Using the Instructional Quality Assessment Toolkit to Investigate the Quality of Reading Comprehension Assignments and Student Work

CSE Report 669

Lindsay Clare Matsumura, Sharon Cadman Slater, Mikyung Kim Wolf,
Amy Crosson, Allison Levison, Maureen Peterson, Lauren Resnick
LRDC/University of Pittsburgh

Brian Junker
Carnegie Mellon University

January 2006

Correction in Authorship: Versions of this report prior to September 30, 2006 contained a cover page listing Lauren Resnick and Brian Junker as authors, with a full list of correct authors on a subsequent page. This error was corrected on September 29th. The complete list of contributing authors is: Lindsay Clare Matsumura, Sharon Cadman Slater, Mikyung Kim Wolf, Amy Crosson, Allison Levison, Maureen Peterson, Lauren Resnick, and Brian Junker.

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
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, Strand 2: Development and Testing of Classroom Practice Indicators. Lauren Resnick and Brian Junker, LRDC, University of Pittsburgh/CRESST Project Directors

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.

The findings and opinions expressed in this report are those of the author(s) and do not necessarily reflect the positions or policies of the National Center for Education Research, the Institute of Education Sciences (IES), or the U.S. Department of Education.

USING THE INSTRUCTIONAL QUALITY ASSESSMENT TOOLKIT TO INVESTIGATE THE QUALITY OF READING COMPREHENSION ASSIGNMENTS AND STUDENT WORK

Lindsay Clare Matsumura, Sharon Cadman Slater, Mikyung Kim Wolf, Amy Crosson, Allison Levison, Maureen Peterson, Lauren Resnick

LRDC/University of Pittsburgh

Brian Junker

Carnegie Mellon University

Abstract

This study presents preliminary findings from research developing an instructional quality assessment (IQA) toolkit that could be used to monitor the influence of reform initiatives on students' learning environments and to guide professional development efforts within a school or district. This report focuses specifically on the portion of the IQA used to evaluate the quality of teachers' reading comprehension assignments and student work. Results are limited due to a very small sample of participating teachers (N = 13, 52 assignments), and indicate a poor to moderate level of inter-rater agreement and a good degree of consistency for the dimensions measuring academic rigor, but not the clarity of teachers' expectations. The rigor of the assignments collected from teachers also was associated with the rigor of observed instruction. Collecting four assignments (two challenging and two recent) from teachers did not yield a stable estimate of quality. Additional analyses looking separately at the two different assignment types indicate, however, that focusing on one assignment type would yield a stable estimate of quality. This suggests that the way in which assignments are collected from teachers should be revised. Implications for professional development are also discussed.

The primary goal of the Instructional Quality Assessment (IQA) toolkit is to develop a set of measurement tools that provide a rich picture of instructional quality, and have the potential to serve as a learning tool (Shepard, 2000) for district

and school personnel (principals, teachers, etc.). Without sacrificing richness, however, the IQA also is intended to be a toolkit that is reasonably parsimonious to use. In other words, our goal is to create a set of measures that can be used to assess instructional quality within a reasonable period of time and at a reasonable cost. These somewhat competing goals form the central challenge of our project: How can a measure of instructional quality be created that is "rich" in content and at the same time "lean" enough to be used in large scale research and evaluation studies?

To address this challenge, we measured instructional quality from multiple perspectives. First, teachers were observed once in their classroom. This observation was intended to provide, among other things, insight into students' opportunities to develop their academic language skills and engage with rigorous content material. A single observation, however, does not reveal enough about a teacher's classroom practice to be considered an adequate assessment of instructional quality. We decided, therefore, to also collect from teachers a sample of assignments reflecting student work in order to gain a more multi-faceted perspective on the quality of instruction.

This paper describes our work so far in looking at the quality of the assignments we collected from teachers. This work was conducted in two content areas: reading comprehension and mathematics. The purpose of this paper is to describe the work undertaken in reading comprehension and, specifically, students' responses to literature.[1] While other aspects of reading comprehension instruction also are important to study (e.g., for example, the support students receive to decode text, develop their vocabulary, etc.), we focused on students' responses to literature, as this would be a likely area or genre for them to demonstrate higher-level academic skills.

The first part of this paper describes the research and theories that underlie the development of our rubrics, or how we went about determining the degree to which an assignment task supports students' engagement in meaningful, challenging work. Preliminary results from a small pilot study then are described that focus on the technical quality of these rubrics. Assignment quality also is investigated from a more qualitative perspective to look at the degree to which our rubrics may capture important differences in students' opportunities to learn and to help explain the statistical findings. Specifically, the following questions are addressed:

[1] Our work conducted in mathematics is described in Boston & Wolf (2004).

1. How reliable and independent are the classroom assignment rating scales?
2. How many assignments and raters might be needed to obtain a stable estimate of the quality of classroom practice?
3. What is the relation of the classroom assignment ratings and observed instruction?

Assignment Tasks as an Indicator of Instructional Quality

"People learn by doing" is an old and familiar maxim. A more up-to-date version, informed by 30 plus years of research on learning and instruction, adds a role for teachers: "People learn by doing with guidance and assistance." This view of learning is rooted in the theoretical work of Lev Vygotsky and is supported by research focused on children's development across diverse cultures. This body of research indicates that children learn skills—such as weaving, sewing, cooking, etc.—by jointly participating in activities with an adult (or other more capable peer). These adult mentors "scaffold" children's participation in the activity by orienting children to the overall goals of a task, breaking the activity down into manageable parts, and focusing children's attention and actions on the steps required to complete the activity. Adult mentors also support and guide children's participation in an activity by demonstrating and modeling the act to be performed, and "marking critical discrepancies between what the child has produced and the idealized version of the activity" (cited in Rogoff, 1990, p. 94). Through engaging children in the appropriate handling of a task, adults "create situations in which children can extend current skills and knowledge to a higher level of competence" (Rogoff, 1990, p. 93). In other words, adults open "zones of proximal development" for children by allowing them to do with assistance what they would not be able to do on their own (Vygotsky, 1978).

In order to become powerful abstract thinkers and consumers of texts, students need the opportunity to participate in social interactions where analytical and abstract thinking is modeled for them and where they have the opportunity to practice their emerging skills in this area. Verbal interactions, or classroom conversations are of critical importance for providing students with the opportunity to be exposed to the modeling of these types of skills, as well as to practice their

thinking and reasoning skills and get immediate feedback on their efforts. Nevertheless classroom conversations alone, however excellent they may be, are not enough to develop students' comprehension skills. Students also need the opportunity to apply their newly emerging thinking and reasoning skills in written forms as well. This is important for monitoring student learning. Additionally, students who have not had the opportunity to develop these skills (the ability to write about text in meaningful ways) are at a distinct disadvantage academically—a disadvantage that limits their chances of being successful in high school, being accepted to a college, and completing college-level course work.

Assignment tasks provide insight into the level and type of support (or scaffolding) a teacher provides to students, and so can be an important source of information for assessing students' opportunity to learn academic skills. Assignment tasks can be a window into the degree to which a teacher makes a task accessible to students (e.g., breaks down the steps of a task or provides explicit directions for how to complete each step); communicates performance expectations (e.g., demonstrates an idealized version of the act to be performed); and provides feedback to students on their efforts (e.g., marks critical features of discrepancies between what a child has produced and the ideal solution).

Assignment tasks can also provide insight into students' opportunities to learn skills and content that are germane to a specific discipline. As described earlier, children learn through joint-participation in real activities. They learn to weave by weaving (with assistance), and to cook by cooking (with assistance). By the same token, children learn academic skills by being assisted to engage in the work of real scholars. This could mean using mathematics to solve real world problems that contain multiple solutions (as a scientist or an engineer might). This could also mean synthesizing, analyzing, interpreting, and evaluating information from texts (as would be required in college).

The question is, how does one determine the degree to which an assignment supports and guides students to develop higher-level academic skills? What would one look for in an assignment to indicate that students had been exposed to a high-quality classroom learning environment?

Defining Assignment Quality

To answer this question we drew in part on research investigating best practices for teaching reading comprehension and research investigating assignment quality (e.g., Clare & Aschbacher, 2001; Newmann, Bryk, & Nagaoka, 2002; Storms, Riazantseva, & Gentile, 2000). The following sections provide a brief overview of this research.

Some elements of effective reading comprehension instruction. The ability to comprehend text is a very complex process. In order to become proficient readers students have to master a number of interrelated skills. These include the ability to construct mental models of a text at various levels; for example, understanding how clauses are related, or how events in a text are temporally sequenced. Besides understanding the specific events or ideas represented in a text, however, proficient readers also are able to construct meaning beyond what is represented on the written page. In other words, they are able to apply higher-order thought processes to infer meaning beyond the surface-level features of the text.

Borrowing from Bloom (1956), the complexity of thought processes one could use to infer meaning from a text could be described in three general levels (described in Snow, 2002, p. 109). At the lowest level are *recognition* and *recall* or the ability to identify specific content verbatim and to reproduce (remember and retrieve) specific content that was explicitly mentioned in a text. The second level is termed *comprehension* and includes the ability to generate a mental model of a text by summarizing, paraphrasing, explaining, or translating a text. At the highest level of complexity are the *application* of knowledge from a text to solve a problem not mentioned in the text, and the *analysis* of a text into its constituent parts that are linked back to each other in new ways. The ability to *synthesize* or construct new patterns or structures from the events in a text, and the *evaluation* of a text based on an external criteria or standard also are considered to be high-level thinking skills.

