

**Beyond Summative Evaluation:  
The Instructional Quality Assessment  
as a Professional Development Tool**

CSE Technical Report 691

Amy C. Crosson  
Harvard Graduate School of Education

Melissa Boston, Allison Levison, Lindsay Clare Matsumura,  
Lauren B. Resnick, and Mikyung Kim Wolf  
University of Pittsburgh, Learning Research and Development Center

Brian W. Junker  
Carnegie Mellon University

July 2006

Correction in Authorship: Versions of this report prior to September 30, 2006 contained a cover page listing Brian Junker and Lindsay Clare Matsumura 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: Amy Crosson, Melissa Boston, Allison Levison, Lindsay Clare Matsumura, Lauren Resnick, Mikyung Kim Wolf, 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 Building, 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

Copyright © 2006 The Regents of the University of California

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), and the U.S. Department of Education.

The findings and opinions expressed in this report are those of the authors 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.

**BEYOND SUMMATIVE EVALUATION:  
THE INSTRUCTIONAL QUALITY ASSESSMENT AS A PROFESSIONAL  
DEVELOPMENT TOOL<sup>1</sup>**

Amy C. Crosson  
Harvard Graduate School of Education

Melissa Boston, Allison Levison, Lindsay Clare Matsumura,  
Lauren B. Resnick, and Mikyung Kim Wolf  
University of Pittsburgh, Learning Research and Development Center

Brian W. Junker  
Carnegie Mellon University

**Abstract**

In order to improve students' opportunities to learn, educators need tools that can assist them to reflect on and analyze their own and others' teaching practice. Many available observation tools and protocols for studying student work are inadequate because they do not directly engage educators in core issues about rigorous content and pedagogy. In this conceptual paper, we argue that the Instructional Quality Assessment (IQA)—a formal toolkit for rating instructional quality that is based primarily on classroom observations and student assignments—has strong potential to support professional development within schools at multiple levels. We argue that the IQA could be useful to *teachers* for analyzing their own and their colleagues' practice; additionally, the IQA could aid the efforts of *principals* in their work as instructional leaders, identifying effective practitioners to help lead professional development within a school and targeting professional development needs that would require external support. Although the IQA was designed for summative, external evaluation, we argue that the steps taken to improve the reliability of the instrument— particularly the efforts to make the rubric descriptors for gradations of instructional quality as transparent as possible—also serve to make the tool a resource for professional growth among educators.

---

<sup>1</sup> PLEASE DO NOT CITE OR QUOTE WITHOUT PERMISSION

Paper presented at the Annual Meeting of the American Educational Research Association Meeting, San Diego, CA, April 2004

The authors are grateful to Nonie K. Lesaux and John Soluri for their thoughtful comments.

We can agree, as a matter of process, to treat all issues of pedagogy as matters of personal taste. But doing so would mean that decisions about professional development would be largely personal also, disconnected from collective knowledge, about best practice in the improvement of students' learning. Thus, the prospects for large-scale improvement would remain dim.

-Elmore, 2002

## **Introduction**

Evaluation tools that are designed for large-scale summative assessment purposes rarely support the kind of detailed feedback and on-going growth that is beneficial to teachers. In most cases, because external assessments must meet stringent requirements of reliability and validity, and because they must be broad enough in scope to align with a variety of curricula, they cannot, by design, simultaneously provide the context-embedded and highly detailed feedback that characterizes useful formative assessment tools (Shepard, 2003). While we recognize that summative evaluation tools typically are not appropriate for formative assessment purposes, we argue that the Instructional Quality Assessment (IQA)—because of its theoretical and research bases, in addition to some of its features designed to support reliability—is uniquely positioned to provide useful formative feedback to instructional leaders and teachers to support school improvement through professional growth.

This conceptual paper lays out a vision for how we believe the IQA could be used by classroom teachers and principals to support instructional improvement specifically in the areas of reading comprehension and mathematics. We begin by explaining our rationale for the need for a tool like the IQA to bolster professional learning for both teachers and administrators. We then ask how the IQA could be used collaboratively by teachers to support their professional growth in the two content areas for which the IQA has been developed to date. Specifically, we consider how the IQA could support teachers in assessing and improving their own practice, highlighting features that are directly relevant to the use of the IQA for professional development purposes by drawing on what we learned from training educators unfamiliar with the IQA in an intensive Instructional Quality Assessment Rater Training Program during a pilot study in the spring of 2003. We then ask how the IQA can be used to support instructional improvement by educators in school leadership positions. Finally, we conclude by considering some of the limitations of the IQA as a formative assessment.

## **Rationale**

The development of instructional leadership and creation of on-going, school-based opportunities to develop professional teaching capacity have been identified as key elements for scaling up educational improvement across school districts and instructional programs (Elmore, 1996; Fink & Resnick, 1999; Resnick & Hall, 1998). Because students across a broad spectrum of background knowledge and interests are being asked to learn more challenging content at deeper levels than ever before, it follows that the demands on educators to learn how to teach students to high levels are more challenging than ever before. Thus, schools must transform themselves into learning organizations not only for the students they serve, but also for the professionals who run them.

