. L. (2001b). mGENOVA (Version 2.1) [Computer software; on-line]. Available: http://www.education.uiowa.edu/casma/computer_programs.htm

Camburn, E., and Barnes, C. (2004). Assessing the validity of a language arts instruction log through triangulation. *Elementary School Journal, 105*, 49-74.

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

Clare, L., & Aschbacher, P. (2001). Exploring the technical quality of using assignments and student work as indicators of classroom practice. *Educational Assessment, 7*(1), 39-59.

Cobb, P., Boufi, A., McClain, K., & Whitenack, J. (1997). Reflective discourse and collective reflection. *Journal for Research in Mathematics Education, 28*(3), 258-277.

Doyle, W. (1983). Academic work. *Review of Educational Research*, *53*, 159-199.

Hamilton, L. & Stecher B. (2004). Responding effectively to test-based accountability. *Phi Delta Kappan*, *85*(8).

Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K. C., Wearne, D., Murray, H., Olivier, A., & Human, P. (1997). *Making sense: Teaching and learning mathematics with understanding*. Portsmouth, NH: Heinemann.

Junker, B. Weisberg, Y. Matsumura, L. C., Crosson, A., Wolf, M., Levison, A., & Resnick, L. (2005). *Overview of the instructional quality assessment* (CSE Tech. Rep. No. 671). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Juvonen, J., Le, V., Kaganoff, T., Augustine, C., & Constant, L. (2004). *Focus on the wonder years: Challenges facing the American middle school*. Santa Monica, CA: RAND Corporation.

Matsumura, L. C., Slater, S. C, Wolf, M., Crosson, A., Levison, A., Peterson, M., & Resnick, L. (2005). *Using the Instructional Quality Assessment Toolkit to investigate the quality of reading comprehension assignments and student work* (CSE Tech. Rep. No. 669). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing. Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.

Matsumura, L. C., Garnier, H., Pascal, J., & Valdés, R. (2001). Measuring instructional quality in accountability systems: Classroom assignments and student achievement. *Educational Assessment*, *8*(3), 207-229.

Midgley, C., & Edelin, K. (1998). Middle school reform and early adolescent well being: The good news and the bad. *Educational Psychologist*, *33*(4), 225-241.

National Center on Education Statistics (NCES). (2004). The condition of education in 2004 in brief. (NCES 2004-076) Washington DC: US Department of Education, Institute of Education Science.

National Commission on Teaching and America's Future (1996). What matters most: Teaching for America's future. Teachers College Columbia University. New York: Author.

O'Connor, M. C., & Michaels, S. (1996). Shifting participant frameworks: Orchestrating thinking practices in group discussions. In D. Ghicks (Ed.), *Discourse, learning, and schooling* (pp. 63-103). New York: Cambridge University Press.

Porter, A. (2002). Measuring the content of instruction: Uses in research and practice. *Educational researcher, 31*(7), 3-14.

Resnick, L. B., & Hall, M. W. (2001). The Principles of Learning: Study tools for educators [CD-ROM, Version 2.0]. Pittsburgh, PA: University of Pittsburgh, Learning Research and Development Center, Institute for Learning. (www.instituteforlearning.org).

Rowan, B., Correnti, R., & Miller, R. J. (2002). What large-scale survey research tells us about teacher effects on student achievement: Insights from the Prospects Study of Elementary Schools. *Teachers College Record, 104*(8), 1525–1567.

Sanders, W., & Horn, P. (1994). The Tennessee value-added assessment system (TVAAS): Mixed methodology in educational assessment. *Journal of Personnel Evaluation in Education, 8*(1), 299-311.

Shavelson, R. J., Baxter, G. P., & Gao, X. (1993). Sampling variability of performance assessments. *Journal of Educational Measurement*, *30*, 215-232.

Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4-14.

Slater, S. C., Matsumura, L. C., & Junker, B. W. (2005, April). *Generalizability of a performance measure of instructional quality.* Paper presented at the Annual Conference for NCME, Montreal.

Snow, C. E. (2002). *Reading for understanding: Toward a research and development program in reading comprehension*. Prepared for the Office of Research and Improvement. Santa Monica, CA: RAND Corporation.

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

Stein, M. K., Smith, M. S., Henningsen, M. A., & Silver, E. A. (2000). *Implementing standards-based mathematics instruction: A casebook for professional development.* New York: Teachers College Press.

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

Appendix A: IQA Rubrics

Rubrics Rating Reading Comprehension: Lesson Observation

Accountable Talk Rubrics

Rubric 1: Participation	
4	Over 75% of the students participated throughout the discussion.
3	50-75% of the students participated in the discussion.
2	25-49% of the students participated in the discussion.
1	Less than 25% of the students participated in the discussion.
0	None of the students participated in the discussion.
N/A	Reason:

Rubric 2: Teacher's Linking	
4	The teacher consistently connects speakers' contributions to each other and shows how ideas/positions shared during the discussion relate to each other by revoicing or recapping students' ideas. (e.g., "What I hear you saying is that the character has changed from the beginning to the end of the book which is similar to Ana's idea that the character has matured throughout the book.")
3	At least twice during the lesson the teacher connects speakers' contributions to each other and shows how ideas/positions relate to each other by recapping or revoicing students' ideas. (e.g., "What I hear you saying is that the character has changed from the beginning to the end of the book which is similar to Ana's idea that the character has matured throughout the book.")
2	At one or more points during the discussion, the teacher links speakers' contributions to each other, but does not show how ideas/positions relate to each other. No follow-up questions are asked after speakers' contributions OR teacher revoices but content is not academically relevant OR only one strong effort is made to connect speakers' contributions to each other.
1	Teacher does not make any effort to link or revoice speakers' contributions.
0	Discussion was not directly related to the text.
N/A	Reason:

Rubric 3: Students' Linking	
4	Students consistently connect their contributions to each other and show how ideas/positions shared during the discussion relate to each other. (e.g., "I agree with Jay because…")
3	At least twice during the lesson the students connect their contributions to each other and show how ideas/positions shared during the discussion relate to each other. (e.g., "I agree with Jay because…")
2	At one or more points during the discussion, the students link students' contributions to each other, but do not show how ideas/positions relate to each other. (e.g., "I disagree with Ana.") OR only one strong effort is made to connect their contributions with each other.
1	Students do not make any effort to link or revoice students' contributions.
0	Discussion was not directly related to the text.
N/A	Reason:

Rubric 4: Asking (Teachers)	
4	The teacher consistently asks students academically relevant questions that may include asking students to provide evidence for their contributions, pressing students for accuracy, OR pressing students to explain their reasoning.
3	At least twice during the lesson the teacher asks students academically relevant questions which may include asking students to provide evidence for their contributions, pressing students for accuracy, OR pressing students to explain their reasoning.
2	There are one or more superficial, trivial efforts, or formulaic efforts to ask students to provide evidence for their contributions, OR there are one or more superficial, trivial, or formulaic efforts to ask students to explain their reasoning OR only one strong effort is made to ask students academically relevant questions or press students to explain their reasoning.
1	There are no efforts to ask students to provide evidence for their contributions, AND there were no efforts to ask students to explain their thinking.
0	Discussion was not directly related to the text.
N/A	Reason:

Rubric 5: Providing (Students)	
4	Students consistently provide accurate and appropriate evidence for their claims, including frequent references to the text or prior classroom experience, OR students explain their thinking, using reasoning in ways appropriate to the discipline.
3	At least twice during the lesson students provide accurate and appropriate evidence for their claims, including frequent references to the text or prior classroom experience, OR students explain their thinking, using reasoning in ways appropriate to the discipline.
2	In general, what little evidence is offered to back up claims is inaccurate, incomplete, or vague, OR there are one or more superficial or trivial efforts to provide evidence. In general, what little attempt to explain reasoning is inaccurate, incomplete, or vague, OR there are one or more superficial or trivial efforts to explain the speaker's reasoning, OR students only make one strong effort to provide evidence or explain their thinking.
1	Students do not back up their claims, OR do not explain the reasoning behind their claims.
0	Discussion was not directly related to the text.
N/A	Reason:

Academic Rigor: Reading Comprehension

Rubric 1: Rigor of the Text	
3	The text contains substantial "grist" for students to grapple with in a group discussion. This grist is seen in the complexity of the content (theme, relationships between characters, etc.) and in the writer's craft (literary language, rich vocabulary, organizational structures).
2	The text contains some "grist" for students to grapple with during group discussion. There may be some degree of complexity in the content (theme, relationships between characters, etc.) and in the writer's craft (literary language, rich vocabulary, organizational structures).
1	There is minimal "grist" for students to discuss to make meaning of the story. It may contain a simple narrative or basic information. The text may be a simplified version of a complex text, or a short excerpt from a workbook.
0	There is nothing about the text that requires extended discussion.
N/A	Reason:

Rubric 2: Discussion	
4	The teacher guides students to engage with the underlying meanings or literary characteristics of a text. Students interpret or analyze a text and use extensive and detailed evidence from the text to support their ideas or opinions.
3	The teacher guides students to construct an enriched and elaborated understanding of the text including analysis of the causes and effects of events and/or character actions. The students may engage with some underlying meanings or literary characteristics of a text, but they provide limited evidence from the text to support their ideas or opinions. The students construct an in-depth summary including the main idea.
2	The teacher guides students to construct a surface-level summary of the text based on straightforward information, OR students engage in perfunctory response to the text (e.g., superficial understanding). Students use little evidence from the text to support their ideas or opinions.
1	The teacher guides students to recall fragmented, isolated facts from a text, OR the teacher guides students to discuss a topic that does not directly reference information from the text.
0	Discussion was not directly related to the text.
N/A	Reason:

Rubric 3: Lesson Activity	
4	The task guides students to engage with the underlying meanings or nuances of a text. Students interpret or analyze a text AND use extensive and detailed evidence from the text to support their ideas or opinions AND the task provides students with an opportunity to fully develop their thinking (e.g., challenging questions, extended responses, and analytical and interpretive responses).
3	The task guides students to engage with some underlying meanings or nuances of a text. Students may interpret or analyze a text, BUT they use limited evidence from the text to support their ideas or opinions. There is some opportunity for students to develop their thinking (e.g., challenging questions but structured responses).
2	The task guides students to construct a literal summary of the text based on straightforward (surface-level) information OR engage with surface-level information about the text only. The task guides students to use little or no evidence from the text to support their ideas or opinions.
1	The task guides students to recall isolated, straightforward (surface-level) facts about a text OR write on a topic that does not directly reference information from the text OR the task guides students in recalling fragmented information about the text.
0	The lesson activity did not pertain to reading comprehension.
N/A	Reason:

Clear Expectations/Self Management of Learning

Rubric 1: Clarity and Detail of Expectations	
4	The teacher provides a detailed list to students regarding what they would need to do, or include in their work, to do well on the task. The teacher explains what high quality work would look like to students, and illustrates this with models of high-quality student work and/or describes the difference between high-quality and lower-quality work (e.g., "Write at least three paragraphs summarizing the story we just read. Remember to include what the story is about, a description of the main characters, and the main problem and solution to the problem. Use lots of examples from the story to support what your write. A high quality summary will include many examples from the story, if there are only one or two examples from the story, then it will be considered low quality.").
3	The teacher provides a detailed list to students regarding what they would need to do, or include in their work, to do well on the task (e.g., "Write at least three paragraphs summarizing the story we just read. Remember to include what the story is about, a description of the main characters, and the main problem and solution to the problem. Use lots of examples from the story to support what your write.").
2	The teacher provides a cursory or very general explanation for what she is looking for in the quality of students' work (e.g., "Write a summary of the story we just read. Remember to include examples from the story in your writing.").
1	The teacher provides directions for the activity, but does not describe what students would need to do, or include in their work, to be successful on the task (e.g., "Return to your seats and write a summary of the story we just read.").
0	Expectations are not shared with students.
N/A	Reason:

Rubric 2: Rigor of Expectations in Reading Comprehension	
4	At least one of the teacher's expectations focuses on analyzing and interpreting the text (e.g., inferring major themes, analyzing character motives, comparing and contrasting two texts or characters, etc.) AND at least one expectation focuses on including evidence or examples to support a position.
3	At least one of the teacher's expectations focuses on analyzing and interpreting the text (e.g., inferring major themes, analyzing character motives, comparing and contrasting two texts or characters, etc.).
2	The teacher's expectations focus on building a basic understanding of the text (e.g., summarizing).
1	The teacher's expectations do not focus on reading comprehension. The expectations may focus solely on procedures (e.g., how well students follow directions, producing neat work, or behavioral norms) or content not directly related to reading comprehension (e.g., writing conventions) OR the expectations focus on isolated, fragmented information from the text.
0	The teacher's expectations were not given.
N/A	Reason:

Rubric 3: Access to Expectations	
4	Criteria for the quality of work expected and how work will be scored is *readily accessible* to ALL students. There is a public record of these criteria.
3	Criteria for quality of work expected have been explicated to ALL students. However, there is no public record of these criteria.
2	Criteria for quality of work expected have been explicated to SOME students. There is no public record of these criteria.
1	Criteria shared with students were procedural only.
0	No criteria were shared with students.
N/A	Reason:

Rubrics Rating Reading Comprehension: Assignments

Rubric 1: Rigor of the Text	
3	The text contains lots of "grist" for students to grapple with in a group discussion. This grist is seen in the complexity of the content (theme, relationships between characters, etc.) and in the writer's craft (literary language, rich vocabulary, organizational structures).
2	The text contains some "grist" for students to grapple with during group discussion. There may be some degree of complexity in the content (theme, relationships between characters, etc.) and in the writer's craft (literary language, rich vocabulary, organizational structures).
1	There is minimal "grist" for students to discuss to make meaning of the story. It may contain a very simple narrative or very basic information, but these are so straightforward that there is nothing about the text that requires extended discussion. For example, the text may simplified version of a complex text, or a short excerpt from a workbook.
N/A	Reason:

Rubric 2: Analyzing and Interpreting the Text through Lesson Activity	
4	The task guides students to engage with the underlying meanings or nuances of a text. Students interpret or analyze a text AND use extensive and detailed evidence from the text to support their ideas or opinions AND the task provides students with an opportunity to fully develop their thinking (e.g. challenging questions, extended responses, and analytical and interpretive responses).
3	The task guides students to engage with some underlying meanings or nuances of a text. Students may interpret or analyze a text, BUT they use limited evidence from the text to support their ideas or opinions. There is some opportunity for students to develop their thinking (e.g. challenging questions but structured responses).
2	The task guides students to construct a literal summary of the text based on straightforward (surface-level) information OR engage with surface-level information about the text only. The assignment task guides students to use little or no evidence from the text to support their ideas or opinions.
1	The task guides students to recall isolated, straightforward (surface-level) facts about a text OR write on a topic that does not directly reference information from the text OR the task guides students in recalling fragmented information about the text.
N/A	Reason:

Rubric 3: Implementation of the Task	
4	Students engaged with the underlying meanings or nuances of a text. Students interpreted or analyzed a text AND used extensive and detailed evidence from the text to support their ideas or opinions.
3	Students engaged with some underlying meanings or nuances of a text. Students interpreted or analyzed a text BUT used limited evidence from the text to support their ideas or opinions. Students constructed an in-depth summary including the main idea.
2	Students constructed a literal summary of the text based on straightforward (surface-level) information OR students engaged with surface-level information about the text only. Students used little or no evidence from the text to support their ideas or opinions. OR the task guides students to engage with interpreting or analyzing a text but provides limited opportunity to develop their thinking.
1	Students recalled isolated, straightforward (surface-level) facts about a text OR wrote on a topic that does not directly reference information from the text.

Rubric 4: Academic Rigor in Teacher's Expectations	
4	At one of the teacher's expectations focuses on analyzing and interpreting the text (e.g., inferring major themes, analyzing character motives, comparing and contrasting two texts or characters, etc.) AND at one expectation focuses on including evidence or examples to support a position.
3	At one of the teacher's expectations focuses on analyzing and interpreting the text (e.g., inferring major themes, analyzing character motives, comparing and contrasting two texts or characters, etc.).
2	The teacher's expectations focus on building a basic understanding of the text (e.g. summarizing).
1	The teacher's expectations do not focus on reading comprehension. The expectations may focus solely on procedures (e.g. how well students follow directions, producing neat work, or behavioral norms) or content not directly related to reading comprehension (e.g., writing conventions) OR The teacher's expectations do not focus on coherent understanding of the text (e.g., recalling fragmented information about a text).

Rubric 1: Clarity and Detail of Expectations	
4	The expectations are very clear and explicit regarding the quality of work expected. The criteria for quality work are appropriately detailed.
3	The expectations are clear regarding the quality of work expected. However, there is no elaboration of what level of quality is expected for each criterion.
2	The expectations for the quality of student's work are broadly stated and unelaborated.
1	The teacher's expectations for the quality of student's work are unclear OR the expectations for quality work are not shared with students.
N/A	Reason:

Rubric 2: Communication of Expectations	
4	Teacher discusses the expectations or criteria for student work (e.g., scoring guide, rubric, etc.) with students in advance of their completing the assignment and models high-quality work.
3	Teacher discusses the expectations or criteria for student work (e.g., scoring guide, rubric, etc.) with students in advance of their completing the assignment.
2	Teacher provides a copy of the criteria for assessing student work (e.g., scoring guide, rubric, etc.) to students in advance of their completing the assignment.
1	Teacher does not share the criteria for assessing students' work (e.g., scoring guide, rubric, etc.) with the students in advance of their completing the assignment. (e.g., Teacher may provide a copy of the scoring rubric to students when giving them their final grade.
N/A	Reason:

Rubrics Rating Mathematics: Lesson Observation

Accountable Talk

Rubric 1: Participation	
4	Over 75% of the students participated throughout the discussion.
3	50-75% of the students participated in the discussion.
2	25-49% of the students participated in the discussion.
1	Less than 25% of the students participated in the discussion.
0	None of the students participated in the discussion.
N/A	Reason:

Rubric 2: Teacher's Linking	
4	The teacher consistently connects speakers' contributions to each other and shows how ideas/positions shared during the discussion relate to each other by revoicing or recapping students' ideas.
3	At least twice during the lesson the teacher connects speakers' contributions to each other and shows how ideas/positions relate to each other by recapping or revoicing students' ideas.
2	At one or more points during the discussion, the teacher links speakers' contributions to each other, but does not show how ideas/positions relate to each other. No follow-up questions are asked after speakers' contributions OR teacher revoices but content is not academically relevant OR only one strong effort is made to connect speakers' contributions to each other.
1	Teacher does not make any effort to link or revoice speakers' contributions.
0	Class discussion was not related to mathematics.
N/A	Reason:

Rubric 3: Students' Linking	
4	The students consistently connect their contributions to each other and show how ideas/positions shared during the discussion relate to each other. (e.g., "I agree with Jay because…")
3	At least twice during the lesson the students connect their contributions to each other and show how ideas/positions shared during the discussion relate to each other. (e.g., "I agree with Jay because…")
2	At one or more points during the discussion, the students link students' contributions to each other, but does not show how ideas/positions relate to each other. (e.g., "I disagree with Ana.") OR only one strong effort is made to connect their contributions with each other.
1	Students do not make any effort to link or revoice students' contributions.
0	Class discussion was not related to mathematics.
N/A	Reason:

Rubric 4: Asking (Teachers)	
4	The teacher consistently asks students academically relevant questions that may include asking students to provide evidence for their contributions, pressing students for accuracy, OR pressing students to explain their reasoning.
3	At least twice during the lesson the teacher asks students academically relevant questions which may include asking students to provide evidence for their contributions, pressing students for accuracy, OR pressing students to explain their reasoning.
2	There are one or more superficial, trivial efforts, or formulaic efforts to ask students to provide evidence for their contributions, OR there are one or more superficial, trivial, or formulaic efforts to ask students to explain their reasoning OR only one strong effort is made to ask students academically relevant questions OR press students to explain their reasoning.
1	There are no efforts to ask students to provide evidence for their contributions, AND there are no efforts to ask students to explain their thinking.
0	Class discussion was not related to mathematics.
N/A	Reason:

Rubric 5: Providing (Students)	
4	Students consistently provide accurate and appropriate evidence for their claims, including frequent references to the text or prior classroom experience, OR students explain their thinking, using reasoning in ways appropriate to the discipline.
3	At least twice during the lesson students provide accurate and appropriate evidence for their claims, including frequent references to the text or prior classroom experience, OR students explain their thinking, using reasoning in ways appropriate to the discipline.
2	In general, what little evidence is offered to back up claims is inaccurate, incomplete, or vague, OR there are one or more superficial or trivial efforts to provide evidence. In general, what little attempt to explain reasoning is inaccurate, incomplete, or vague, OR there are one or more superficial or trivial efforts to explain the speaker's reasoning, OR students only make one strong effort to provide evidence or explain their thinking.
1	Students do not back up their claims, OR do not explain the reasoning behind their claims.
0	Class discussion was not related to mathematics.
N/A	Reason:

Academic Rigor

Rubric 1: Potential of the Task	
4	**The task has the potential to engage students in exploring and understanding the nature of mathematical concepts, procedures, and/or relationships, such as:** • Doing mathematics: using complex and non-algorithmic thinking (i.e., there is not a predictable, well-rehearsed approach or pathway explicitly suggested by the task, task instructions, or a worked-out example); OR • Procedures with connections: applying a broad general procedure that remains closely connected to mathematical concepts. The task must explicitly prompt for evidence of students' reasoning and understanding. For example, the task **MAY** require students to: • solve a genuine, challenging problem for which students' reasoning is evident in their work on the task; • develop an explanation for why formulas or procedures work; • identify patterns and form generalizations based on these patterns; • make conjectures and support conclusions with mathematical evidence; • make explicit connections between representations, strategies, or mathematical concepts and procedures. • follow a prescribed procedure in order to explain/illustrate a mathematical concept, process, or relationship.
3	**The task has the potential to engage students in complex thinking or in creating meaning for mathematical concepts, procedures, and/or relationships. However, the task does not warrant a "4" because:** • the task does not explicitly prompt for evidence of students' reasoning and understanding. • students may be asked to engage in doing mathematics or procedures with connections, but the underlying mathematics in the task is not appropriate for the specific group of students (i.e., too easy or too hard to promote engagement with high-level cognitive demands); • students may need to identify patterns but are not pressed for generalizations; • students may be asked to use multiple strategies or representations but the task does not explicitly prompt students to develop connections between them; • students may be asked to make conjectures but are not asked to provide mathematical evidence or explanations to support conclusions
2	The potential of the task is limited to engaging students in using a procedure that is either specifically called for or its use is evident based on prior instruction, experience, or placement of the task. **There is little ambiguity about what needs to be done and how to do it.** The task does not require students to make connections to the concepts or meaning underlying the procedure being used. **Focus of the task appears to be on producing correct answers rather than developing mathematical understanding (e.g., applying a specific problem solving strategy, practicing a computational algorithm).** **OR** The task does not require student to engage in cognitively challenging work; the task is easy to solve.
1	The potential of the task is limited to engaging students in **memorizing or reproducing** facts, rules, formulae, or definitions. The task does not require students to make connections to the concepts or

	meaning that underlie the facts, rules, formulae, or definitions being memorized or reproduced.
0	The task requires no mathematical activity.
N/A	Reason

Rubric 2: Implementation of the Task	
4	**Students engaged in exploring and understanding the nature of mathematical concepts, procedures, and/or relationships, such as:** • Doing mathematics: using complex and non-algorithmic thinking (i.e., there is not a predictable, well-rehearsed approach or pathway explicitly suggested by the task, task instructions, or a worked-out example); OR • Procedures with connections: applying a broad general procedure that remains closely connected to mathematical concepts. There is explicit evidence of students' reasoning and understanding. For example, students may have: • solved a genuine, challenging problem for which students' reasoning is evident in their work on the task; • developed an explanation for why formulas or procedures work; • identified patterns and formed generalizations based on these patterns; • made conjectures and supported conclusions with mathematical evidence; • made explicit connections between representations, strategies, or mathematical concepts and procedures. • followed a prescribed procedure in order to explain/illustrate a mathematical concept, process, or relationship.
3	**Students engaged in complex thinking or in creating meaning for mathematical concepts, procedures, and/or relationships. However, the implementation does not warrant a "4" because:** • there is no explicit evidence of students' reasoning and understanding. • students engaged in doing mathematics or procedures with connections, but the underlying mathematics in the task was not appropriate for the specific group of students (i.e., too easy or too hard to sustain engagement with high-level cognitive demands); • students identified patterns but did not make generalizations; • students used multiple strategies or representations but connections between different strategies/representations were not explicitly evident; • students made conjectures but did not provide mathematical evidence or explanations to support conclusions
2	Students engaged in using a procedure that was either specifically called for or its use was evident based on prior instruction, experience, or placement of the task. **There was little ambiguity about what needed to be done and how to do it. S**tudents did not make connections to the concepts or meaning underlying the procedure being used. **Focus of the implementation appears to be on producing correct answers rather than developing mathematical understanding (e.g., applying a specific problem solving strategy, practicing a computational algorithm).** OR **Students did not engage in cognitively challenging work; the task was easy to solve.**
1	Students engaged in memorizing or reproducing facts, rules, formulae, or definitions. Students do not make connections to the concepts or meaning that underlie the facts, rules, formulae, or definitions being memorized or reproduced.
0	Students did not engage in mathematical activity.

N/A	Reason:

Rubric 3: Student Discussion Following Task	
4	Students show/describe written work and provide complete and thorough explanations of why their strategy, idea, or procedure is valid. Students explain why their strategy works and/or is appropriate for the problem by making connections to the underlying mathematical ideas (e.g., "I divided because we needed equal groups"). OR Students show/discuss more than one strategy or representation for solving the task, and provide explanations of why the different strategies/representations were used to solve the task.
3	Students show/describe written work BUT the explanations are not complete and thorough (e.g., student responses often require extended press from the teacher, are incomplete, lack precision, or fall short making explicit connections). OR Students show/discuss more than one strategy or representation for solving the task, and do provide explanations of how the different strategies/representations were used to solve the task but do not explain why they were used.
2	Students show/describe written work for solving the task (e.g., the steps for a multiplication problem, finding an average, or solving an equation; what they did first, second, etc) but do not explain why their strategy or procedure works and/or was appropriate for the problem; OR Students show/discuss only one strategy or representation for solving the task.
1	Students provide brief or one-word answers (e.g., fill in blanks);
0	Student's responses are non-mathematical.
N/A	Reason:

Clear Expectations/Self Management of Learning

Rubric 1: Clarity and Detail of Expectations*	
4	The teacher provides an explicit explanation to students regarding what they would need to do, or include in their work, to do well on the task. The teacher explains what high quality work would look like to students, and illustrates this with models of high-quality student work and/or describes the difference between high-quality and lower-quality work
3	The teacher provides an explicit explanation to students regarding what they would need to do, or include in their work, to do well on the task.
2	The teacher provides a cursory or very general explanation for what she is looking for in the quality of students' work.
1	The teacher provides directions for the activity, but does not describe what students would need to do, or include in their work, to be successful on the task.
0	Expectations for quality work are not shared with students.
N/A	Reason:

Rubrics Rating Mathematics: Assignments

Rubric 1: Potential of the Task	
4	**The task has the potential to engage students in exploring and understanding the nature of mathematical concepts, procedures, and/or relationships, such as:** • Doing mathematics: using complex and non-algorithmic thinking (i.e., there is not a predictable, well-rehearsed approach or pathway explicitly suggested by the task, task instructions, or a worked-out example); OR • Procedures with connections: applying a broad general procedure that remains closely connected to mathematical concepts. **<u>The task must explicitly prompt for evidence of students' reasoning and understanding.</u>** For example, the task **MAY** require students to: • solve a genuine, challenging problem for which students' reasoning is evident in their work on the task; • develop an explanation for why formulas or procedures work; • identify patterns and form generalizations based on these patterns; • make conjectures and support conclusions with mathematical evidence; • make explicit connections between representations, strategies, or mathematical concepts and procedures. • follow a prescribed procedure in order to explain/illustrate a mathematical concept, process, or relationship.
3	**The task has the potential to engage students in complex thinking or in creating meaning for mathematical concepts, procedures, and/or relationships. However, the task does not warrant a "4" because:** • the task does not explicitly prompt for evidence of students' reasoning and understanding. • students may be asked to engage in doing mathematics or procedures with connections, but the

	underlying mathematics in the task is not appropriate for the specific group of students (i.e., too easy <u>or</u> too hard to promote engagement with high-level cognitive demands); • students may need to identify patterns but are not pressed for generalizations; • students may be asked to use multiple strategies or representations but the task does not explicitly prompt students to develop connections between them; • students may be asked to make conjectures but are not asked to provide mathematical evidence or explanations to support conclusions
2	The potential of the task is limited to engaging students in using a procedure that is either specifically called for or its use is evident based on prior instruction, experience, or placement of the task. **There is little ambiguity about what needs to be done and how to do it.** The task does not require students to make connections to the concepts or meaning underlying the procedure being used. **Focus of the task appears to be on producing correct answers rather than developing mathematical understanding (e.g., applying a specific problem solving strategy, practicing a computational algorithm).** OR The task does not require student to engage in cognitively challenging work; the task is easy to solve.
1	The potential of the task is limited to engaging students in **memorizing or reproducing facts, rules, formulae, or definitions**. The task does not require students to make connections to the concepts or meaning that underlie the facts, rules, formulae, or definitions being memorized or reproduced. OR The task requires no mathematical activity.

*Representations include numbers and/or symbols, diagrams/pictures, use of written/verbal language , graphs, tables/charts, concrete materials.]

Rubric 2: Implementation of the Task	
4	**Students engaged in exploring and understanding the nature of mathematical concepts, procedures, and/or relationships, such as:** • Doing mathematics: using complex and non-algorithmic thinking (i.e., there is not a predictable, well-rehearsed approach or pathway explicitly suggested by the task, task instructions, or a worked-out example); OR • Procedures with connections: applying a broad general procedure that remains closely connected to mathematical concepts. <u>There is explicit evidence of students' reasoning and understanding in their written work.</u> For example, students may have: • solved a genuine, challenging problem for which students' reasoning is evident in their written work; • provided an explanation for why formulas or procedures work; • identified patterns and formed generalizations based on these patterns; • made conjectures and supported conclusions with mathematical evidence; • made explicit connections between representations, strategies, or mathematical concepts and procedures, • followed a prescribed procedure in order to explain/illustrate a mathematical concept, process, or relationship.
3	Students engaged in complex thinking or in creating meaning for mathematical concepts, procedures, and/or relationships. However, the implementation does not warrant a "4" because: • there is no explicit evidence of students' reasoning and understanding. • students engaged in doing mathematics or procedures with connections, but the underlying

	mathematics in the task was not appropriate for the specific group of students (i.e., too easy <u>or</u> too hard to sustain engagement with high-level cognitive demands); • students identified patterns but did not make generalizations; • students used multiple strategies or representations but connections between different strategies/representations were not explicitly evident; • students made conjectures but did not provide mathematical evidence or explanations to support conclusions
2	Students engaged in using a procedure that was either specifically called for or its use was evident based on prior instruction, experience, or placement of the task. **There was little ambiguity about what needed to be done and how to do it. S**tudents did not connections to the concepts or meaning underlying the procedure being used. **Focus of the implementation appears to be on producing correct answers rather than developing mathematical understanding (e.g., applying a specific problem solving strategy, practicing a computational algorithm).** OR **Student did not engage in cognitively challenging work; the task was easy to solve.**
1	**Students engage in memorizing or reproducing facts, rules, formulae, or definitions. Students do not make connections to the concepts or meaning that underlie the facts, rules, formulae, or definitions being memorized or reproduced.** **OR Students did not engage in mathematical activity.**

Rubric 3: Rigor in Students' Responses to the Task	
4	Students show written work and provide complete and thorough explanations of why their strategy, idea, or procedure is valid. Students explain why their strategy works and/or is appropriate for the problem by making connections to the underlying mathematical ideas (e.g., "I divided because we needed equal groups"). OR Student work displays use of more than one strategy or representation* for solving the task, and provides a written explanation of how the different strategies/representations were used to solve the task.
3	Students show written work and provide explanations BUT the explanations are incomplete or are procedural in nature. Students explain the steps of their work (e.g., what they did first, second, etc.) but do not explain why their strategy or procedure works and/or was appropriate for the problem; OR Student work displays use of more than one strategy or representation* for solving the task.
2	Students show written work for solving the task (e.g., the steps for a multiplication problem, finding an average, or solving an equation) with no written explanation; OR Student work displays use of only one strategy or representation* for solving the task.
1	Students provide brief or one-word answers (e.g., fill in blanks); OR Student's responses are non-mathematical.

Rubric 4: Academic Rigor in Teacher's Expectations*	
4	The majority of the teacher's expectations are for students to: • use complex and non-algorithmic thinking (i.e., there is not a predictable, well-rehearsed approach or pathway explicitly suggested by the task, task instructions, or a worked-out example); • explore and understand the nature of mathematical concepts, procedures, and/or relationships. [The expectations for mathematical content are stated explicitly in one of the sources indicated by the * below.] For example, the teacher may expect students to: • solve a genuine, challenging problem; • develop an understanding for why formulas or procedures work; • identify patterns and form generalizations based on these patterns; • make conjectures and support conclusions with mathematical evidence; • make connections between representations, strategies, or mathematical concepts and procedures.
3	At least some of the teacher's expectations are for students to engage in complex thinking or in understanding important mathematics. However, the teacher's expectations do not warrant a "4" because: • the expectations are appropriate for a task that lacks the complexity to be a "4"; • the expectations do not reflect the potential of the task to elicit complex thinking (e.g., identifying patterns but not forming generalizations; using multiple strategies or representations without developing connections between them; providing shallow evidence or explanations to support conclusions). • the teacher expects complex thinking, but the expectations do not reflect the mathematical potential of the task.
2	The teacher's expectations focus on skills that are germane to student learning, but these are not complex thinking skills (e.g., expecting use of a specific problem solving strategy, expecting short answers based on memorized facts, rules or formulas; expecting accuracy or correct application of procedures rather than on understanding mathematical concepts).
1	The teacher's expectations do not focus on substantive mathematical content. The teacher's focus may be solely on activities or classroom procedures (e.g., following directions, producing neat work, or following norms for cooperative learning).