Effective reading comprehension instruction supports students to answer higher-level questions about a text (Snow, 2002). In addition to understanding the surface level features of a story (i.e., constructing a mental model of the events), effective reading comprehension instruction provides students with an opportunity to construct meaning beyond what is represented on the page. Instruction of this type guides students to analyze, synthesize, evaluate or apply knowledge from a text in the service of more deeply comprehending what they read.

Another element of effective reading comprehension instruction concerns the curricular materials used by a teacher. Ideally, students would be exposed to a curriculum that is in-depth and challenging, and exposes students to a wide variety of genres (Snow, 2002). This includes having students read texts that contain themes or ideas that are complex enough (or illustrate sufficient "grist") to support meaningful writing topics and classroom discussions (Beck, McKeown, Hamilton, & Kucan, 1997). Such texts could convey information to students about other places, times, or cultures. These texts also could contain interesting dilemmas where there is no obvious right or wrong answer, or other themes that broaden students' thinking. Grist could also be evidenced in the writer's craft, for example, in the language use, vocabulary and organizational structures employed by an author.

Finally, effective reading comprehension instruction exposes students to a wide variety of intentionally applied comprehension strategies (National Reading Panel, 2000). Specifically, effective teachers guide and/or model for the reader the actions that the reader needs to take to improve comprehension. S/he does this by clearly explaining the reason for the task (and standards for completing the task), breaking the task down in smaller parts for their students, and activating prior knowledge. Teachers then have students practice these strategies and provide them with assistance and feedback on their performance until the student internalizes the skill and is able to independently carry out the comprehension task on their own. This type of explicit instruction appears to be especially beneficial for lower-achieving students (Snow, 2002, p. 33).

Research in assignment quality. Most of the research on assignment quality has included a focus on students' opportunity to apply higher-level thinking skills. For example, Stein, Smith, Hennigsen, and Silver (2000) considered mathematics tasks that required students to recall information only, or apply an algorithm or procedure without any reference to an underlying mathematical concept to represent a lower level of cognitive demand. Tasks that had students apply procedures and engage with the underlying conceptual ideas, or that required students to apply complex problem-solving (i.e., non-algorithmic) thinking were considered to represent a higher level of cognitive demand.

With specific regard to English language arts, two separate efforts comprise the bulk of the research on assignment quality: studies conducted by the Chicago Consortium for Quality Schools, and the National Center for Research in Evaluation, Standards and Student Testing at UCLA. Fred Newmann, Anthony Bryk, and their

colleagues in Chicago defined high quality assignments as "authentic intellectual work." They operationalized the characteristics of authentic intellectual work in three scales. The first scale, construction of knowledge, focused on the extent to which an assignment required students to organize, interpret, evaluate, or synthesize prior knowledge to solve new problems. The second scale, disciplined inquiry, focused on the extent to which the assignment required students to use a prior knowledge base, strive for in-depth understanding, and express their ideas with elaborated communication. The last scale, value beyond school, focused on the applicability of the task to life outside the school setting (Newmann, Bryck, Nagaoka, 2001).

Similar to Newmann and Bryk's work, (Clare) Matsumura (the first author of this report) and her colleagues at CRESST looked at assignment quality in terms of its level of cognitive challenge, or the degree to which students had the opportunity to apply higher order reasoning, engage with academic content material and produce extended responses. They did not consider the applicability of a task to contexts outside of schools. They also looked at how clearly a teacher articulated the specific skills, concepts or content knowledge students were to gain from completing the assignment in order to ascertain teachers' intentions for a task (specific learning versus activity for activity's sake). The clarity and specificity of the grading criteria used to assess students' work, and the alignment between the learning goals and the assignment task, and the learning goals and the grading criteria, were considered as well. The purpose of these dimensions was to produce more diagnostic information about assignment quality that could be used to guide professional development efforts. In other words, to consider at what point in the assignment activity (e.g., the conception, implementation, assessment of student performance, etc.) teachers might need additional support (Clare, 2000; Clare & Aschbacher, 2001; Matsumura, Garnier, & Pascal, 2002; Matsumura, Garnier, Pascal, & Valdés, 2002).

Results from both projects indicated that students produced higher quality work and scored higher on standardized tests of achievement when they were exposed to higher-quality assignments (see Clare & Aschbacher, 2001; Matsumura, Garnier, Pascal & Valdés, 2002; Newmann, Bryk, & Nagaoka, 2002). This result was arrived at after controlling for students' SES, ethnicity, language status and prior level of achievement. These results supported the decision to include measures of assignment quality in the IQA toolkit as it appears that teachers' assignments have

the potential to yield important information about the quality of classroom practice that is associated with differential student achievement.

IQA assignment quality dimensions. The IQA rubrics are structured around the Principles of Learning (Institute for Learning, 2002). These principles are a comprehensive, standards-based framework that includes both instructional processes and the external supports intended to support high-quality teaching and learning. They are comprised of nine interrelated constructs.

We focused on two Principles of Learning in developing our rubrics that we believed would be most proximal to assignment quality. These are *academic rigor in a thinking curriculum* and *clear expectations*. The principle of academic rigor holds that student success depends on their exposure to a rich knowledge core that is organized around the mastery of major concepts. This curriculum also should provide students with the regular opportunity to pose and solve problems, formulate hypotheses, justify their reasoning, construct explanations, interpret text, and test their own understanding. Additionally, students should have the opportunity to construct their own understandings of concepts based on the synthesis of several sources of information including their experiences outside of school.

The second principle of learning upon which we based on our work, clear expectations, holds that students need to have access to the performance expectations for their work. Teachers can communicate these expectations to students by posting or distributing standards and rubrics or by discussing with students the criteria for work that meets a specific standard. Providing students with models of high quality work that outline a sequence of expected concepts and skills students are to master in the process of accomplishing a larger standard, and discussing these models with students are also important for communicating expectations to students. Other important means for making expectations clear to students include involving students in judging their own work with respect to the standards, and communicating to parents what students are supposed to accomplish.

For the IQA toolkit these principles of learning were operationalized in five rubrics and a checklist for evaluating assignment quality. Specifically, to assess the degree to which an assignment promoted academic rigor in a thinking curriculum we looked at the rigor of the text used for an assignment in terms of the complexity

of its themes and content. Similar to the research on assignment quality described earlier, we also looked at the degree to which an assignment provided students with an opportunity to develop their analytical and interpretation skills, and engage with the deeper meanings of a text (i.e., go beyond describing surface-level details). The degree to which students were supported to realize the potential of the task during implementation, that is, that the collection of student work evidenced that students had analyzed and interpreted the deeper meanings of the text and supported their responses with evidence, also was assessed. Additionally, we looked at the rigor of a teacher's expectations for the quality of student work. By this we meant the degree to which a teacher's expectations (expressed in his/her grading criteria or assignment directions) supported students to apply higher-level comprehension skills and supported their responses with extensive evidence from a text.

To assess the degree to which a teacher communicated clear expectations to students regarding the quality of their work, we first considered the specificity and amount of information a teacher provided to students for what they would need to do to successfully complete the task. We also considered the teacher's efforts (based on self-reported information) to ensure that all students had access to the performance expectations for a task.

The following sections describe the results of a small pilot study investigating the technical quality of these rubrics in terms of interrater reliability, stability of the assignment ratings and relation to observed instruction. These results must be interpreted with a great deal of caution, however, as they were based on a very small sample of teachers. These analyses have utility, however, for providing information that will be used to guide future development work. Additionally, variation in assignment quality is explored from a more qualitative perspective to take a closer look at the degree to which our ratings capture meaningful distinctions in students' opportunity to develop higher-level academic thinking and writing skills, and to better understand our findings from the statistical analyses.

Methods

Sample

Second- and fourth-grade teachers (N = 30) were recruited from 11 elementary schools across two demographically similar school districts.[2] Of these 30 teachers, 14 participated in the reading comprehension portion of the study, and 13 of these teachers turned in assignments with samples of student work. These schools served a diverse population of students (26% African American, 6% Asian, 47% Latino, 15% White, 6% other) 20% of whom were English language learners. Teachers who participated in the study had been teaching for an average of 14 years, and had been at their school an average of 4 years.

Procedures

District personnel suggested schools that might be interested in participating in the study. A member of the IQA research team then contacted the principals of these schools. Principals who were interested in participating were asked to explain the study to all of the second- and fourth-grade teachers at their school.

In April 2003, a member of the IQA research team visited each school to discuss the study with interested teachers and distribute the assignment collection materials. Teachers were asked to submit four reading comprehension assignments—two recent assignments and two assignments they considered to be challenging for their students (N = 52 assignments). For each assignment teachers filled out a two-page cover sheet describing the assignment task, their assessment criteria for grading student work, and how they shared these criteria with students. Teachers also submitted six samples of student work for each task—two samples of work they considered to be of high, medium and low quality, respectively. The assignments were collected later in May when their classroom was observed. Teachers were given $100 gift certificate as a token of appreciation for completing the assignment coversheets and assembling the samples of student work.

The assignments were rated by graduate students who were recruited to participate in the data collection and were not part of the team who developed the rubrics (N = 2). We hired "naïve" raters in order to assess the quality of our rating

[2] One third-grade teacher and one fifth-grade teacher were recruited as well.

training program and rubrics as evidenced by the degree to which people that were external to the project could agree on the different ratings. The assignments were also rated by members of the IQA development team (N = 4), though these ratings were not included in the analyses reported here. The assignments were randomly ordered for scoring and were rated independently by each of the naïve raters (N = 52 assignments). The dimension measuring the academic rigor of the text was rated a few weeks later by these same raters, after the research team located the relevant texts. Because of the difficulty locating the books or articles, it was possible to rate this dimension for only 37 assignments.