Principals play a critical role in leading their schools' continuous professional learning (Fink & Resnick, 1999). While it would not be reasonable to expect principals to become as deeply knowledgeable about effective instruction as the level of expert teachers, they do need to be knowledgeable enough about good curriculum and instructional practice to make sound decisions about professional development to lead the learning of school faculty (Spillane, 2001, Stein & D'Amico, 2002). There are two central reasons why principals need to be knowledgeable about effective curriculum and instruction. First, principals need to be able to identify the most effective teachers in a given domain so that they can capitalize on those teachers' expertise to "build capacity" within the school. Indeed, Richard Elmore (2002) claims that "systems need to be able to identify people who know what to do, and to create settings in which people who know what to do can teach those who do not" (p. 26). Second, when the capacity is not already present within the school, principals need to set specific goals to improve curriculum and instruction. This entails recognizing when the current curriculum does not provide opportunities for rigorous learning and therefore must be enriched or replaced. Additionally, this involves identifying specific areas of instructional practice that could be improved upon to better support student learning.

For teachers, on-going learning opportunities are essential to develop new ways of teaching that support students in meeting the new demands of rigorous curricula. Yet while teachers and schools have been held accountable for student success, it is widely acknowledged that most professional development opportunities available to teachers are inadequate for supporting the improvement

of classroom practice (e.g., Cohen, McLaughlin, & Talbert, 1993; Lieberman, 1994; Saunders, Goldenberg, & Hamann, 1992). For example, professional development characterized by fragmented and generic efforts, typical of the workshops and conferences experienced by most teachers during district in-service days, do not improve day-to-day classroom practice (Little, reviewed in Dole 2003). Instead, effective professional development is based in schools and focuses on day-to-day practice around a specific curriculum (e.g., Ball and Cohen, 1999; Elmore, 2002). Rather than “tacking on” new practices, professional development should remain “proximal to practice” (Elmore, 2002) and indeed should assist teachers to fit new ideas to improve existing practice (Dole, 2003).

Effective professional development in both mathematics and reading comprehension enables teachers to generalize and apply ideas about good practice and appears to depend on assisting educators in understanding the thinking behind why specific goals in mathematics or reading comprehension are considered good content or pedagogy. Boston and Wolf (2004) point out that the most important first step in supporting reform-oriented practice in mathematics instruction is identifying the *tasks* that provide opportunities to learn mathematical reasoning and problem-solving. They argue that this is likely to affect how teachers choose tasks, the kinds of opportunities they give students when implementing those tasks, and their lens on students’ work and how they assess this work to think about the effectiveness of their own instruction (Boston & Wolf, 2004).

The knowledge base about professional development in reading comprehension lags behind what is known about how to support teachers’ pedagogical content knowledge in math. For example, in a review of the effects of professional development in reading comprehension instruction, Dole (2003) found that most studies examined structures (e.g., cooperative learning groups) rather than looking at the quality of instruction through interactions between teachers and students, and they failed to look at outcomes in terms of student learning. Despite a dearth of research on professional development in reading comprehension, she argues that teachers need an understanding of the theory of the active comprehension processes behind what they are asked to do. Otherwise, “if teachers learn strategies apart from theoretical underpinnings, they are unlikely to use them effectively or reliably” (p. 182). Dole argues that effective professional development in reading comprehension therefore actively engages teachers in study groups and provides opportunities to observe other teachers using instructional strategies.

Indeed, she argues that a critical aspect of professional development in reading comprehension is to ensure that all teachers have the opportunity to “see” instructional strategies in practice and to receive feedback.

Given this tremendous need to support principals and teachers engaged in on-going learning about effective instruction, we believe that a crucial part of creating such continuous learning opportunities for administrators and teachers is providing tools to structure and support improvement efforts. We consider certain aspects of the IQA to be especially promising for promoting professional growth. Firstly, the IQA is built on an underlying assumption that there is a tremendous body of accumulated knowledge about what expert instruction looks like (both in terms of general pedagogical principals and within the disciplines) and about how to teach content-specific information and ways of thinking.<sup>2</sup> Secondly, the definition of “instructional quality,” for texts, tasks, teacher-student interactions, in addition to the support structures provided to students, enables the instrument to define far more precisely than other tools what is meant by “quality instruction.” Other tools, in contrast, require extensive rater training in order to understand at a reliable level what the different scale points mean (as is the case with Likert scales) or to explain what is meant by rigorous content (as is the case in the example of the Queensland School Reform project which includes high-inference language in the rubrics such as “worthwhile content”). Thirdly, the tool’s protocols assist observers in citing key evidence for rubric ratings and are features that have been used primarily to support inter-rater reliability, but we believe are equally valuable to support learning and instructional improvement.