*Rate this dimension based on Coversheet Q 4 and the attached rubric.

Rubric 1: Clarity and Detail of Expectations	
4	The expectations for the quality of students' work are very clear and elaborated. Each dimension or criterion for the quality of students' work is clearly articulated. Additionally, varying degrees of success are clearly differentiated.
3	The expectations for the quality of students' work are clear and somewhat elaborated. Levels of quality may be vaguely differentiated for each criterion (i.e., little information is provided for what distinguishes high, medium and low performance.)
2	The expectations for the quality of student's work are broadly stated and unelaborated.
1	The teacher's expectations for the quality of student's work are unclear OR the expectations for quality work are not shared with students.

Rubric 2: Communications of Expectations	
4	Teacher discusses the expectations or criteria for student work (e.g., scoring guide, rubric, etc.) with students in advance of their completing the assignment and models high-quality work.
3	Teacher discusses the expectations or criteria for student work (e.g., scoring guide, rubric, etc.) with students in advance of their completing the assignment.
2	Teacher provides a copy of the criteria for assessing student work (e.g., scoring guide, rubric, etc.) to students in advance of their completing the assignment.
1	Teacher does not share the criteria for assessing students' work (e.g., scoring guide, rubric, etc.) with the students in advance of their completing the assignment. (e.g., Teacher may provide a copy of the scoring rubric to students when giving them their final grade.
N/A	Reason:

Appendix B:

High and Low Quality Examples of Observed Lessons and Assignments

In the following sections, examples of observed lessons and classroom assignments with student work are described that illustrate the range of instructional quality captured by our ratings—both high and low.

High quality ELA lesson. Mrs. Jones' lesson illustrates the highest scores (4's) for the IQA rubrics. She began the class discussion by reviewing with the students her expectations for their participation. Mrs. Jones referred students to a chart posted in the front of the room and read aloud:

> Look at the chart. Everyone participates; makes connections to characters, other books, or self; cites text to back up statements; brings others in; stays on topic; validates contributions of others; analyzes what the author is doing; is respectful; asks questions; looks for clarification; takes turns; doesn't yell; uses conversation starters; [there are] no right or wrong answers.

These expectations (or class rules) are exemplary in that they focus both on the substance of students' comments (e.g., cites evidence to back up statements; makes connections to characters, other books, or self; analyzes what the author is doing; etc.) as well as the manner in which students contribute and relate to their peers (e.g., brings others in, validates contributions of others, is respectful, takes turns, etc.). It is notable as well that nearly every student in the class spoke at some point during the discussion. An excerpt from the discussion is shown below:

1 T: Discuss among yourselves why Mamasita is so unhappy.
2 S7: I think she's depressed because she dropped out of school and only knows a few words of English and she went from a home in the country to the city.
3 T: Let me correct you there, in that she came from another country.
4 S8: I agree with him and I think she's mad because her niece or son is speaking English and she doesn't.
5 S9: I respectfully disagree, I think she's jealous of the boy.
6 S10: I think she's homesick.
7 S11: I agree with Paul, Andy, and Kevin, and respectfully disagree with Mike. On page 78 "….Pepsi…" She's disappointed and sad that he's learning English and not … [Spanish]
8 S12: I agree and also think that she doesn't want to forget home.

9 T: Yes, she thinks it's a betrayal that her son is learning English. She's resisting learning, like in Sea Folks, some didn't want to assimilate.
10 S14: What do you think? (to another student)
11 S15: Only eight words…
12 S16: I think this is a great example of how kids learn English easier.
13 T: Can you think of another story where kids learned faster?
14 S16: Sea Folks…Taiwan (uncle, son)
15 T: That's a good example of kids picking it up easier. Anyone's parents have a tough time?
16 S: My mom was born in Japan.
17 S: My parents have a tough time.
18 S: My dad is still struggling, going to school. My mom can't speak English well.
19 S: My parents learned English in Africa.
20 T: Who else learned English from TV?
21 S5: The kid from Sea Folks.

In this excerpt Mrs. Jones both builds on students' contributions and presses them to make connections between different texts. For example, in turn 9, Mrs. Jones agrees and builds on a student's contribution and links the experience of the character to the experience of other characters the class had read about ("Yes, she thinks it's a betrayal that her son is learning English. She's resisting learning, like in Sea Folks, some didn't want to assimilate."). Later, in turn 13, she presses a student to build on his statement (in turn 12) that "this is a great example of how kids learn English easier" ("Can you think of another story of how kids learn English faster?"). In turn 15, she also prompts the students to connect the story to their own families' experiences ("Anyone's parents have a tough time [learning English]?") and again connects the story they are reading to another text (in turn 20) ("Who else learned English from TV?").

Students, in turn, built on and validated each other's contributions throughout the discussion. This is exemplified in turns 4 through 8. In turn 4, a student validates a students' earlier comment, ("I agree with him…") and adds to it, ("and I think she's mad because her niece is speaking English and she doesn't."). This statement is challenged in turn 5 and 6 by students who commented "I respectfully disagree, I think she's jealous of the boy" and "I think she's homesick." These exchanges from multiple students were summarized in turn 7 by a student who commented, "I agree with Paul, Andy, and Kevin, and respectfully disagree with Mike…" and added to in turn 8 by a student who said, "I agree and also think that she doesn't want to forget home." What is exemplary here, and markedly different from Mrs. Smith's classroom, was that throughout the conversation students listened attentively to one another, and were able to both agree and disagree with one another in a constructive, respectful manner. No one was criticized for their contributions.

Additionally, Mrs. Jones asked the students probing questions and guided students to provide evidence for their contributions. This was evident in the students' responses as illustrated in the following excerpt:

41 T: Is he an uncaring husband or is he frustrated and doing his best?
42 S: I think he is frustrated (reading from text)…"man paints the walls pink."
43 S: I don't think.
44 S: I'd like to hear what she has to say.
45 S: The husband is frustrated.
46 T: To build upon that, the husband is doing his best like Tracy said. I moved here from [another country] and learned that I could adapt to my new country and still love the old country. That's what the husband has done…the wife thinks it's a betrayal to her home country to make friends here. She just wants to go home that's why she's saying on page 78 "when, when, when?" I learned things about Ezparanza. What can we learn about her from this?
47 S: I thought it was his mom, not his wife.
48 S: I disagree with you because Ezparanza would know that it's his wife.
49 T: I think Paul is right. Heidi?
50 S: Where it says, (reading from text)…
51 T: But what can we tell about <u>Ezparanza</u>?
52 S: That she's very observant.
53 S: I agree with Paul and Anna and would like to see if Patrick does also?
54 S: Yes, shows that she understands.
55 T: Ezparanza is non-judgmental and shows understanding.