Measures

Assignment quality is assessed on a four-point scale (1 = poor, 4 = excellent) for the following dimensions, with the exception of the dimension measuring the rigor of the text which is instead assessed on a three-point scale (1 = poor, 3 = excellent):

- Rigor of the Text – The purpose of this dimension is to measure the degree to which the text that is the focus of a reading comprehension assignment contains literary or informational content that is complex and engaging enough to warrant extended writing. Additionally, this dimension considers the richness and variety of the language (vocabulary and sentence structures) in the text. To receive a high score on this dimension, a text would have to contain a complex plot or elaborated information, and the text would have to contain rich or highly specific vocabulary.

- Potential of the Task – The purpose of this dimension is to describe the degree to which an assignment provides students with an opportunity to develop their analysis and interpretation skills, and to engage with the deeper meanings of a text. Specifically, this dimension considers the extent to which students are supported to apply higher-level skills in the service of deepening their comprehension of a text, as opposed to recalling, describing, or identifying basic information. To receive a high score on this dimension, students would be required to go beyond surface-level description, detail, or theme identification, and to engage with subtle nuances of the text or the overarching or larger significance of the work (e.g., discussion of story themes) with the opportunity to develop and elaborate their ideas. Additionally, the task would require students to provide evidence from a text to support their ideas.

- Implementation of the Task – The purpose of this dimension is to describe the degree to which students are supported to realize the potential of the task during implementation. To receive a high score on this dimension, the collection of student work would evidence that students analyzed and interpreted the deeper meanings of the text, and that students provided extensive evidence for their positions. Additionally, the collection of student work would demonstrate that students were supported to develop and elaborate their ideas through extended written response.

- Rigor of Expectations – The purpose of this dimension is to describe the degree to which a teacher's expectations for the quality of students' work support students to analyze and interpret the deeper meanings of a text. An assignment that received a high score for this dimension would focus on students' attainment of these higher-level skills.

- Clarity and Detail of the Expectations – The purpose of this dimension is to assess the specificity and elaborateness of a teacher's expectations for the quality of students' work for the assignment task. A high score for this dimension would indicate that a teacher provided a great deal of information to students for what they would need to do to successfully complete the task. Each of the teacher's criteria for success would be clearly articulated, and within these criteria, detail would be provided for the varying levels of success (e.g., what a student would need to do to get an A, a B, etc.).

In addition to the five-point scales, teachers also completed a checklist reporting how they shared their expectations with students (e.g., discussed criteria in class, posted criteria charts, shared models of high quality work, etc.).

As shown in Table 1, on average, the assignments we collected were considered to be of fair quality (i.e., were rated a '2' on a four-point scale) on all of the dimensions. The exception to this was the dimension measuring the rigor of the text. This dimension was assessed on a three-point scale, so a mean score of 2.38 indicates a somewhat higher level of quality than the other dimensions.

Table 1

Description of reading comprehension assignments (N = 52)

AR Dimension	Mean	SD	Range
Academic rigor of the text*	2.38	0.72	1-3
Potential of the task	2.31	0.98	1-4
Implementation of the task	1.79	0.89	1-4
Rigor of the expectations	2.29	1.07	1-4
Clarity of the expectations	2.39	1.02	1-4

*Note: N = 37 for this dimension, as it was not possible to rate the rigor of the text for every assignment. Also, this dimension was assessed on a three-point scale, as opposed to a four-point scale.

Teachers for approximately half of the assignments (51.9%) reported that they discussed their criteria for high quality work with students in class and shared models of high quality work with them in advance of their completion of the assignment. For slightly more than a quarter of the assignments (26.9%) the teachers reported that they discussed their criteria for high quality work with the students, but did not provide them with models of high quality assignments. For nearly a quarter of the assignments (21.2%), teachers reported that they did not share their criteria for assessing students' work with their students.

Analyses

Descriptive statistics were used to characterize the teachers' assignments. Cohen's kappa coefficients were calculated to investigate the level of agreement between five raters on each dimension when controlled for chance agreement. Cronbach's alpha coefficients were calculated to estimate the consistency of these ratings at the teacher level. Correlations also were computed to measure the strength of agreement between the rater pair.

Generalizability studies were conducted to investigate whether our design yielded a stable estimate of quality and decision studies were conducted to explore options for future research design. Correlations also were computed at the teacher

level to investigate the interrelationship of the assignment ratings, and the relation of the assignment ratings to observed instruction.

Results

The results of this study (limited by the small sample size) are presented in the following sections organized by each of the research questions.

How reliable and independent are the classroom assignment rating scales?

To address the first part of this question we investigated the interrater reliability of the rating scales, the degree to which different people can independently look at the same phenomenon (in this case, teachers' assignments) and agree on a score. The percent agreement between raters was calculated on assignment ratings within each grade level. Results indicated that there was a fair level of agreement between the two raters who scored the assignments for the dimensions measuring the academic rigor of the assignments (see Table 2). The percent agreement ranged from 63.5% to 81.1% and the correlation between raters ranged from (r = 0.81 to r = 0.83) for each of the dimensions measuring the academic rigor of the assignment task. The dimension measuring clear expectations, however, had poor inter-rater agreement (44.2%), and a relatively low correlation between the raters (r = 0.56).

Cohen's kappa coefficients were calculated to investigate the level of agreement when controlling for chance agreement. Significant kappas for each of the academic rigor dimensions indicated that the level of rater agreement was better than chance. The magnitude of the kappas ranged from 0.51 to 0.59, indicating a moderate level of agreement between the raters for these dimensions. The exception to this pattern was, again, the dimension measuring clear expectations (κ = 0.24).

Cronbach's alpha coefficients also were calculated to investigate the consistency of the ratings within each assignment for each dimension. This statistic considers the trend in rater agreement, and ranged from 0.88 to 0.91, confirming a high degree of consistency within each dimension for each assignment. The clarity of the expectations rubric showed a lower level of consistency (α = 0.71).

Table 2
Inter-rater reliability of assignment ratings for the reading comprehension assignments

Dimension	% of exact agreement	Spearman *r*	Kappa	Alpha
Rigor of the text*	81.1	0.76	0.66	0.87
Potential of task	71.2	0.84	0.59	0.91
Implementation of task	69.2	0.84	0.56	0.88
Rigor of expectations	63.5	0.83	0.51	0.90
Clarity of expectations	44.2	0.56	0.24	0.71

*Note: $N = 37$ for this dimension. It was not possible to rate the rigor of the text for every assignment

To investigate the independence of the assignment ratings, we examined the relation of the different scales to each other. Our reasons for this were twofold: First, evaluating large-scale reform efforts can be quite costly, so it is imperative that measurement tools be as efficient and streamlined as possible. We examined the interrelation of the rating scales, therefore, to reduce possible redundancy in our rating scheme by investigating whether certain scales may be so highly correlated with one another that they could be eliminated. Additionally, we were interested in looking at the interrelation of the scales within each construct/principle of learning to examine how consistent these were with one another, as well as the relationship of these constructs (academic rigor and clear expectations) to each other.

Results indicated that most of the dimensions measuring academic rigor were significantly associated with one another; specifically, the potential and implementation of the assignment tasks ($r = 0.70$, $p < 0.05$) and the potential of the task and the rigor of the expectations ($r = 0.85$, $p < 0.01$). The exception to this pattern was the dimension measuring the academic rigor of the texts read by students for the assignment. This dimension was not significantly associated with any of the other assignment quality rubrics—within academic rigor or those rubrics measuring the clarity of the expectations.

Table 3
Interrelation of assignment ratings for the reading comprehension assignments

	Rigor of the text	Potential of task	Implementation of task	Rigor of expectations	Clarity of expectations	Comm. of expectations
Rigor of the text	1	.35	.20	.40	.45	.45
Potential of task		1	.70*	.85**	.25	.51
Implementation of task			1	.49	.01	.31
Rigor of expectations				1	.58*	.78**
Clarity of expectation					1	.82**
Comm. of expectations						1

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

The two dimensions measuring clear expectations (the clarity of the expectations and the communication of the expectations to students) were significantly associated ($r = 0.82$, $p < 0.01$). Additionally, the dimensions measuring the clarity and rigor of the expectations for an assignment task ($r = 0.58$, $p < 0.05$) and the rigor and communication of the expectations to students were significantly associated ($r = 0.82$, $p < 0.01$). For the most part, however, the two constructs (clear expectations and academic rigor) did not show a high level of association (see Table 3).

How many assignments and raters would be needed to obtain a stable estimate of the quality of classroom practice?

Generalizability and decision studies were conducted to determine how many raters and assignments might be necessary to obtain a stable estimate of the quality of classroom practice. Results indicated that our design based on two raters and four

teacher assignments yielded a dependability coefficient[3] of only 0.48 (0.80 and above is considered to be acceptable).

As shown in Table 4, the Teacher x Assignment component made the greatest contribution to error variance (68%). This represents differences in judgments made about the same teacher based on the different assignments submitted. The source of variance associated with the Teacher represented far less of the total variance (18%). Individual Rater and Assignment variance components, as well as the Teacher x Rater and Rater x Assignment interactions were null or negligible. Decision study results showed that adding an additional rater did not improve estimated generalizability ($\hat{\phi}$ = 0.49) and that the estimated phi-coefficient with even 10 assignments would only be 0.68, not nearly close to an acceptable level of 0.80. In fact, results indicated as many as 22 assignments would be needed to attain an acceptable level of stability based on this design.