Although the IQA has only been tested as an external, summative evaluation tool to date, we do have some indication—especially from our rater training program—that the tool also has strong potential to support practitioners in learning about best instructional practices. Technical qualities of the IQA concerning inter-rater reliability have been discussed in previous papers in this symposium (Boston & Wolf, 2004; Junker et al., 2004, Clare Matsumura, Wolf, Crosson, Levison,

---

<sup>2</sup> The general pedagogical principals that underlie good instruction, that, as was explained in the introductory paper to this symposium (Junker et al., 2004), we refer to as the Principles of Learning (Appendix A). This set of principles, distilled from several decades of research in cognitive, social, and developmental psychology, “rings true” for educators because they describe a widely established vision of good instruction (e.g., Cohen, McLaughlin, & Talbert, 1993). In addition to general pedagogical principles, more and more is known within the disciplines about how to teach content-specific information and ways of thinking, or “pedagogical content knowledge” (Shulman, 1987).

Peterson, Resnick, and Junker, 2004; Wolf et al., 2004). In general, the rubrics make important distinctions about gradations in quality of tasks, texts, lessons, and assignments (Boston & Wolf, 2004; Junker et al., 2004, Clare Matsumura et al., 2004; Wolf et al., 2004). Exact agreement for each four point rubric, while variable, is for the most part moderate. We are not only interested in strengthening reliability to make the instrument more technically robust, but also to help transform the nature of professional development.

### **Using the IQA to Guide and Enrich Teacher Self-Assessment**

For teachers, the IQA could guide the selection or modification of curricular tasks, decisions about instructional moves, and support for students to meet the demands of challenging tasks. There are two primary ways we envision teachers might use the IQA rubrics and protocols. First, the IQA could be utilized to analyze artifacts of teaching in ways that link the quality of student work with the opportunities provided via assignments and instruction. Second, teachers could use the IQA to observe each others' instruction and provide specific feedback about interactions between teachers and students focused on the effectiveness of the instruction.

### **The IQA as a Lens for Analyzing Artifacts of Practice**

The following example, in which students were asked to write a summary of the story, *Jumanji*, by Chris Van Allsburg, illustrates how the IQA would provide added value to teachers' analysis of classroom artifacts. For this assignment, students were asked to "write a summary including all important events in the story (without writing every little detail)." They had to "include names of characters and write events that took place, therefore showing comprehension of the text." Clare Matsumura and her colleagues (2004) point out that although an assignment asking students to summarize the text is no easy task given the complexity of this particular story, the assignment only asked students to convey surface-level understanding of the text. Clare Matsumura et al. presented this example as a benchmark "2" on the IQA four point rubric (with 1=poor and 4=excellent).

The level of precision and detail embedded within the IQA rubrics—a feature that intended to increase rater reliability—may make the tool especially useful for teachers' analysis of classroom artifacts, as well as instruction. The content of the rubrics, themselves, by specifying what is meant for gradations of "instructional



quality” for this content area at this grade band, would guide teachers engaged in analyzing their reading comprehension assignments to determine that the *Jumanji* task characterized a score of “2” (See Figure 1). Further, the rubric descriptors, by specifying the qualitative differences between the score levels, could be used by teachers to consider many of the essential elements of more rigorous assignments. These elements could be incorporated into a subsequent version of the task, or into other assignments to ensure that students have opportunities to engage in highly challenging reading comprehension tasks.

1	The assignment task guides students to recall isolated, straightforward facts about a text OR write on a topic that does not directly reference information from the text.
2	The assignment task guides students to construct a literal summary of the text based on straightforward (surface-level) information OR the assignment task guides students to engage with surface-level information about the text only. The assignment task guides students to use little or no evidence from the text to support their ideas or opinions.
3	The assignment task guides student to engage with some underlying meanings or nuances of a text. The assignment task guides students to interpret or analyze a text, BUT 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).
4	The assignment task guides student to engage with the underlying meanings or nuances of a text. The assignment task guides students to interpret or analyze a text AND use extensive and detailed evidence from the text to support their ideas or opinions.

Figure 1: Rubric descriptors for scoring reading comprehension assignments in the upper elementary grades

By sharing as transparently as possible the qualitative differences between the rubric scale points, the IQA makes explicit the features of rigorous assignments. This feature of the tool would enable teachers not only to assess the rigor of a given task, but more importantly, it would assist teachers to determine how to improve the task. In this example, the teacher might revise the assignment to include a question that advances analysis of the text that emphasizes a judgment about themes and the author’s messages. Thus, the tool has the potential to play an instrumental role in enabling teachers to self-monitor improvement in their own practice.

In addition to precise rubric descriptors, many of the rubrics are further supported through a more detailed checklist to guide understanding of each score level (Table 1). For example, in the case of the *Rigor of the Text* rubric for primary grades, a checklist is used to assist raters in determining a valid score on the

accompanying rubric (1=poor, 3=excellent). The checklists do not directly translate into rubric scores, but they provide additional guidance about what to look for when analyzing texts, tasks, and classroom interactions.