In turn 41 Mrs. Jones asks a student to consider if the husband in the story is "uncaring" or "frustrated and doing his best"—a challenging question that opens the door for the students to infer the character's motivations (or "read between the lines") of the events described in the story. Later in the exchange (turn 48), the teacher prompts the students to think about what the events in the story reveal about the main character (Esperanza)—another question that guides students to consider the subtle nuances of the story (versus recalling surface-level, basic information). And in turn 51, Mrs. Jones presses a student to focus her response on the main character. The student responded to these questions by directly referring to the text (turns 42 and 50), using evidence to support their responses. For these reasons, both the quality of the questions Mrs. Jones asked her students and the amount of pressing she did to elicit thoughtful, supportive responses, this lesson illustrates a high score for the rigor of the discussion.

Low quality ELA lesson. Mrs. Smith's lesson, in contrast, illustrates the lowest scores for the quality of the lesson (primarily 1's for each of the IQA rubrics). The lesson began chaotically. Students milled around talking and laughing with each other well after the bell had rung signaling the beginning of the class period. Mrs. Smith spent this time passing back graded homework to the students. After a time she called the class to attention and began a short class discussion about character traits. The purpose of the discussion (and following class activity) was to prepare students to describe the main characters from the Jerry Spinelli book they were reading in class. The following is an example of dialogue from the very brief discussion:

1 T: Character traits? Who remembers [what these are]?
2 S1: Something about a person, about their personality, something they like to do.
3 S2: Description of someone.
4 T: Does that sound good to you Joanne?
5 S3: Don't know about that second one [a description of someone].
6 T: (Gives hair color examples.)
7 T: "Tina is a freak." Is that a description? (Teacher is referring to one of the students in the class, Tina.)
8 S4: No, but it's how she acts.
9 S5: Rosie is two-faced. That's a trait. (Points to Rosie, the girl sitting across from S5.)

Following the discussion, Mrs. Smith handed out the lesson activity—a list of adjectives for students to define with a partner (e.g., "bright," "honest," "prim," "ugly," "pretty," etc.), using a dictionary as needed. Students spent the large majority of the class period working on this activity. For the final ten minutes of the class period Mrs. Smith reviewed with the students the definition of each word.

This lesson received illustrates scores (1's) overall for the quality of the class discussion. Students were not asked academically challenging, rigorous questions, and were not pressed to provide elaborated responses or explain their thinking. The discussion did not focus on the text the students were reading—for example, analyzing the qualities or traits of the characters from the Jerry Spinelli book—nor did they deal substantively with the difference between a character trait and a physical description.

The lesson activity also illustrates a '1' for rigor. While developing vocabulary knowledge is important for furthering comprehension and writing skills, defining words out of context is a poor way to build students' capacity in this area (see, for example, Beck, McKeown, & Kucan, 2002) and the words given to students to define were not rigorous for sixth grade. Additionally, the task did not build on, or deepen, students' knowledge of the text they were reading, or require any degree of writing or elaborated response.

Finally, the teacher did not communicate her expectations for high quality work to students. In other words, she did not discuss with the students what she was looking for in their work, or what they would need to do to get a high score on the task. For this reason, this lesson also received low scores on the rubrics measuring the quality and communication of the teachers' expectations for student learning.

High-quality math lesson. This lesson received 3's and 4's on the IQA rubrics. Mr. Green's class started promptly. The first 12 minutes of class time was spent reviewing a previous lesson. Students took turns reading the problems aloud and explaining how they came to their answer. For example:

1 Student raises hand and reads problem 1.
2 T: What are the facts, and what's your plan of attack?
3 Student states relevant facts.
4 T: What was your plan?
5 S: 1.80 x 10.
6 T: Why did you do that?
7 S: Explains.
8 T: What does the 10 represent?
9 S: Explains.
10 T: [Another problem] Tell me what you did to solve it?
11 S2: Explains.
12 T: Answer? 15. How many people got 15? Anyone do it differently?
13 S3: I did [explains how arrived at answer].

Teacher records students' solution and shows how it's connected to the previous solution.

14 S4: I did $1.50 and added it 10 times.

Teacher explains that multiplication is the same as repeated addition, records solutions on board and shows how the two strategies are linked.

Discussion continues with students volunteering their answers and how they arrived at their solutions and the teacher recording strategies and showing how they are related.

Teacher then passes out rubric and asks students to grade themselves in the remaining 3 minutes of the activity.

After completing this activity, Mr. Green asked a student to read the "Baking Brownies" problem aloud from their textbook. He then asked:

15 T: Who are the 3 kids?
16 S: [Answers]
17 T: How many people are gonna be at the camp?
18 Ss: 240
19 T: Will this recipe feed all those people?
20 Ss: No.
21 T: [Directs students to the recipe in the text book] Should the kids make small, medium, or large brownies, and why?
22 S5: I think small ones because [you] need to make 8.
23 T: That's interesting, why 8?
24 S5: 8 x 30 equals 240
25 T: Interesting, anyone think differently?
26 S6: I think medium. [You would] need 12 pans, 12 X 20 equals 240

Teacher restates students' answer and connects to repeated addition concept discussed earlier.

27 T: Anyone think large?
28 S7: To find large, all you do is double it because…they're cut in half.

The class discussion continued along these lines. After a short while, Mr. Green broke the students into groups of four and introduced the lesson activity:

29 T: Let me tell you exactly what I want you to do. Decide with your group if you want to make small, medium, or large brownies. Okay, so now chose a size. Determine first, how many batches. After you determine size and batches the recipe tells you exactly how much you'll need [of each ingredient] to feed the whole camp. If making 10 batches, how many eggs?
30 S: Ten.

The students went right to work while Mr. Green circulated around the room. One group had a disagreement about the size of brownies. Mr. Green intervened, and asked the class if anyone was doing “small brownies.” He then reassigned the student to a small brownie group. Everyone got right back to work. After a timer sounded (signaling the end of the activity), he asked students to rearrange their desks from small groups back into rows. The students did this very quickly and efficiently. Mr. Green then said:

31 T: Let’s review what we’ve done so far…who did small [batches]? How many batches did you decide to make?
32 S: 8
33 T: So you’re telling me that 8 x 30 equals 240?
34 T: [To students who chose medium sized brownies] How many batches?
35 S: 12
36 T: So you’re telling me that 12 x 12 equals 240?
37 T: I had the pleasure of working with the group making large brownies. How many batches?
38 S: 15
39 S: 16
40 T: Why does 15 x 15 not seem sensible? What digit is this going to end with?
41 S: 5
42 T: So I think you goofed up. How many batches?
43 S: 16
44 T: 15 x 16 equals 240. Anyone see anything weird? So is that right?
45 Ss: Yes
46 T: Now the eggs. How many for one batch?
47 Ss: One egg equals one batch
48 T: So if you want to make 12 batches how many eggs do you need?
49 Ss: 12
50 T: Pretty simple. Sugar equals 1/4 cup for one batch. How do you find [the answer]?
51 S: Multiply by 12
52 T: Even if you don’t want to multiply you can do repeated addition. [Demonstrates this on the board for the class]

This lesson received high scores overall for the quality of instruction. Mr. Green consistently pressed students to explain their reasoning without actively endorsing a single strategy for finding the correct answer (e.g., turn 23, “That’s interesting, why 8?” and turn 50, “How did you find [the answer]?”) and at least a few times in the lesson showed how students' contributions (solutions) linked to one another. The quality of the lesson activity also received a high score because it required students to solve a genuine, challenging problem and to make conjectures and support their conclusions with mathematical evidence. Mr. Green encouraged students to find alternative ways to solve problems, consistently showing the links between strategies—multiplication and repeated addition (e.g., turn 15, "How many people got 15? Anyone do it differently?"). He also, while not sharing models of high quality work, provided a very clear and detailed explanation regarding what they would need to do to get a high score on the task by walking the students through an example (not detailed in the field notes), and by sharing with them a rubric that had been developed for the task.

Low-quality math lesson. Mr. Brown's sixth-grade classroom, literally a few doors down from Mr. Green, illustrates a stark contrast in students' opportunity to learn. This lesson was characterized by mostly 1's and 2's on the IQA rubrics. The class began with Mr. Brown helping students arrange themselves into groups, and passing out menus to students who were noisily talking to one another. Mr. Brown then presented the following lesson activity to students (these directions were shown to students on an overhead projector):

A. Select your meal

B. List it and cost of each item

C. Total cost and sales tax

D. Total for the table

The verbal explanation he gave to students for this activity was as follows:

1 T: In order to help us make sure what we want is ready, I'm going to ask you to prepare a menu (teacher passes out sheet). Since the place is in Connecticut, Connecticut has a 6% sales tax. [I want you to write the] name of the item, add up the cost and sales tax. This is right out of Investigations 1 (there is a book on each desk for reference)…For each table you are going to total all items and give me something that looks like this (holds up work from the previous period's class). Any questions? This should take about 20 minutes, 25 minutes at the most—then I have questions for you from Investigations 1.

Teacher passes out small sheets of scrap paper to each group of students, and then reviewed the directions again.

2 T: Number 1, read thru menu, number 2 make your selection—appetizer, soup or salad, main course, dessert…then find the tax, keeping in mind that the tax is an add-on.

The teacher then began working with individual tables of students. To one group:

3 T: What is the meal total? This looks really high [Teacher rounds and adds] $7, $11, $18—so that's the total, its gonna be 6%—its not gonna be that much.
4 S: Do you take tax for each item and add them up?
5 T: You could do it that way, but I want you to total them up first and then find the tax.
6 -------
7 T: [Later to another student] That's pretty close, how did you get it [the answer]?
8 S2: [No response]
9 T: Show how you got it [walks away]
10 S2: …Used the calculator [student follows teacher, teacher says he will be right back and sends student back to his seat.]
11 T: [to another group of students] I'm suggesting that you total them up then take the tax. [A student] came up with another way to take the tax on each item, but why do it that way when you can do it my way?
12 T: [to S2] How'd you get it [the tax], did you follow our little script? [Later]
13 S3: [At another table] I don't get the 6%. [Teacher sits down with him]
14 T: Add these up first and get a total. Let's say your total is $35.95 [teacher punches total into calculator—multiplies by 1.06 to find the total with tax].

Another student approaches the teacher who doesn't know how to add the tax. The teacher again performs the procedure on the calculator.

Table of 3 boys is finished with their work. Teacher holds up packets of work from previous period and reiterates to the students that they should look like this, and should be stapled.

This lesson received a basic score (a 2) for the potential of the task, and a 1 for its implementation. Students were explicitly directed to use a single (essentially non-mathematical) procedure to solve the problem—punching a simple formula into a calculator. And when certain students struggled with doing this, the teacher did the procedure for them rather than helping them figure out how to do the problem on their own. Students were not required to make connections to the concepts or procedures underlying the procedure being used, and in fact, as illustrated in turns 5, 11, and 12, the teacher actively discouraged students from exploring alternative ways to solve the problem (e.g., "I'm suggesting that you total them up then take the tax. [A student] came up with another way to take the tax on each item, but why do it that way when you can do it my way?"). It is notable that this activity was from *Connected Math* and had the potential to be quite rich. The teacher altered the task, reducing its level of rigor from the published version, and reduced the level of cognitive demand still further in the implementation.

The expectations for the task expressed by the teacher focused solely on procedures—for example, that students staple their paper—rather than on what "good" work should look like. Finally, while the teacher went around and spoke with individual groups, the class did not engage in a whole class discussion following the task (though we had asked to see a discussion and a problem-solving activity), which potentially would have been a chance for students to share and discuss their work.

It is notable as well that a tremendous amount of instructional time was wasted. About a half hour into the fifty-minute class period many students had completed the task. Some of these students left the classroom for the library. Other students (who had or had not finished their work) remained in the classroom noisily talking or running around. Some students began leaping off the tables. For these reasons—the focus of the task on enacting a single procedure, the lack of discussion, and focus of the teachers' expectations on procedures rather than the content of students' work, and the general chaos of the room—this lesson received low scores overall for the quality of instruction.

High quality ELA assignment. This seventh grade assignment illustrates the highest ratings (4's) on the IQA rubrics. Students wrote an analytic essay based on Shakespeare's "Macbeth" choosing from the following writing prompts:

A. What do you think is Macbeth's tragic flaw? In other words, what is the defect in his character or personality that causes him to do the things he does (murder the king, murder the guards, have his best friend and the family of another friend killed…). He is not a purely evil man, but a good man who has done horrible things. What is it about him that made him capable of such horrible deeds? Is there a lesson to be learned by the events in this play?

B. How does fate play a role in the tragedy of Macbeth? How much of what happens is caused by Macbeth himself? How much is caused by the witches and their prophecies? Who/what is really responsible for what happens in the play?

C. What roles do guilt and conscience play in the tragedy of Macbeth? How do they affect the characters and the action of the play? Is there a lesson to be learned by how they work in the play?

D. "Don't grab, don't say 'gimme,' don't cut in line." (Tara, age 7). What is the wisdom behind this little girl's statement? How does this relate to the tragedy of Macbeth? Is there a lesson to be learned by the events in the play?

E. According to Shakespeare in Macbeth, what does it really mean to be a man? Explore the various definitions of manhood in the play as you respond to this big question.

Prior to beginning the assignment, the students were provided with a rubric, and a criteria sheet describing the elements of an "excellent" essay, and model