Table 4
Variance Components (and Percent of Total Variance) for Assignment Ratings for English Language Arts [*t x r x a* Design] Includes All Four Assignments*

Source of Variance	*Challenging and Recent Assignments*
Teacher (*t*)	1.932 (17.9%)
Rater (*r*)	0
Assignment (*a*)	0
Teacher x Rater (*tr*)	0.204 (1.9%)
Teacher x Assignment (*ta*)	7.274 (67.8%)
Rater x Assignment (*ra*)	0.040 (0.4%)
Residual (*tra*,*e*)	1.282 (12.0%)
Dependability Coefficient $\hat{\phi}$	0.48

* Negative variance components set to zero.

[3] Throughout the paper, generalizability results are expressed in terms dependability coefficients (ϕ), rather than generalizability coefficients (ϱ^2). Interpretation is generally the same, however, we chose to use dependability coefficients because we are more interested in absolute judgments, rather than relative ones, for considering teacher quality independent of the quality of other teachers.

We suspected that the large variation in assignment quality within teachers was due to the fact that we had asked for two different types of assignments: challenging and recent. To investigate this, we conducted additional generalizability studies for the two different assignment types (Slater, Matsumura, & Junker, 2005)

Table 5
Variance Components (and Percent of Total Variance) for Assignment Ratings for English Language Arts [*t x r x a* Design] Includes Two Challenging Assignments Only*

Source of Variance	*Challenging Assignments Only*
Teacher (*t*)	3.830 (37.9%)
Rater (*r*)	0
Assignment (*a*)	0
Teacher x Rater (*tr*)	0.500 (5.0%)
Teacher x Assignment (*ta*)	4.964 (49.1%)
Rater x Assignment (*ra*)	0
Residual (*tra,e*)	0.817 (8.0%)
Dependability Coefficient $\hat{\phi}$	0.57

* Negative variance components set to zero.

As shown in Table 5, the variance component results based on the two assignments that teachers considered to be challenging indicated again that the greatest source of variation was within teacher by assignment (49.1%). In other words, teachers varied significantly in the quality of the two challenging assignments they submitted. It is likely that collecting more assignments per teacher could minimize this source of variance, however. The three-way interaction term made a much smaller contribution to measurement error (8%); and the remaining sources of variance made little to no contribution to the total variance. The dependability coefficient ($\hat{\phi}$ = 0.57), is higher here than for the combined analysis ($\hat{\phi}$ = 0.48), which is the opposite of what one might expect given that these results are based on half the number of assignments (two assignments rather than four).

Next, we repeated the analysis using only the two recent assignment tasks submitted by teachers (see Table 6). In contrast to the challenging assignments, these results indicated that the vast majority of total variance was due solely to the Teacher (74%), and the three-way interaction term represented nearly all of the remaining variance (23%). The Teacher x Assignment interaction term, which in previous analyses accounted for a significant portion of the measurement error, contributed virtually nothing to measurement error; the remaining sources of variance also made little to no contribution to the total variance. The dependability coefficient ($\hat{\phi}$ = 0.90) based on these two assignments is quite a bit higher than for the four assignments combined ($\hat{\phi}$ = 0.48).

Table 6
Variance Components (and Percent of Total Variance) for Assignment Ratings for English Language Arts [*t x r x a* Design] Includes Two Recent Assignments Only*

Source of Variance	*Recent Assignments Only*
Teacher (*t*)	8.348 (73.6%)
Rater (*r*)	0.045 (0.4%)
Assignment (*a*)	0.357 (3.2%)
Teacher x Rater (*tr*)	0
Teacher x Assignment (*ta*)	0
Rater x Assignment (*ra*)	0
Residual (*tra,e*)	2.603 (22.9%)
Dependability Coefficient $\hat{\phi}$	0.90

* Negative variance components set to zero.

What is the relation of the classroom assignment ratings and observed instruction?

To address our third research question, we compared the ratings of assignment quality to the quality of a teacher's observed instruction. The purpose of this was to assess the degree to which the classroom assignment ratings yielded meaningful and appropriate information about students' learning environments that were commensurate with other measures of quality practice.

Results indicated that the degree to which students were asked to analyze and interpret text (potential of the task) and the rigor of a teacher's expectations for student work were associated with the rigor of the observed lesson ($r = 0.66$, $p < .01$ and $r = 0.60$, $p < 0.05$ respectively). Contrary to expectations, however, the implementation of the classroom task was not associated with the level of observed rigor in the observation.

A Closer Look at Assignment Quality and Student Work

We returned to our corpus of assignments and student work to better understand the differences in assignment quality within teachers for their "challenging" and "recent" assignments, and to look at whether our assignment ratings appeared to yield meaningful differences in students' opportunity to learn. The following section describes the portfolio of a fourth-grade teacher whose assignments we considered to be typical for our sample.

Ms. Smith's[4] **assignment portfolio**. For her portfolio of assignments and student work Ms. Smith, a fourth-grade teacher, submitted two "recent assignments" that were intended to build students' comprehension strategies. For the first assignment, students read an excerpt from a book about gorillas by Seymour Simon (2003) and generated a series of questions about the text. Her expectations for high quality work for this assignment were as follows:

Students were asked to jot down any initial questions that they had. Two formats were presented. Students selected which format they preferred. As students read they were expected to jot answers to questions and generate new questions.

- High performance – Students generated a reasonable number of questions (at least five) and were able to answer if the answer was present in the text. Students demonstrated high performance if they asked questions about things they didn't already know. Several questions were also critical.

[4] Pseudonyms are used to protect the identity of the participants.

- Middle performance – Students asked at least five questions. A few were obvious. Students answered the questions. Most answers were copied verbatim from the text.

- Low performance – Few questions, many questions had obvious answers. Missed many answers presented in the text.

These criteria were not explicitly shared with students, but reportedly were modeled for students during a mini-lesson that followed this assignment. The following is an example of student work considered by the teacher to be of high quality for the class for this task.

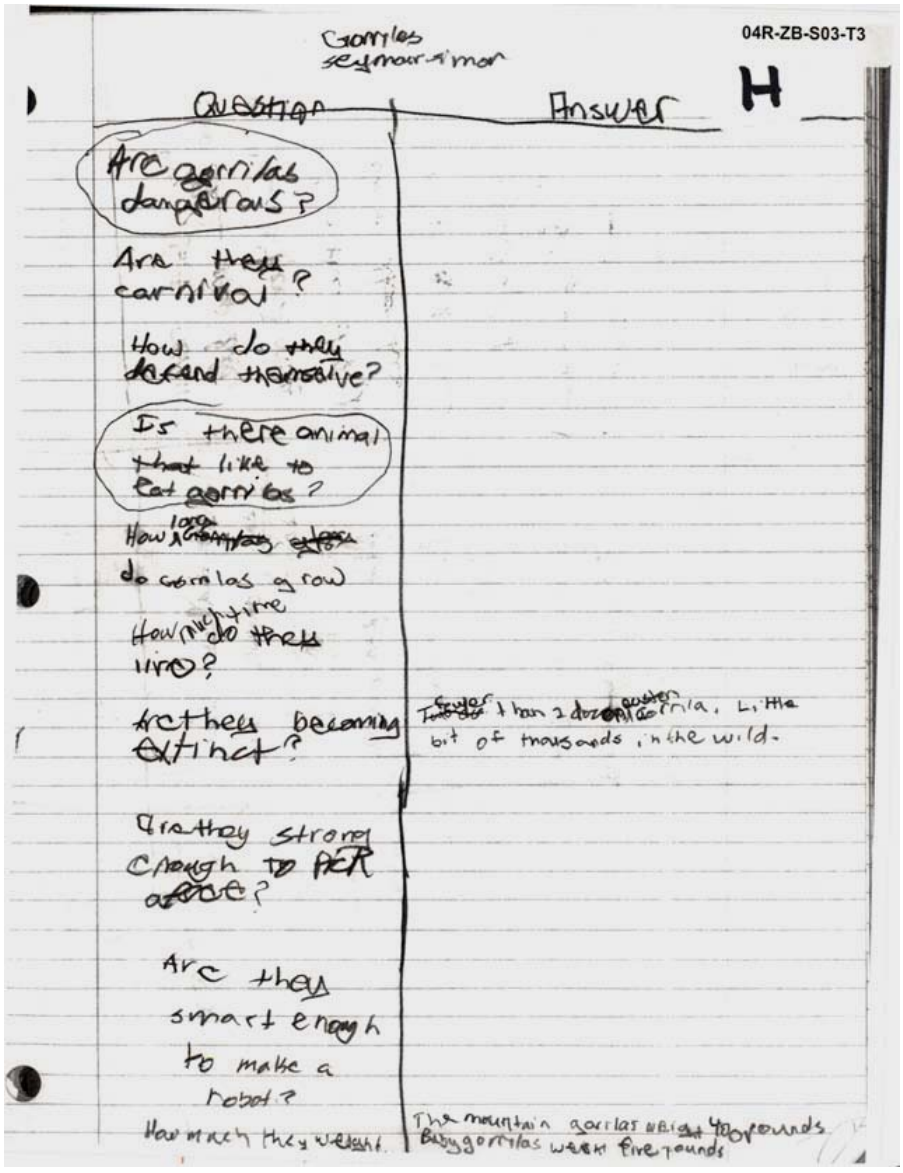

Figure 1. Example 1 of fourth-grade student work for a recent assignment.

This assignment received low ratings overall for both rigor and for the quality of the teacher's expectations. The questions generated by students were very similar (nearly identical) to one another. Furthermore, the questions generated by students, even those students whose work was considered by the teacher to be of high quality, required only basic recall of isolated facts. In contrast to the teacher's stated expectations for the task, students were clearly not guided to generate or answer questions that required them to think "critically," or even to know very much about gorillas beyond very basic, surface details (e.g., that they have five fingers and toes, 32 teeth, etc.).