Table 1  
Checklist to Guide Raters in Scoring *Rigor of the Text* Rubric

<b>FICTION</b>	<b>3</b>	<b>2</b>	<b>1</b>
Language of text	<ul style="list-style-type: none"> <li>_ rich vocabulary throughout the story</li> <li>_ literary language</li> <li>_ many complex sentence structures</li> </ul>	<ul style="list-style-type: none"> <li>_ some rich vocabulary appear sporadically throughout the story</li> <li>_ some diverse sentence structures</li> </ul>	<ul style="list-style-type: none"> <li>_ decodable vocabulary</li> <li>_ all simple (oral language) sentence structures</li> <li>_ highly patterned sentences</li> <li>_ predictable plot or no plot</li> <li>_ story is conveyed more through illustrations than through the text</li> </ul>
	<ul style="list-style-type: none"> <li>_ complex theme</li> <li>_ complex relationships among characters</li> </ul>	<ul style="list-style-type: none"> <li>_ straightforward storyline</li> <li>_ conventional, familiar theme</li> <li>_ predictable problems and solutions</li> <li>_ straightforward characters</li> </ul>	
<b>NON-FICTION</b>	<b>3</b>	<b>2</b>	<b>1</b>
Language of text	<ul style="list-style-type: none"> <li>_ specialized/technical vocabulary throughout the text</li> <li>_ many complex sentence structures</li> </ul>	<ul style="list-style-type: none"> <li>_ some specialized/technical vocabulary appear sporadically throughout the text</li> <li>_ some diverse sentence structures</li> </ul>	<ul style="list-style-type: none"> <li>_ decodable vocabulary</li> <li>_ all simple (oral language) sentence structures</li> <li>_ highly patterned sentences</li> </ul>
Complexity of content	<ul style="list-style-type: none"> <li>_ much information for students to understand and synthesize</li> <li>_ major themes are elaborated</li> <li>_ complex concepts are explained</li> </ul>	<ul style="list-style-type: none"> <li>_ major themes are presented without supporting details</li> <li>_ simple explanations, or no interconnection of explanations</li> </ul>	<ul style="list-style-type: none"> <li>_ conveys no or limited information</li> <li>_ labeling book</li> </ul>

These checklists appear to be especially useful because they alert raters to primary sources of evidence and key characteristics of instruction for scoring rubrics in the midst of observing the complexity of real time interactions between teachers

and students.<sup>3</sup> In the case of the *Jumanji* assignment, teachers would be guided by the IQA rubrics and checklist to consider how the various features of text enhance or constrain the degree to which the text is useful for reading comprehension instruction. This is important because the richness of the text can place a ceiling on the potential for rigorous comprehension work if it does not contain enough “grist” for students to work at making sense of the text (Beck & McKeown, 2001).

Perhaps most importantly, making the IQA rubrics and checklists available to teachers to support and self-monitor their instruction may strengthen opportunities for teachers to become involved in professional development structures that engage them more deeply in core issues about rigorous content and pedagogy than most structures currently afford. For example, many educators are already involved in school-based professional development structures, but these tend to focus on helpful—but not sufficient—norms of participation. The IQA, by clearly outlining and focusing on “quality instruction,” might center conversation in study groups on evidence of worthwhile student learning and the academic opportunities afforded in the work itself. The IQA would therefore be a tremendous asset to teacher study groups by framing conversation about student work in the collective knowledge about rigorous content and effective teaching practices.

The IQA also has strong potential to support the professional growth of teachers because components of the tool address the connections between student work and teacher assignments and instruction. As the papers by Clare Matsumura, et al. (2004) as well as Boston and Wolf (2004) pointed out earlier in this symposium, previous research demonstrates a direct relationship between the rigor of assignments and the quality of student work generated by those tasks (e.g., Clare & Aschbacher, 2001; Newmann, Lopez & Bryk, 1998; Stein, Smith, Henningson, & Silver, 2000). The IQA is especially well-suited to support an increased awareness of the connection between what is taught and what students learn because it offers a coherent lens on the kinds of opportunities provided through both task/text and instruction to support student learning.

---

<sup>3</sup> The correlations between six “naïve” raters who were trained to use the checklists during the Spring 2003 pilot demonstrated a strong relationship between their rubric scores and checklist items, strongly supporting our hypothesis that the checklist is helpful in assisting raters in selecting rubric scores (Appendix B).

## The IQA as a Lens for Focusing Feedback on Instruction

In addition to analyzing artifacts of classroom practice, the IQA could also be used to support professional growth by teachers who observe each others' instruction to provide specific feedback about interactions between teachers and students focused on the effectiveness of the instruction. Not only are precise rubric descriptors that concretize what is meant by instructional quality very helpful for this purpose, but also, the IQA's data collection protocols may assist teachers in providing feedback to peers, as well as in assessing their own instruction. The IQA's data collection protocols focus raters' attention on rich evidence for scoring rubrics, and they shape how raters record evidence.

The following example in which students engage in a book discussion about *Home Run* by Robert Montgomery, illustrates how the IQA data collection protocols guide observers to pay attention to pivotal conversational moves during whole group discussions (Appendix C). The classroom observation protocol alerts observers to specific kinds of talk moves that are likely to support students in listening to and building on one another's ideas (*Accountability to the Learning Community*), citing appropriate and accurate evidence from the text (*Accountability to Accurate Knowledge*), and explaining the reasoning behind their claims (*Accountability to Rigorous Thinking*) (AT CD-ROM). Wolf and her colleagues (2004) demonstrated how specific kinds of conversational moves that teachers make while facilitating discussions around texts have a strong relationship with the overall rigor of the lesson. A "weak" *Accountability to Rigorous Thinking* example discussed in Wolf and her colleagues' (2004) paper illustrates how the IQA may assist teachers in not only identifying learning needs, but also in determining specific pedagogical moves to enhance student learning:

T: Let me read this again so you can discuss this with a partner. [re-reads stanza and students discuss ideas with a partner]

T: Okay, let's hear a few ideas.