For the second "recent" assignment, Ms. Smith had students "sketch what they visualized and jot a few words identifying the character's feelings" as the teacher read aloud from the text, *How Tia Lola Came to Stay* by Julia Alvarez (2002). The purpose of this was to "help generate and deepen talk" during "turn and talks" (discussions with a student partner). The teacher's expectations for this assignment were as follows:

> I explained to students that I would be looking for accuracy in their sketches and explanations. They were expected to focus on character feelings. The first sketch should focus on setting. [High performance sketches] students accurately captured character feelings and articulated feeling in words. Sketches tended to show greater elaboration of char.'s (sic) internal thinking.

The following is a sample of student work for this assignment:

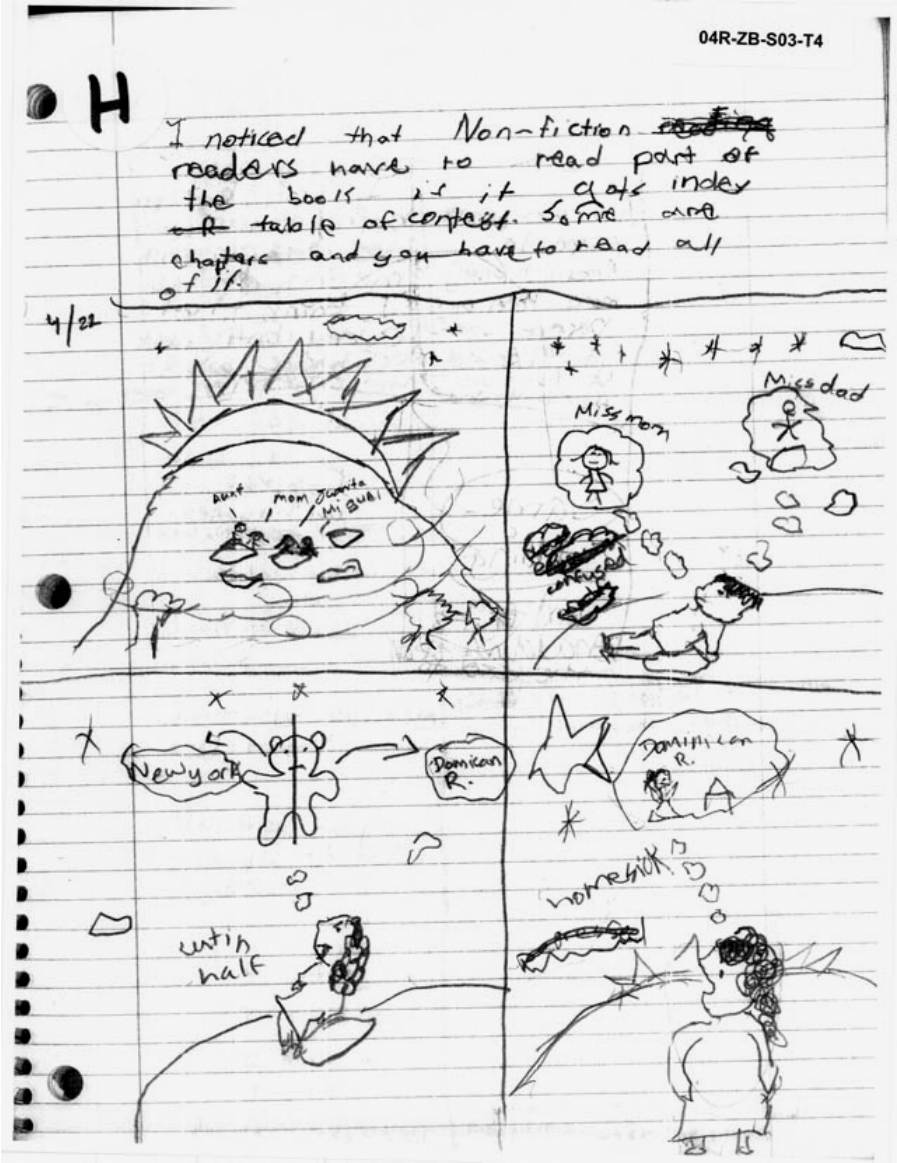

Figure 2. Example 2 of fourth-grade student work for a recent assignment.

Similar to the first recent assignment, this assignment also was considered to be of low quality overall. While the purpose of the task was to focus and spur discussion, by the time students are in the fourth-grade they should be writing more extensive responses to what they read—at least a few paragraphs. Drawing pictures with captions is a more appropriate task for students in the primary grades.

In contrast to the recent assignments, the challenging assignments (while not as rigorous as they could be) supported students to write more extensively and think more analytically about what they were reading. For the first assignment, Ms. Smith submitted letters exchanged between her and her students. These letters were

written on a weekly basis as part of the student's reader response journals. The following are the directions for the assignment she provided to the students:

> Dear Class,
>
> For the rest of the year we will write to each other about books, reading, writers and writing. Our letters will help us learn together. The letters will help you become stronger readers and thinkers.
>
> When you write your letters in your response journals do your best work and share your best thinking. For example, you might:
>
> - Tell what you noticed about the characters like how they changed or why the acted the way they did.
> - Tell about the connections to the text you made and how they help you understand.
> - Tell about the message or theme of the book.
> - Explain a theory or idea you have about the text.
> - Tell about what you liked or disliked in the book and why.
> - Write about your predictions and whether or not they came true.
> - Write about the author's craft and how it helps you to visualize as your read.
> - Write about the author's style and how it makes you feel.
> - Write about something in the text that you found interesting and surprising and why.
>
> Write a letter to me once each week. The letter is due on the day indicated on the journal list. Use letter form and please include the title and author of your book. It's also important that your letters are neat and easy to read, so I can understand your thinking. Reread your letter to make sure it's clear and makes sense.
> I can't wait to read all of your smart thinking. We are going to have a great time learning from each other!
>
> Love, Ms. S.

The following is an excerpt of student work discussing the "inside" and "outside" character traits of the book's protagonist (Arthur, an Aardvark). This letter was in response to comments the teacher had made to the student in a previous letter:

I'm wondering what you think about Arthur as a character. What is he like? For example, is he friendly, loyal, determined, a hard worker? Why do you think this? Remember to explain with examples from the Arthur books you read.

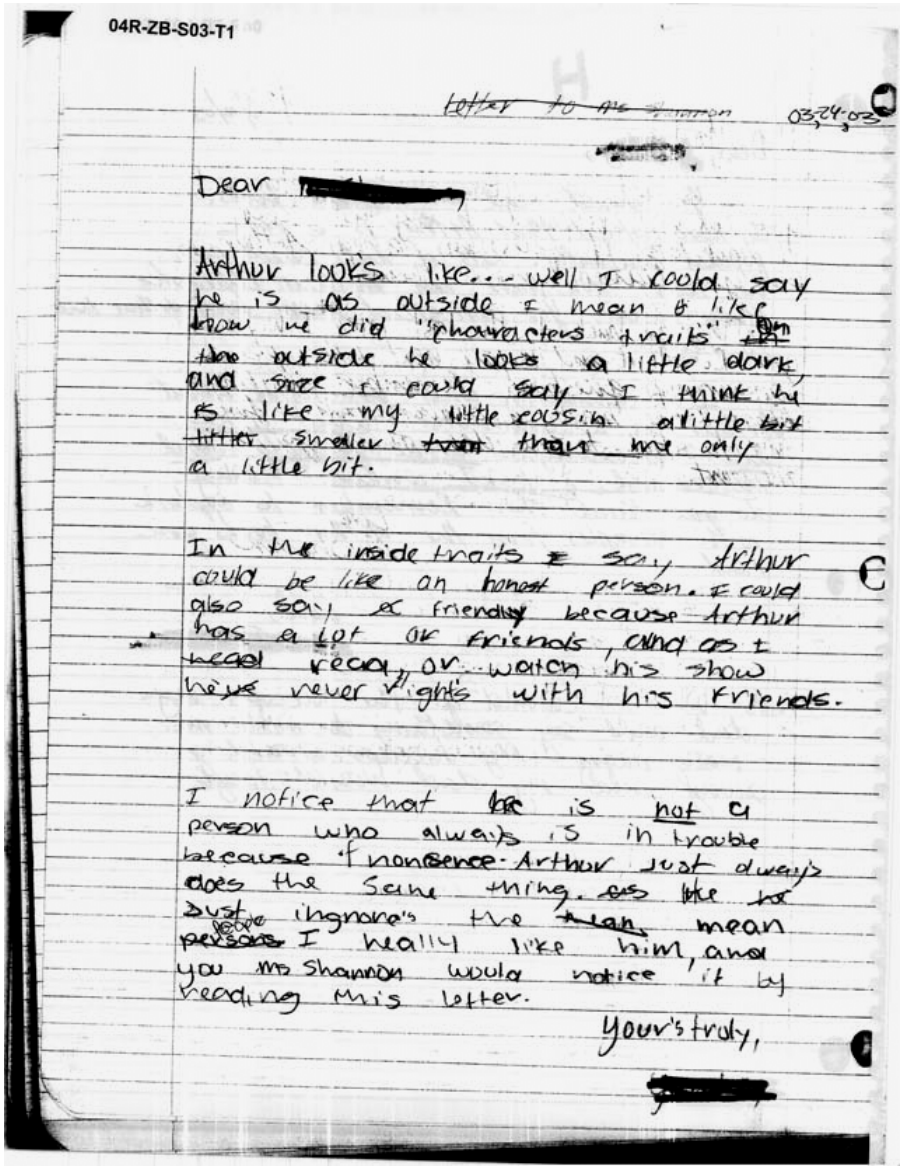

Figure 3. Example 1 of fourth-grade student work for a challenging assignment.

The second "challenging" assignment similarly focused on character analysis, and specifically, inferring character traits from textual evidence. Students chose a character and were guided to look at the character's actions, dialogue, and internal thinking and infer from this the character's "inside" traits. The following is a sample of student work (a graphic organizer) for this assignment:

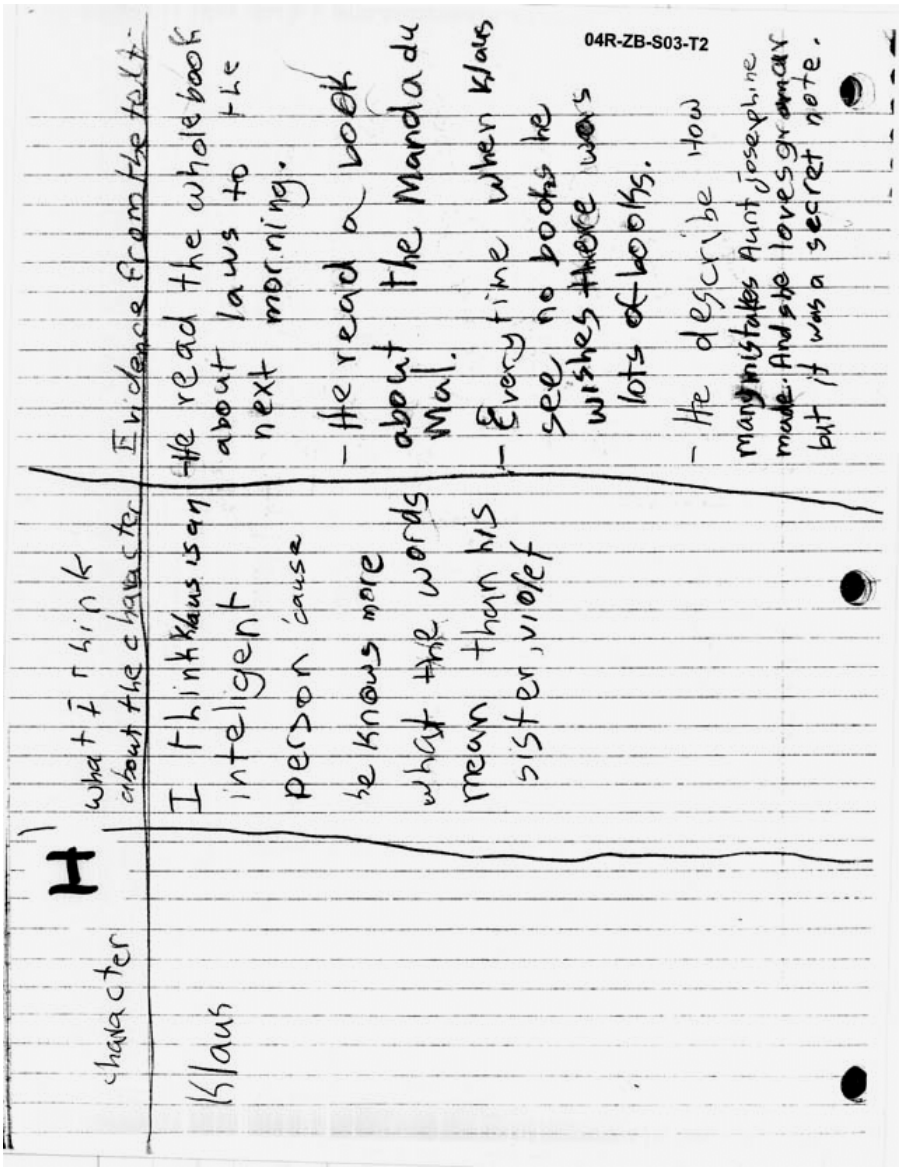

Figure 4. Example 2 of fourth-grade student work for a challenging assignment.

While neither of these assignments could be considered exceptionally rigorous for students at this grade level, they are clearly more challenging than the first two assignments in that they support students to be somewhat analytical about what they are reading and use evidence from the text to support their assertions. This can partially explain the results of the generalizability analyses. It highlights the reason for the large amount of variability between recent and challenging assignments. A puzzling finding was that the results for the analysis of recent assignments alone were so different from the analysis of the challenging assignments alone. Closer examination of the assignments suggests a possible "floor

effect" for the recent assignments that could explain the consistency of the ratings that is seen in the higher dependability coefficient for the recent rather than for the challenging assignments.

Summary and Conclusions

In summary, while these results are quite limited by the small sample size, the reliability of the dimensions measuring the academic rigor of an assignment (as assessed by the level of interrater agreement) appeared to be moderate overall, and showed a good level of internal consistency. The dimension measuring the clarity of a teacher's expectations (clear expectations), in contrast was poor. Additional development work will need to be undertaken to revise this rubric in order to improve the interrater agreement rubric. It is possible that including more benchmark samples of clear expectations in the rater-training program could help improve the reliability of that dimension as well.

Most of the dimensions measuring academic rigor were significantly and positively associated with one another, specifically, the potential and implementation of the assignment tasks, and the potential of the task and the rigor of the expectations. The exception to this pattern was the dimension measuring the academic rigor of the texts read by students for the assignment. This dimension was not significantly associated with any of the other assignment quality rubrics—within academic rigor or those rubrics measuring the clarity of the expectation—suggesting that this rubric provides unique information regarding instructional quality.

We took a closer look at the assignments we received from teachers to better understand why the rigor of the text was not associated with the other academic rigor dimensions. It appears that some teachers who received low scores for the potential and implementation of their assignment tasks had assigned high quality texts for their students. It did not appear that any of the teachers who received high scores for these dimensions had assigned low quality texts. In other words, it is unlikely that a low quality text would provide the material necessary for a high quality response to literature. At the same time, teachers did not always exploit the potential of the texts they assigned to students by supporting them to analyze and interpret what they read at a deep level. This was illustrated in the assignment described earlier in this paper where students read a text about gorillas, but only were asked to generate simple questions about the text. It is possible then, that a

high quality text could be considered a necessary, but on its own insufficient factor, for a high-quality response to literature assignment task. This issue would need to be explored in future research, however, to draw more definitive conclusions.

The two dimensions measuring clear expectations (the clarity of the expectations and the communication of the expectations to students) also were significantly associated. The dimensions measuring the clarity and rigor of the expectations for an assignment task and the rigor and communication of the expectations to students were significantly associated as well. For the most part, however, the two constructs (clear expectations and academic rigor) did not show a high level of association. This suggests that they are measuring two independent facets of instructional quality. These results also raise questions as to the grouping of the different dimensions, notably, if the rubric measuring the rigor of a teacher's expectations is best situated with the academic rigor dimensions, or if it should be grouped with the clear expectations rubrics.

We also compared our ratings of assignment quality to the quality of observed instruction in order to assess the degree to which the classroom assignment ratings yield information about students' learning environments that were commensurate with other measures of quality practice. Results indicated that the degree to which students were asked to analyze and interpret text (potential of the task) and the rigor of a teacher's expectations for student work were associated with the rigor of the observed lesson. Contrary to expectations, however, the implementation of the classroom task was not associated with the level of observed rigor in the observation. It is not clear to us why this was the case, and raises questions about how we defined potential and implementation in the reading comprehension assignments. These constructs appear to be more difficult to disaggregate in this content area (as opposed to mathematics, see for example, Boston and Wolf, 2004). Again, future research is necessary with larger samples of classrooms to draw more definitive conclusions.

Generalizability studies were conducted to determine how many raters and assignments would be necessary to obtain a stable estimate of the quality of classroom practice. Results indicated that our design based on two raters and four teacher assignments did not yield a stable estimate of quality (G-coefficient = 0.48). This means that individual teachers provided assignments that varied in quality—some poor, some fair, and some good. Interestingly, we found a large difference in the stability of assignment quality when we looked separately at the challenging and

recent assignments. Results for both of the separate analyses are better than those from the combined assignments analysis even though there were half as many assignments included in each of the separate analyses ($\hat{\phi}_{challenging}$ = 0.57, $\hat{\phi}_{recent}$ = 0.90, $\hat{\phi}_{combined}$ = 0.48). (A different, more expected, pattern was seen for mathematics assignments, see Slater, Matsumura, & Junker, 2005.)

We returned to the original portfolios submitted by teachers to gain a better understanding of what this variation meant. In fact, while a few teachers did submit assignments that were consistent in quality (e.g., four poor quality assignments), most of the teachers showed quite a bit of variation in their portfolios. For example, the fourth-grade teacher who submitted the assignment about gorillas described earlier (rated a '1' for academic rigor) also submitted a reading response journal assignment where students wrote a letter to the teacher describing what they were reading (scored a '2' as this was mostly a summary of surface-level events), an analysis of a character's traits with some evidence from the text (scored a '3'), and an assignment where students drew their impressions of a text (scored a '1').

It is possible (even probable) that teachers would submit assignments that were more consistent in quality if we asked for all challenging or all typical work. Considering that we asked for both, our results are hardly surprising. Being more specific with regard to the type of reading comprehension assignment we ask teachers to submit (e.g., all responses to literature) likely would increase the stability of our ratings at the teacher level. Our results from this study are interesting, however, for showing the wide degree of variation *within classrooms* in students' opportunities to develop their comprehension skills, and more importantly, suggests that teachers may have a broad (perhaps fuzzy) idea of how to support student development in this area. This appears to be especially true in terms of creating assignments that teachers believe are challenging in quality.