S1: I was thinking they hit the ball...he is going to get it.

T: I think I heard something different from this group over here. Would you like to share your ideas with us?

S2: I think he is going to hit a homerun and the team starts to yell because they win the game.

S3: They were yelling because the ball was hit far and they were yelling for him to catch it.

T: So, you agree with Michele, that they were yelling to tell him to catch the ball.

S3: Yes, they were yelling at him to pay attention.

T: Well, let's find out...[Teacher continues to read the book]

As a formative assessment, the IQA observation protocols and rubrics would focus discussion on specific evidence of talk moves that are “on the way” toward facilitating productive discussions to support comprehension as well pointing specifically towards next steps. For example, while the teacher clearly made initial efforts to get students to share their thinking about how events are unfolding in the text by asking “would you like to share your thinking with us?” this move could have been further developed. The IQA rubrics, by describing gradations in quality on rubric descriptors, would make apparent tangible “next steps,” including asking follow-up questions that encourage students to explain the reasoning behind their predictions guided by a menu of potential talk moves (e.g., “What makes you say that?” and “Say more about that.”) Dole (2003) observes that “most adult readers do not think about how they strategically process text; they just do it. Therefore, it is difficult to make teachers cognitively aware of what they do so that they can communicate that awareness to their students” (p. 186). Thus, use of the IQA would assist in professional development that supports teachers in making explicit with students the strategies they use when making sense of a text.

In our experience in training raters to use the protocols for scripting discussion, coding exchanges, and scoring rubrics, raters were able to describe evidence of Accountable Talk with increasingly precision over the course of the lesson observations. For example, during the initial days of the pilot, raters' evidence for their Accountable Talk scores tended to be vague references to the lesson discussion (e.g., “um, it was something she said in the beginning of the discussion, just after she stopped reading”), whereas by the end of the pilot, raters provided far more specific evidence for their scores (e.g., “As for teacher's Accountability to Knowledge, she asked the student the meaning of a word. *Quench*. She kept asking, ‘What do you mean? Can you give me an example of *quench*?’”)

As was explained in Wolf and her colleagues' (2004) paper, the critical distinctions in coding Accountable Talk are a) recognizing discussion exchanges that qualify as Accountable Talk; b) coding them into the correct category, and c)

determining whether the exchange is a low or high-quality example. We have found that the latter of these distinctions is the most difficult to make, and plan to develop additional rater training materials to help raters make these distinctions.

The IQA as a formative self-study tool could assist teachers to identify the patterns of conversation in the discussions they facilitate, and may provide guidance about specific talk moves to probe students' ideas and ability to cite evidence during discussion. Thus, the lesson observation and coding protocol coupled with the precise descriptors offered by IQA rubrics may enable practitioners to work in collaboration with colleagues to figure out not only their strengths and weaknesses, but also to identify "next steps" for instructional improvement. Further, the IQA rubrics could be used to track growth over time. Moreover, teachers could use the IQA observation protocols to observe colleagues who are working on improving their ability to facilitate discussions to support reading comprehension (i.e., to develop strategies for deepening comprehension of text) or in mathematics (i.e., to construct understanding of mathematical concepts and procedures). Not only do the observation protocols teach observers how to identify appreciable clusters of verbal exchanges in classroom discourse, but also directs them to determine whether the exchange is "strong" or "weak" in its potential to support deep thinking and engagement about important content.

### **Supporting Instructional Leadership for Principals**

The features of the IQA protocols and rubrics intended to increase rater reliability, in addition to serving as a self-assessment tool for teachers, may also benefit principals as they strive to become instructional leaders in their schools. In fact, the IQA protocols and rubrics could be especially valuable for supporting instructional leadership given that it is not possible to hold deep pedagogical content knowledge in all disciplines across the content areas. Specifically, the concrete and precise definition of instructional quality outlined in IQA rubrics may assist principals in making decisions about how to allocate professional development resources toward needs of specific teachers or subsets of faculty, as well as figuring out where instructional expertise already exists in the school to help build capacity.

IQA items can be used to yield important diagnostic information about the quality of curriculum. The lesson observation protocols and rubrics may help principals sharpen their focus on specific aspects of instructional practice for a more

precise vision of rigorous instruction. Additionally, the classroom observation protocols will help them provide specific evidence for their assessments.

Boston and Wolf's (2004) finding from the IQA Spring 2003 pilot that included mathematics tasks from lessons observed in District D, even with some expected variability in quality, never afforded students the opportunity to engage with tasks that deepen students' mathematical understanding of concepts and procedures at the highest level on our rubrics, and has profound implications for school and district leadership. Such a finding points to the need to examine existing curriculum as well as the need to provide professional development on how to modify tasks so that they provide the potential for students to construct deep mathematical understanding.

The detailed and explicit definition of instructional quality within specific content areas offered by the IQA may be an ideal tool for assisting principals in figuring out which teachers are most effective within different domains of instruction.

However, a caveat is in order here. Since the IQA was not developed for formal evaluation of individual teachers, it should not be used for that purpose. We plan to develop a modified rater training program for use with principals that will include professional ethics training. While this training is important in general, the most important aspect in this case is that it is unethical to use protocols for instructional observation that have only moderate reliability at the classroom level. Evaluative or punitive actions would never be warranted or ethical due to the low classroom-level reliability and consequent risk of false positive or false negative inferences about individual teachers.

### **Conclusion**

In this paper, we have argued that the IQA distinguishes itself from most external evaluation tools in that, while it has been designed primarily as a summative evaluation tool, the very technical qualities that we have emphasized for external assessment have provided an advantage for use as a formative internal tool to support school improvement through professional growth. By sharing as transparently as possible the characteristics of rigorous tasks and rich texts, as well as the features of effective classroom interactions, the IQA would enable teachers themselves to determine the strengths and weakness in curriculum materials and practice, and to figure out next steps for professional growth. Thus, the tool might

have the key supporting role in enabling teachers to self-monitor improvement in their own practice.

Indeed, the IQA, when used for formative purposes, supports “dynamic assessment” (Shepard, 2000), in that it can be used to provide effective feedback that points out differences between learner’s performance and the “ideal” (citing Pressley p. 30). Sharing with teachers the criteria by which their practice will be assessed is not only fair, but necessary. Undoubtedly, efforts that relegate teachers to a passive role, rather than an active role in leading improvement, are likely to generate resistance (Darling-Hammond, 1997; Dole, 2003).

Of course, as a tool, the IQA could be used—or misused—in many ways. Our objective in modifying the IQA for internal use to provide on-going, formative feedback is to support the development of a collaborative learning culture for on-going improvement. Tools alone cannot do this work. This can only be accomplished by communities of educators. Nonetheless, the IQA may strengthen the social networks critical for school improvement by facilitating school-based professional development that it is deeply focused on core issues of content and teaching practices.

While the IQA focuses on rich aspects of curriculum and instruction, there are, of course, other important aspects of teaching that are not the focus of the tool. For example, the IQA is not adequate for examining how instruction is differentiated for heterogeneous groups of students with varying needs. We are currently in the process of modifying the tool for the purpose of defining instruction for English-language learners (ELL) and designing a study to consider how practices captured by the tool differentially support various profiles of ELLs. Additionally, the tool is not currently designed to assess or guide instruction as it builds over time in the course of a unit of study.

On the other hand, there are some aspects of instruction that can be incorporated into a formative version of the IQA with minimal modification. Over the last two years, some IQA rubrics have been pared back in an effort to make them more reliable and affordable for large-scale assessment. As Clare Matsumura and her colleagues (2004) discussed in an earlier paper on reading comprehension assignments, one of the driving goals has been to create a tool that is both a rich measure of instruction and at the same time parsimonious. Therefore, as part of the development of the external tool, we have eliminated some items that would be



important for professional development purposes. For example, in an earlier version of the IQA, the amount and quality of teacher feedback to students was part of the classroom observation measurement. However, we found that raters could not simultaneously attend to the kind of feedback students received from teachers and at the same time interview individual students about their understanding of expectations and how they judged their own work. In the interest of a reliable and parsimonious tool, we eliminated the teacher feedback items. However, there is evidence that the quality of feedback that teachers provide to students has an impact on student learning (e.g., Elawar and Corno as cited in Shepard, 2000). Teacher feedback is critical for communicating Clear Expectations specific to the students' work and also for creating opportunities for students to learn to self-monitor their learning. It would not be necessary to limit use of the internal version of the tool to a single visit, however. In fact, ideally the IQA would be used over the course of time. Thus, it would be possible, with minimal modification, to include more items than are part of the present, "external" version of the IQA.

Perhaps the most daunting potential limitation of the IQA as a formative assessment tool is that the explicit definition of instructional quality that it offers may be perceived as a controlling effort that imposes a unitary, inflexible approach to teaching. As we take part in this effort that requires educators at all levels to learn new pedagogical approaches and to embrace collaborative improvement processes, it is essential that teachers are professionally engaged in the process. If not, such efforts could fail miserably. Indeed, mechanical application of the IQA criteria will probably not, in and of itself, enable professional learning and refinement of excellent teaching practices. Instead, educators should have the opportunity to engage with the criteria—to apply them to their own work to challenge and to improve them. Therefore, it is critical that the IQA as a formative assessment build-in flexibility—including ways to enter into dialogue with teachers.

Even given this risk and the tool's limitations, the framework it offers for formative assessment is a rich starting point. To date, the IQA has only been tested for its external, summative function. Thus applications for self-assessment by teachers to support professional growth and the extent to which the IQA can scaffold principals' efforts at determining the strengths and needs of their faculties and making related decisions about professional development resources remain to be empirically tested. Combined with summative use, the IQA offers rich supporting materials for professional development and self-study makes it possible

for districts, schools and teachers not only to be the objects of study (or evaluation) of instructional practice, but also to be active participants in improving daily instructional quality in ways that focus on high-level student achievement.

## References

- Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as a learning profession* (pp. 3-31). San Francisco, CA: Jossey-Bass Publishers.
- Beck & McKeown, (2001). Text talk: Capturing the benefits of read-aloud experiences for young children. *Reading Teacher*, 55(1), 10-20.
- Boston, M. and Wolf, M. K. (2004). *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., & Aschbacher, P. (2001). Exploring the technical quality of using assignments and student work as indicators of classroom practice. *Educational Assessment*, 7(1), 39-59.
- Clare Matsumura, L., Wolf, M. K., Crosson A., Levison, A. Peterson, M., Resnick, L., and Junker, B. (2004). *Assessing the quality of reading comprehension assignments and student work*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.
- Cohen, D. K, McLaughlin, M. W., & Talbert, J. E. (Eds). (1993). *Teaching for understanding: Challenges for policy and practice*. San Francisco: Jossey-Bass.
- Darling-Hammond, L. (1997). *The right to learn: A blueprint for creating schools that work*. San Francisco: Jossey-Bass.
- Dole, J. A. (2003). Professional development in reading comprehension instruction. In A.P. Sweet & C.E. Snow (Eds.), *Rethinking reading comprehension*. New York: Guilford Press.
- Elmore, R. F. (2002). *Bridging the gap between standards and achievement: The imperative for professional development in education*. Washington, DC: The Albert Shanker Institute.
- Elmore, R. F. (1996). Getting to Scale with Good Educational Practice. *Harvard Educational Review*, 66(1), 1-26.
- Fink, E. and Resnick, L. B. (2001). Developing principals as instructional leaders. *Phi Delta Kappan*, 598-606.

- Junker, B. W., Clare Matsumura, L., Crosson, A. C., Wolf, M. K., Levison, A., Weisberg, Y. & Resnick L.B. (2004) Overview of the Instructional Quality Assessment. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, April 2004.
- Lieberman, A. (1994). Teacher development: Commitment and challenge. In P. Grimmett & J. Neufield (Eds.), *Teacher development and the struggle for authenticity* (pp. 15-30). New York: Teachers College Press.
- Newmann, F. M., Lopez, G. & Bryk, A. (1998). *The quality of intellectual work in Chicago schools: A baseline report*. Chicago: Consortium on Chicago School Research.
- Resnick, L. B. and Hall, M. W. (1998). Learning organizations for sustainable education reform. *Daedalus*, 217(4), 89-118.
- Saunders, W., Goldenberg, C., & Hamann, J. (1992). Instructional conversations beget instructional conversations. *Teaching and teacher education*, 8, 199-218.
- Shepard, LA. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29, 23-28.
- Shepard, LA. (2003). Reconsidering large-scale assessment to heighten its relevance to learning. In J. M. Atkin & J. E. Coffey (Eds.). *Everyday assessment in the science classroom*. Arlington, VA: NSTA Press.
- Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1-22.
- Spillane, J. P. (2001). Investigating school leadership practice: A distributed perspective. *Educational Researcher*, 23-28.
- Stein, M. K. & D'Amico, L. (2002). Inquiry at the crossroads of policy and learning: A study of a district-wide literacy initiative. *Teachers College Record*, 104(7).
- Stein, M.K., Smith, M. S., Henningson, M. A., & Silver, E. A. (2000). *Implementing Standards-based mathematics instruction*. NY: Teachers College Press.
- Wolf, M. K., Crosson, A. and Resnick, L. (2004). *Classroom talk for rigorous reading comprehension instruction*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.

## APPENDIX A

### Abridged Version of the Principles of Learning

#### Clear Expectations

If we expect all students to achieve at high levels, then we need to define explicitly what we expect students to learn. These expectations need to be communicated clearly in ways that get them "into the heads" of school professionals, parents, the community and, above all, students themselves. Descriptive criteria and models of work that meets standards should be publicly displayed, and students should refer to these displays to help them analyze and discuss their work. With visible accomplishment targets to aim toward at each stage of learning, students can participate in evaluating their own work and setting goals for their own effort.

- Standards that include models of student work are available to, and discussed with, students.
- Students judge their work with respect to the standards.
- Intermediate expectations leading to the formally measured standards are specified.
- Families and community are informed about the accomplishment standards that children are expected to achieve.

#### Academic Rigor in a Thinking Curriculum

Thinking and problem solving will be the "new basics" of the 21st century. But the common idea that we can teach thinking without a solid foundation of knowledge must be abandoned. So must the idea that we can teach knowledge without engaging students in thinking. Knowledge and thinking are intimately joined. This implies a curriculum organized around major concepts that students are expected to know deeply. Teaching must engage students in active reasoning about these concepts. In every subject, at every grade level, instruction and learning must include commitment to a knowledge core, high thinking demand, and active use of knowledge.

## **Commitment to a Knowledge Core**

The ability to think well goes hand-in-hand with rich stores of knowledge. In each field of learning, there is a core of knowledge and conceptual understanding that all students should learn. This knowledge core should be specified in rigorous academic standards. The standards can then serve as the basis for an articulated curriculum in which core concepts are taught and learned in considerable depth, along with skills and tools of the discipline.

- There is an articulated curriculum in each subject that avoids needless repetition and progressively deepens understanding of core concepts.
- The curriculum and instruction are clearly organized around major concepts specified in the standards.
- Teaching and assessment focus on student mastery of core concepts.

## **High Thinking Demand**

Students will learn thinking abilities best when thinking is infused throughout the curriculum. Each subject should be taught in ways that press students to pose and solve problems, to formulate conjectures and hypotheses and to justify their arguments, and to construct explanations and test their own understanding. These high thinking demands, normal in programs for the gifted and talented, should be the daily fare of all students.

- In every subject, students are regularly expected to raise questions, to solve problems, to think, and to reason.
- Students are doing challenging, high-level assignments in every subject.
- Assignments in each subject include extended projects in which original work and revision to standards are expected.
- Students are challenged to construct explanations and to justify arguments in each subject.
- Instruction is organized to support reflection on learning processes and strategies.

## **Active Use of Knowledge**

People only acquire robust, lasting knowledge if they themselves do the mental work of making sense of it. Good teaching is a matter of arranging for students to do their own knowledge construction, while assuring that the ideas students develop will be in good accord with known facts and established concepts.

- Each subject includes assignments that require students to synthesize several sources of information.
- Students in each subject are challenged to construct explanations and to test their understanding of concepts by applying and discussing them.
- Students' prior and out-of-school knowledge is used regularly in the teaching and learning process.
- Instructional tasks and classroom discourse require students to interpret text and construct solutions.

## **Accountable Talk**

Talking with others about ideas and work is fundamental to learning. But not all talk sustains learning. For classroom talk to promote learning it must be accountable—to the learning community, to accurate and appropriate knowledge, and to rigorous thinking. Accountable talk seriously responds to and further develops what others in the group have said. It puts forth and demands knowledge that is accurate and relevant to the issue under discussion. Accountable talk uses evidence appropriate to the discipline (e.g., proofs in mathematics, data from investigations in science, textual details in literature, documentary sources in history) and follows established norms of good reasoning. Teachers should intentionally create the norms and skills of accountable talk in their classrooms.

## **Accountability to the Learning Community**

- Students actively participate in classroom talk.
- Students listen attentively to one another.

- Students elaborate and build upon ideas and each others' contributions.
- Students work toward the goal of clarifying or expanding a proposition.

### **Accountability to Knowledge**

- Students make use of specific and accurate knowledge.
- Students provide evidence for claims and arguments.
- Students identify the knowledge that may not be available yet which is needed to address an issue.

### **Accountability to Rigorous Thinking**

- Students synthesize several sources of information.
- Students construct explanations.
- Students formulate conjectures and hypotheses.
- Students test their own understanding of concepts.
- Classroom talk is accountable to generally accepted standards of reasoning.
- Students challenge the quality of each other's evidence and reasoning.
- Classroom talk is accountable to standards of evidence appropriate to the subject matter.

### **Self-management of Learning**

If students are going to be responsible for the quality of their thinking and learning, they need to develop—and regularly use—an array of self-monitoring and self-management strategies. These metacognitive skills include noticing when one doesn't understand something and taking steps to remedy the situation, as well as formulating questions and inquiries that let one explore deep levels of meaning. Students also manage their own learning by evaluating the feedback they get from others, bringing their background knowledge to bear on new learning, anticipating



learning difficulties and apportioning their time accordingly, and judging their progress toward a learning goal. These are strategies that good learners use spontaneously and all students can learn through appropriate instruction and socialization. Learning environments should be designed to model and encourage the regular use of self-management strategies.

- Within the context of instruction and learning in the various subject areas, metacognitive strategies are explicitly modeled, identified, discussed, and practiced.
- Students are expected and taught to play an active role in monitoring and managing the quality of their learning.
- Teachers scaffold students' performance during initial stages of learning, and then gradually remove supports.

## APPENDIX B

### Relationship between Checklist Ratings and Rubric Scores

Table 2  
Correlations between the reading comprehension and checklist items.

	AR1	Checklist A	Checklist B	Checklist C	Checklist D
AR1	-	-.60*	-.47	.01	.56*
Checklist A		-	.45	-.13	-.37
Checklist B			-	.39	-.28
Checklist C				-	.32
Checklist D					-

\*  $p < .05$ .

Note: AR1- Reading comprehension discussion; Checklists A and B include low-level thinking that correspond to rubric scores 1 and 2 respectively. As expected, these checklists have negative correlation with scores on the rubric. Checklist C and D include high-level thinking that correspond to rubric scores 3 and 4. As expected, these checklists have a positive correlation with scores on the rubric.

## APPENDIX C

### Accountable Talk Function Checklist

**Check all that apply and script relevant contributions.**

*(script here)*

#### **1. Linking contributions**

- Getting students to relate to one another's ideas

"Jay just said...and Susan, you're saying..."

"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..."

#### **2. Accountability to knowledge**

- 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?"

#### **3. Accountability to rigorous thinking**

- 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."