Finally, while additional refinement of the rating scales is needed, it appears that our ratings of individual assignments captured meaningful differences in students' opportunities to develop higher-level thinking and writing skills. Furthermore, it appears from the quality of the assignments we collected that there is room for improvement in many of the assignments given to students within classrooms. Specifically, as found in other research on reading comprehension instruction, it appears that the tasks assigned to students do not frequently provide students with an opportunity to develop more complex thinking skills, or apply

strategies in a way that supports students to look beyond the surface-level features of a text. Future research is being planned that will focus on the relation of these ratings to student achievement. Looking at the relation of specific dimensions to student learning will help us further refine our scales, and will also provide additional information that could be useful in terms of developing the IQA as a tool for instructional leadership, professional development, and teacher self-assessment.

References

Alvarez, J. (2002). *How Tia Lola Came to Stay*. Dell-Yearling.

Beck, I., McKeown, M., Hamilton, R., & Kucan, L. (1997). *Questioning the author.* Newark, DE: International Reading Association.

Bloom, B.S. (1956). *Taxonomy of educational objectives. Handbook I: Cognitive domain.* New York: McKay.

Boston, M., & Wolf, M.K. (2004, April). *Using the Instructional Quality Assessment (IQA) toolkit to assess academic rigor in mathematics lessons and assignments.* Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.

Clare, L. (2000). *Using teachers' assignments as an indicator of classroom practice*. (CSE Technical Report #532). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

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

Institute for Learning (IFL, 2002). *Principles of Learning*. Overview available at http://www.instituteforlearning.org/pol3.html. University of Pittsburgh, Pittsburgh PA: Author.

Matsumura, L.C., Garnier, H., & Pascal, J. (2002). *Measuring instructional quality in accountability systems: Classroom assignments and student achievement*. (CSE Technical Report #582). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Matsumura, L.C., Garnier, H., Pascal, J., & Valdés, R. (2002). Measuring instructional quality in accountability systems: Classroom assignments and student achievement. *Educational Assessment, 8*(3), 207-229.

National Reading Panel (2000). *Teaching children to read: An evidence based assessment of the scientific research literature on reading and implications for reading instruction.* Washington, DC: National Institute of Child Health and Human Development.

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

Newmann, F.M., Bryk, A.S., & Nagaoka, J.K. (2001). *Authentic intellectual work and standardized tests: Conflict or coexistence?* Chicago: Consortium on Chicago School Research.

Rogoff, B. (1990). *Apprenticeship in thinking*. New York: Oxford University Press.

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

Simon, S. (2003). *Gorillas*. Harper Trophy Publishers.

Slater, S.C., Matsumura, L.C., & Junker, B.W. (2005, April). *Generalizability of a performance measure of instructional quality.* Paper presented at the annual meeting of the National Council on Measurement in Education: Montreal, Canada.

Snow, C. (2002). *Reading for understanding: Toward and R&D program in reading comprehension.* RAND Reading Study Group. Santa Monica, CA: RAND.

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, NY: Teachers College Press.

Storms, B.A., Riazantseva, A, & Gentile, C. (2000). Focusing in on content and communication (writing assignments that work). *California English, 5*(4), 26-27.

Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

2003 draft, please contact the project directors for the revised 2005 version.

Appendix A: 2003 Draft Observation and Assignment Rubrics for Reading Comprehension

For revised 2005 version of the rubrics, please contact:

Dr. Lindsay Clare Matsumura, lclare@pitt.edu
Dr. Brian Junker, brian@stat.cmu.edu

Accountable Talk Observation Rubrics, 2003

Consider talk from the whole-group discussion only.

1. How effectively did the lesson-talk build <u>Accountability to the Learning Community</u>?

Low-inference dimensions, to be rated after observing all teacher-facilitated discussions of the lesson:

A. Participation

Was there widespread participation in teacher-facilitated discussion?

4	Over 50% of the students participated consistently throughout the discussion.
3	25-50% of the students participated consistently in the discussion OR over 50% of the students participated minimally.
2	25-50% of the students participated minimally in the discussion (i.e, they contributed only once).
1	Less than 25% of the students participated in the discussion.
N/A	Reason:

2003 draft, please contact the project directors for the revised 2005 version.

B. Linking contributions

Did speakers' contributions link to and build on each other? (i.e., Was there "local coherence" during the discussion?)

4	At at least 3 points during the discussion, the teacher/student explicitly connects speakers' contributions and shows how ideas/positions shared during the discussion relate to each other.
3	At 1-2 points during the discussion, the teacher / student links speakers' contributions to each other and shows how ideas/positions relate to each other.
2	At one or more points during the discussion, the teacher / student links speakers' contributions to each other, but does not show how ideas/positions relate to each other.
1	Teacher / student does not make any effort to link speakers' contributions.
N/A	Reason:

1. Teacher contributions 4 ___ 3 ___ 2___ 1___

2. Student contributions 4 ___ 3 ___ 2___ 1___

2. How effectively did the lesson-talk build <u>Accountability to Knowledge</u>?

Asking: Were contributors asked to support their contributions with evidence?

4	There are 3 or more efforts to ask students to provide evidence for their contributions, including questions that seemed academically relevant.
3	There are 1-2 efforts to ask students to provide evidence for their contributions that seemed academically relevant.
2	There are one or more superficial, trivial efforts, or formulaic efforts to ask students to provide evidence for their contributions.
1	There are no efforts to ask students to provide evidence for their contributions.
N/A	Reason:

Providing: Did contributors support their contributions with evidence? (This evidence must be appropriate to the content area—i.e., evidence from the text; citing an example, referring to prior classroom experience.)

4	At at least 3 points, speakers provide accurate and appropriate evidence for their claims, including frequent references to the text or prior classroom experience.
3	At 1-2 points, speakers provide accurate and appropriate evidence for their claims, including references to the text or prior classroom experience.
2	In general, what little evidence is offered to back up claims is inaccurate, incomplete, or vague.
1	Speakers do not back up their claims.
N/A	Reason:

2003 draft, please contact the project directors for the revised 2005 version.

3. How effectively did the lesson-talk build <u>Accountability to Rigorous Thinking</u>?

Asking: Were speakers asked to explain their thinking during the lesson?

4	There are 3 or more efforts to ask students to explain their reasoning, including questions that seemed academically relevant.
3	There are 1-2 efforts to ask students to explain their reasoning that seemed academically relevant.
2	There is at least one superficial, trivial, or formulaic efforts to ask students to explain their reasoning.
1	There were no efforts to ask students to explain their thinking.
N/A	Reason:

Providing: Did contributors explain their thinking during the lesson?

4	There are 3 or more examples of speakers explaining their thinking, using reasoning in ways appropriate to the discipline.
3	There are 1-2 examples of speakers explaining their thinking, using reasoning in ways appropriate to the discipline.
2	In general, what little attempt to explain reasoning is vague or inappropriate.
1	Speakers do not explain the reasoning behind their claims.
N/A	Reason:

Academic Rigor: Reading Comprehension Rubrics, 2003

I. Discussion

A. Active Use of Knowledge: Analyzing and Interpreting the Text	
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 specific examples from the text and/or cite examples 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.
2	The teacher guides students to construct a surface-level summary of the text based on straightforward information. 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.
N/A	Reason:

II. Lesson Activities

B. Active Use of Knowledge: Analyzing and Interpreting the Text (Grades 3-5)	
4	During the lesson activity, students 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.
3	During the lesson activity, students engage with some underlying meanings or nuances of a text. Students may interpret or analyze a text, BUT use limited evidence from the text to support their ideas or opinions.
2	During the lesson activity, students construct a literal summary of the text based on straightforward (surface-level) information OR students engage with surface-level information about the text only. Students use little or no evidence from the text to support their ideas or opinions.
1	During the lesson activity, students recall isolated, straightforward (surface-level) facts about a text OR write on a topic that does not directly reference information from the text.
N/A	Reason:

2003 draft, please contact the project directors for the revised 2005 version.

III. Expectations-(Only consider expectations for the task as they were explained to students during initial set-up of lesson activities.)

C. Rigor of Expectations (Grades 3-5)	
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).
N/A	Reason:

II. Lesson Activities

B. Active Use of Knowledge: Analyzing and Interpreting the Text (Grades 1-2)	
4	During the lesson activity, students interpret or evaluate a text AND make explicit references to the text.
3	During the lesson activity, students interpret or evaluate a text. Students to make general references to the text. OR During the lesson activity, students demonstrate a comprehensive understanding of the text through a detailed summary.
2	During the lesson activity, students demonstrate a superficial understanding of the text. Students summarize basic information about a text.
1	During the lesson activity, students do not engage with a text. Students write on a topic (or draw a picture) that does not directly reference information from the text (in other words, the assignment could have been completed without ever having heard or read a specific text).
N/A	Reason:

III. Expectations (Only consider expectations for the task as they were explained to students during initial set-up of lesson activities.)

C. Rigor of Expectations (Grades 1-2)	
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 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).
N/A	Reason:

2003 draft, please contact the project directors for the revised 2005 version.

Clear Expectations/Self- Management of Learning Observation Rubrics, 2003

Rate these dimensions holistically (not by individual student response)

I. Discussion (Lesson Task)

A. 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 un elaborated.
1	The teacher's expectations for the quality of student's work are unclear and/or unelaborated. OR the expectations for quality work are not shared with students.
N/A	Reason:

B. Access to expectations	
4	Criteria for the quality of work expected and how work will 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	The expectations for quality work are not shared with students.
N/A	Reason:

Rate these dimensions for each student interview

C. Understanding of expectations (Student Interview: Grade 1-2 only)	
4	Student clearly explains directions and expectations of quality for the task with details or examples. • Student explains what high, middle, and low-level performance looks like.
3	Student explains directions and expectations of quality for the task without much detail. • Student names a list of expectations.
2	Student vaguely explains directions and quality of expectations for the task. • Student just explains directions.
1	Student knows neither directions nor quality of expectations for the task
N/A	Reason:

Student A ____ Student B ____ Student C ____ Student D ____

II. Past Tasks (Student interview: all grades)

Rate these dimensions for each student interview

D. Judging work based on expectations	
4	Student clearly judges his/her own work based on the specific examples in the work. • Student demonstrates application of expectations to his/her own work (compares expectations to his/ her work) in detail. • Student translates general expectations to the task specifically.
3	Student judges his/her own work based on criteria in general terms. • Student attempts to apply expectations to his/her own work but general comparisons. • Students says, "I included this expectation."
2	Student vaguely judges his/her own work based on general terms. • Student points to expectations (e.g. scoring guide) but is unable to compare expectations to his/her work.
1	Student does not use the criteria to judge his own work
N/A	Reason:

Student A ____ Student B ____ Student C ____ Student D ____

2003 draft, please contact the project directors for the revised 2005 version.

E. Revising work based on expectations	
4	Student clearly explains his/her revision based on expectations with specific examples. • Student explains why s/he revised the work based on expectations and shows previous drafts and points to specific examples of revisions.
3	Student explains his/her revision based on expectations in general terms. • Student shows revisions and explains the reason in general terms based on expectations.
2	Student vaguely explains his/her revision without expectations. • Student shows revisions but doesn't explain the reasons based on expectations (e.g., "I did it to get a better grade or because the teacher told me to do so.")
1	Student is unable to explain his/her revisions or did not have the opportunity to revise his/her work.
N/A	Reason:

Student A ____ Student B ____ Student C ____ Student D ____

Academic Rigor – Reading Comprehension

F. Rigor of Expectations (Grades 3-5)	
4	At one of the expectations described by the student 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 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 expectations focus on building a basic understanding of the text (e.g., summarizing).**
1	The 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).
N/A	Reason:

Past Task: Student A ____ Student B ____ Student C ____ Student D ____

2003 draft, please contact the project directors for the revised 2005 version.

F. Rigor of Expectations (Grades 1-2)	
4	At least one of the expectations described by the student 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 least one of the 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 expectations focus on building a basic understanding of the text (e.g. summarizing).**
1	The 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).
N/A	Reason:

Current Lesson Task: Student A ____ Student B ____ Student C ____ Student D ____

Past Task: Student A ____ Student B ____ Student C ____ Student D ____

Observation Checklists, 2003

<u>Accountable Talk Function Checklist, 2003</u>: Check all that apply and script relevant contributions.

Most of these moves will be made by the teacher, but in some cases, students might make them. In recording the actual moves, note T for Teacher move, S for Student move.

(script here)

<u>1. Linking contributions</u>

☐ Getting students to relate to one another's ideas
"Jay just said…and Susan, you're saying that…"
"Who wants to add on to what Ana just said?"
"Who agrees and who disagrees with what Ana just said?"
"How does what you're saying relate to what Juan just said?"
"I agree with Sue, but I disagree with you, because…"
S- "I agree with Fulano because…"

<u>2. Accountability to knowledge</u>

☐ Pressing for accuracy
"Where could we find more information about that?"
"Are we sure about that? How can we know for sure?"
"Where do you see that in the text?"
"What evidence is there?"
T revoices S contribution and checks for accuracy

☐ Building on prior knowledge / recalling prior knowledge
T or S links present work to past work
"How does this connect with what we did last week?"
"Do you remember when we read another book by this author?""

<u>3. Accountability to rigorous thinking</u>

☐ Pressing for reasoning
"What made you say that?"
"Why do you think that?"
"Can you explain that?"
"Why do you disagree?
"Say more about that."
"Let's let Fulano think."

2003 draft, please contact the project directors for the revised 2005 version.

Clear Expectations/ Self- Management of Learning Checklist, 2003

Clear Expectations / Self-Management of Learning (CE/SML)

Means of communicating expectations <u>during the lesson</u>

Check all below that were used to communicate expectations during the lesson.

☐ Criteria chart
☐ Process chart
☐ Rubric
☐ Model of student performance that meets standard
☐ Model of intermediate expectation
☐ Counter-model of unacceptable performance
☐ Template- outlines all the steps and information necessary to complete the task
☐ Oral explanation of expectations
☐ Other: _____

Means of communicating expectations <u>during the student interviews</u>

Check all below that were used to communicate expectations during student interviews. Ask students about these means of communicating expectation with students during interviews.

Photograph relevant charts, handouts, etc.

☐ Criteria chart
☐ Process chart
☐ Rubric
☐ Model of student performance that meets standard
☐ Model of intermediate expectation
☐ Counter-model of unacceptable performance
☐ Template- outlines all the steps and information necessary to complete the task
☐ Oral explanation of expectations
☐ Other: _____

Academic Rigor: Reading Comprehension Checklist, 2003

Academic Rigor – Reading Comprehension

Text title: _____ *Author:* _____

Engagement with text:	
__ Teacher reads aloud to class __ Teacher reads from text as student read along __ Student(s) read aloud to class __ Student(s) read from text as peers read along __ Students read with peer __ Students read silently __ Other:	□ **1st reading** □ **Subsequent reading**

Check each box that applies.

Recall Fragmented, Isolated Facts	Construct a Surface-level Summary of the Text	Construct an Enriched & Elaborated Understanding of the Text	Engage with the Underlying Meanings or Literary Characteristics of a Text
Fiction & Nonfiction: _ Answer questions that have a single correct answer (questions are not open-ended) _ Provide "bits" of information _ Describe life experiences without explaining how these help them understand the text _ Describe other books read without explaining how these help them understand the text	Fiction: _ Retell events in sequence _ Identify the characters and/or setting of a text	Fiction: _ Discuss character motives _ Describe the causes and effects of specific events	Fiction: _ Analyze symbols _ Discuss themes _ Compare and contrast texts _ Evaluate a text _ Adopt the perspective of a character _ Discuss the author's craft techniques _ Extend the story (consider alternative outcomes to the ending)
	Nonfiction: _ Describe information learned organized by topic (facts are "chunked" not fragmented)	Nonfiction: _ Explain how information learned from text is interrelated (causes and effects) _ Draw generalizations from or about content not explicit in the text	Nonfiction: _ Support an idea or conclusion from the information learned in the text _ Connect content learned from text to information already known

2003 draft, please contact the project directors for the revised 2005 version.

Reading Comprehension Assignment Rubrics, 2003

Dimension 1
Academic Rigor: Rigor of the Text

Rubric 1a: Rigor of the Text (Grades 3-5)	
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 1b: Rigor of the Text (Grades 1-2)	
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	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. The themes are conventional that there is little about the text that requires extended discussion.
1	There is no "grist" for students to discuss to make meaning of the story. The text does not contain a narrative, information, or interesting language. It may, for example, be a decodable text or a highly patterned book that was designed for teaching print-sound code or fluency.
N/A	Reason:

2003 draft, please contact the project directors for the revised 2005 version.

Dimension 2
Academic Rigor: Potential

Rubric 2b: Analyzing and Interpreting the Text: (Grades 3-5)	
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.
N/A	Reason:

Rubric 2a: Analyzing and Interpreting the Text: (Grades 1-2)	
4	The task guides students to interpret or evaluate a text AND make explicit references to the text. AND students have ample opportunity to develop their thinking (e.g. challenging questions, extended responses, and analytical and interpretive responses).
3	The task guides students to interpret or evaluate a text. The lesson task requires students to make general references to the text. OR The task requires students to demonstrate a comprehensive understanding of the text through a detailed summary. There is some opportunity for students to develop their thinking (e.g. challenging questions but structured responses).
2	The task guides students to demonstrate a superficial understanding of the text. Students summarize basic information about a text. OR students engage in perfunctory responses and have no opportunity to develop higher level thinking skills.
1	The task does not require students to engage with a text. Students write on a topic (or draw a picture) that does not directly reference information from the text (in other words, the assignment could have been completed without ever having heard or read a specific text). OR The task guides students in recalling fragmented information about the text.
N/A	Reason:

2003 draft, please contact the project directors for the revised 2005 version.

Dimension 3

Academic Rigor: Implementation

Rubric 3a: Implementation of the Task: (Grades 3-5)	
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.
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 3b: Implementation of the Task: (Grades 1-2)	
4	Students interpreted or evaluated a text AND made *explicit* references to the text.
3	Students interpreted or evaluated a text AND to made *general* references to the text. OR Students demonstrated a comprehensive understanding of the text through a detailed summary.
2	Students demonstrated a superficial understanding of the text. Students summarized basic information about a text.
1	Students did not engage with the text. Students wrote on a topic (or drew a picture) that does not directly reference information from the text (in other words, the assignment could have been completed without ever having heard or read a specific text).

Dimension 4
Academic Rigor: Expectations

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).

Dimension 1
Clear Expectations: Clarity and Detail of Expectations

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.

2003 draft, please contact the project directors for the revised 2005 version.

Dimension 2
Clear Expectations: Communication of Expectations

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: