

Running head: MEASURING CHANGE

Measuring Change in the Instructional Practices of Literacy Teachers

Anthony Bryk, Stanford University

David Kerbow, University of Chicago

Gay Su Pinnell, The Ohio State University

Emily Rodgers, The Ohio State University

Carrie Hung, The Ohio State University

Patricia L. Scharer, The Ohio State University

Irene Fountas, Lesley University

Emily Dexter, Lesley University

Measuring Change in the Instructional Practices of Literacy Teachers

Abstract

We report on a study to develop a reliable and valid tool to examine changes in teachers' practice in the context of comprehensive literacy instruction. The Developing Language and Literacy Teaching (DLLT) rubrics are grounded in the interactive literacy theories of Clay (2001) and Fountas and Pinnell (2006), and in the summary findings on effective literacy practices from the National Research Council (Snow, Burns, & Griffith, 1998). The DLLT is also anchored in a conceptualization of teaching as a problem of expertise development (Bransford, Brown, & Cocking, 1999). The study presents empirical evidence that examines the construct validity of the DLLT observational system. A field trial was carried out with 78 teachers who were observed on 275 occasions. The internal consistency reliability for each of the eight scales that comprise the DLLT ranged from 0.63 to 0.91. Item-level adjacent category agreement exceeded 90% on every scale. Paired observations were conducted on 33 lessons to assess inter-rater reliability for each scale, which ranged from 0.78 to 0.91. Analyses were undertaken to examine the consistency between rubric item behavior and an hypothesized developmental pathway from novice toward more expert practice. We found large differences on rubric scale scores between novice teachers and those more experienced within comprehensive literacy instruction as predicted by the underlying conceptual framework. The results presented in this paper suggest that the DLLT is a promising new tool, useful both for scientific research and as a formative resource for guiding literacy professional development activities in comprehensive literacy instruction.

Measuring Change in the Instructional Practice of Literacy Teachers

This study was undertaken to create an instrument to measure teachers' development of expertise within the context of comprehensive literacy instruction in kindergarten through second grade. The Developing Language and Literacy Teaching (DLLT) rubrics, reported on here, are grounded in the interactive literacy theories of Clay (2001) and Fountas and Pinnell (2006), and a conceptualization of teaching improvement as a problem of expertise development (Bransford, Brown, & Cocking, 1999; Ericsson, Krampe, & Tesch-Romer, 1993; Ericsson, 1996, 2006; Glaser & Baxter, 2002). Building on these theories, as well as on prior small-scale research and clinical observations from literacy teacher coaching, we sought to create a set of observational rubrics that map instructional practice both within specific literacy activities and as teachers orchestrate students' experiences across multiple instructional activities that combine to form the literacy block in a typical comprehensive literacy classroom. Moreover, the instrument is intended to assess teachers' expertise development over time. In this regard, the DLLT sought to conceptualize and empirically test possible pathways along an instructional continuum toward this end.

Previous Research Using Observation Rubrics

A number of classroom observation systems have been created to assess literacy instruction. Most have focused on examining the frequencies of particular teacher behaviors at a fixed point in time rather than conceptualizing changes in skill level and how this might evolve over time. For example, in Juel and Minden-Cupp's study (2000), observers wrote running narratives while observing teachers during language arts periods,

and then coded their narratives in terms of instructional activities and the strategies the teachers taught for word-solving. From these coded narratives, they generated quantitative measures such as the percentages of activities that were phonics-focused. Other literacy-focused observational systems that are based primarily on frequency measures include Edmunds and Briggs (2003), Estrada (2004), Foorman and Schnatschneider (2003), and Greenwood, Abbott, and Tapia (2003). In all of these instruments, the variables of interest are defined in terms of the frequency of specific activities such as teaching phonics or asking non-literal questions. No direct efforts are made to assess the quality with which the activities are undertaken nor their appropriateness for the particular group of students being instructed. Most important in the context of this paper, teacher improvement over time is not directly conceptualized. Presumably it would be indicated by changes in the frequency of some observed teaching behaviors but how this should be scored is unclear.

A different approach has been to use Likert scales to rate instructional quality rather than frequency. An instrument developed for English Language Learners (Graves, Gersten, & Haager, 2004; Gersten, Baker, Haager, & Graves, 2005; Haager, Gersten, Baker, & Graves, 2003), for example, requires that observers rate items, such as “vocabulary and vocabulary concept development”, on a 1-4 scale from “not effective” to “very effective.” While this instrument directly rates the quality rather than the quantity of teacher behaviors and strategies, it does not include detailed descriptions of the kinds of behaviors that would warrant a score of 1, 2, 3, or 4. Teacher development means becoming “more effective” in an activity, but the rubric does not provide a descriptive

pathway as to what teachers might actually be doing differently over time as their instruction develops.

A few observation rubrics have developed sequenced descriptions of teacher behaviors along a continuum, with qualitative accounts of exemplary teacher behaviors as well as of less exemplary, presumably predecessor, behaviors. While two of these rubrics have included literacy items, none has been developed specifically to focus on K-2 literacy instruction. The Standards Performance Continuum (SPC) (Doherty, Hilberg, Epaloose, & Tharp, 2002), for example, rates teachers on five standards of effective pedagogy, one of which is language and literacy development. The SPC has a 5-point scale that assesses teacher behaviors as “not observed,” “emerging,” “developing,” “enacting,” and “integrating,” with descriptions provided for each rating. Similarly, the Instructional Quality Assessment (IQA) instrument (Crossen et al., 2006; Resnick, Matsumura, & Junker, 2006; Wolf, Crosson, & Resnick, 2006) rates teachers on a continuum of behaviors with most of its items written in a general form so as to be applicable to lessons in any subject.

Thus, the Developing Language and Literacy Teaching rubrics differ from extant instrumentation in several regards. First, the DLLT focuses specifically on comprehensive literacy instruction in kindergarten through grade 2, drawing on theories and research on reading development, notably Clay (2001), National Reading Panel (NICHD, 2000), and Snow, Burns, and Griffin (1998). It is organized around the activities typically associated with such instruction, such as guided reading lessons and writing workshops. Second, the DLLT seeks both to assess differences among teachers in their classroom instruction and to discern differences in how this activity might change

over time. Third, and closely related to this latter point, the DLLT is grounded in a perspective on the development of teaching as a process of moving from novice to expert understandings and behaviors. It is informed in this regard by core findings from research on expertise developments in a variety of different domains (Ericsson 1996; 2006). Finally, the DLLT also draws on the clinical experiences of the members of our team who train and support school-based professional developers and classroom literacy coaches through the Literacy Collaborative®, a K-6 school reform project. (For more information about Literacy Collaborative see www.literacycollaborative.org). Their clinical observations suggest that many teachers move through a development pathway of experimenting with new practices, focusing initially on getting a command over new procedures and instructional materials, developing some efficiency in their regular use, and gradually moving over time toward expertise in both planning lessons and in making moment-to-moment decisions that allow them to scaffold effectively the literacy learning of each individual student.

In sum, the DLLT seeks to differentiate among the forms of teaching activity likely to be seen in comprehensive literacy classrooms and to test whether it is sensible to conceive of these instructional practices as organized along a developmental trajectory from novice toward more expert teaching.

Conceptual Frameworks Anchoring the DLLT

The classroom observation system described in this paper is anchored in a theory of students' literacy learning and basic principles about a responsive instructional pedagogy to advance this learning. It also draws on major findings from research on

expertise to shape our perspective on the development of teachers' expertise in orchestrating literacy instruction to advance student learning. Finally, this teaching activity is contextualized within a framework of comprehensive literacy instruction. We proceed to detail each of these conceptual organizers for the DLLT, and some key interconnections among them. Woven together, these theoretical frameworks guided our choice of the specific instructional activities around which to organize the DLLT and informed the standpoints that we eventually adopted for discerning evidence of expertise development in the enactment of these activities.

A Perspective on Students' Literacy Learning

The study of language development provides convincing evidence that children are constructive learners (Lindfors, 1999). Through interaction with more expert speakers that scaffolds learning, they acquire words but more importantly derive the system of underlying rules that they can apply to construct their own sentences. Every child, all over the world, in about the same sequence and about the same time, develops language (DeStefano, 1978; Lindfors, 1987). Without formal instruction, speakers acquire a phonological system (the sounds that are meaningful within a language); a semantic or meaning system and the lexicon that represents ideas and concepts; and, a syntactic system that requires the arrangement of words in rule-governed ways to express meanings (Pinker, 2000). Acquiring literacy skills, however involves additional learning that for most children entails intentional instruction.

Letters, sounds, and words. Building on a basic foundation of oral language, literacy learning requires the additional acquisition of the letters and words that represent the sounds of a language (Adams, 1990). As they engage with written texts, children

expand their knowledge of the rules of syntax to accommodate new patterns and more complex rules. Exposed to reading and writing at home, young children often begin spontaneously to learn letters and words and to experiment with them; their early attempts provide evidence of constructive learning (Bissex, 1980; Ferriero & Teberosky, 1982; Read, 1971). But most children require supportive teaching to acquire understanding of letters and their relationship to the sounds of oral language (Ehri, 1991, 1998; Goswami & Bryant, 1990). Thus, literacy learning is based on the natural process of acquiring language but also requires both direct and embedded instruction in order for students to learn the complex relationships between graphic signs and the language systems (Dahl, Scharer, Lawson & Grogan, 1999). Knowledge of phonemic awareness and the relationships between sounds, letters, and words is essential in the process of learning to read (Juel, 1988; Kamil & Manning, 2002). Building on this, Clay's interactive theory of literacy processing, which is informed by Rumelhart's (1994) and Singer's (1994) models of the reading process, describes how young learners use these understandings in an orchestrated way while reading and writing continuous text.

Reading. Accomplished readers have acquired a processing system that consists of an integrated set of strategic actions by which readers extract and construct meaning from written language (Ruddell, Ruddell, & Singer, 1994). Readers understand and access many different kinds of information as they employ systems for making sense of print (Rumelhart, 1994). Readers are engaged in complex thinking that is largely invisible to them as a process. They concentrate on ideas without conscious awareness of what is happening inside the human brain, as text processing and problem-solving is taking place. This cognitive activity requires "getting access to and working with several different

types of information to arrive at a decision” (McCoon and Ratcliff in Clay, 2001, p. 80). It utilizes all aspects of relevant knowledge—letter and word recognition; connections to language knowledge; and accessing of content, personal, and text knowledge.

Readers must process a wide range of information, both visible and invisible, to read a text successfully. Visible information includes everything that can be seen on the page: the punctuation, layout of text, and the art. Readers must also draw on invisible relationships while reading. These invisible relationships exist in the reader’s brain and include such things as knowledge of the world; knowledge about texts; phonological, semantic and syntactic information; as well as all of the experience the reader brings to reading the text. Texts provide the opportunity and actually demand that readers “mix” visible and invisible information. The reader is constantly processing both kinds of information (Clay, 1998, 2001) in what Rumelhart (2004) refers to as a process that is at once perceptual and cognitive.

These complex operations occur simultaneously as readers meet the demands of processing texts. In many ways, the processing system is built as readers employ strategies for decoding and understanding a range of texts (Meek, 1988). Thus, when we say that a text places demands on a reader, we are really asking: “What must the reader do in order to read this text with understanding?” No matter how simple the text, the answer always entails much more than just decoding the words.

If reading ability is developed through effective processing of texts, then it makes sense that instruction takes into account what the learner presently knows in the selection of appropriate texts and the teaching that surrounds them. Readers who are struggling may be employing ineffective strategies and building emotional blocks (Lyons, 2003).

Targeted small group instruction aims to match books to readers and provide teaching designed to help readers effectively process ever more challenging texts (Johnston, 1997).

Writing. Young writers, too, are engaged in an extremely complex process of skill development (Calkins, 1994). They are challenged to compose written language using syntax that is different from oral language. Keeping a meaningful message in mind while constructing written text letter-by-letter requires coordination of many different kinds of information—the particular sequence of words arranged in syntax to convey a message; the specific letters needed to form each word (as related in both simple and complex ways to the sounds of language); word parts, sometimes unrelated to sounds, that nevertheless appear as building blocks of words; and the directional movements needed to write the letters in order. In short, writers too must orchestrate large amounts of information simultaneously in this processing activity.

Some aspects of writing may be directly demonstrated to learners; but, ultimately, there is a need for students to engage in the processes of writing so that they experience and practice processes of rehearsing/planning, drafting/revising, editing and proofreading, with supportive instruction in every phase (Graves, 1983, 1994; Short, Harste, & Burke, 1996). Like reading, writing is a process of creating and communicating meaning (Johnston, 1997), and by engaging in the process with this goal in mind, writers have an opportunity to learn more with each piece that is written.

The reading and writing relationship. Reading and writing are different but highly interrelated and complementary processes, and this applies to the building blocks of written language as well as to the organization and structure of written texts.

In terms of learning about how written language works, writing supports reading by revealing the building blocks of words and helping students attend closely to the way sentences are constructed (Elkonin, 1971, 1973; Clay, 2005). It encourages young children to say words carefully and consider sound-to-letter relationships. In writing words, they think about syllables and larger word parts; they make connections between words to solve them (Rodgers, 2007).

The texts that students read or that are read aloud to them provide models of language that can enrich students' writing. Ray (2004) refers to a mentoring relationship as students examine the language and structure of texts and actively learn from them as resources for their own writing. Reading supports writing by providing thousands of models of written language that help students expand their knowledge of complex syntax as well as vocabulary and concepts. Conversely, the composition process can make students more aware of aspects of the writer's craft so that they can think analytically about the texts they are reading.

In addition, learners can use writing as a tool to deepen comprehension and extend thinking. Writing about reading helps students carefully consider texts and compose sentences and longer pieces that express their thinking; in the process, they extend their understandings.

Implications for Literacy Instruction

Quality literacy teaching involves integrating, across several bodies of knowledge about the cognitive processes shaping development in reading and writing, an understanding of children's varied trajectories in acquiring these literacy competencies, and developing a repertoire of effective classroom techniques for advancing various

literacy learning objectives. Accomplished literacy teachers deftly engage students across the multiple activities that typically comprise comprehensive literacy instruction, while constantly making adjustments and tailoring interactions in ways that recognize students' strengths, instructional needs and the immediate goals for learning.

In most general terms, this view of instruction is anchored in Vygotsky's conceptualization of cognitive development as progress through a zone of proximal development (1978). Drawing on these ideas, Rogoff (1997), for example, describes student learning (and the teaching that is needed to support it) as a process of guided participation through increasingly complex activity. Similarly, Tharp and Gallimore (1988) refer to this phenomenon as assisted performance. Instruction is continually adjusted as informed by the evidence of understanding observed in learners' responses (Wood, Bruner, & Ross, 1976).

This general perspective on instruction, in combination with the observations offered above about students' literacy learning, has several implications for focusing the DLLT's observation of instruction in literacy classrooms. It directs attention to:

Teachers as careful observers of their students' reading and writing behaviors, gathering data continuous to inform decision making. The first and most essential element in excellent instruction is the "noticing teacher" (see Clay, 2001). Anchored in an understanding of the reading and writing processes and how these skills develop, teachers systematically observe and interpret student behavior in every literacy learning context. They form tentative theories about each student's current processing system and design teaching plans, as well as make decisions on-the-run, drawing on these evolving ideas.

Based on such observation, teachers differentiate subsequent instruction in a way that recognizes what students already know and supports them in further learning.

Noticing teachers have a wide repertoire of ways of working with students. They not only implement instructional approaches with skill, but also use information gathered “on the run” to plan lessons and modify instructional interactions in ways that make learning experiences powerful for individuals and groups.

Teachers follow a principle of guided practice, aiming to assure student success while simultaneously working to advance student independence. Teachers initially introduce and demonstrate new skills, strategies, and processes to students and then support them as they engage in reading, writing, and talking. This kind of support makes it possible for students to work on the edge of what they know, expanding their understandings as they do so (Vygotsky, 1978). Gradually, support is withdrawn so that the student move toward working independently and instruction then moves to a higher level of sophistication and the process continues.

Teachers engage students in reading a range of texts for different purposes.

Students need to read texts at a level that will allow for maximum learning (Fountas & Pinnell, 2006). They engage in a great deal of reading at an independent level because quantity of reading matters (Anderson, Wilson, & Fielding, 1998). In addition, students receive supportive teaching that enables them to take on more complex texts, thus expanding their processing systems. Teachers enrich students’ exposure to text through selecting and reading aloud a wide variety of texts. Engagement with many different genres provides the foundation for extended discussion that supports student in expanding their thinking about texts.

Instruction combines reading and writing so that students can use the connection between the two to strengthen learning in both areas. As noted earlier, when students are engaged in writing, children are also learning a great deal about how written language works. Conversely, when they engage in reading, they are encountering the syntactic patterns and the forms of written language. When they hear and discuss texts, such as during an interactive read aloud, children are gaining vocabulary and learning how texts are structure and organized, which adds to the resources they employ while reading and writing. Teachers who understand these connections can better advance student learning through orchestrating their diverse teaching activities across the multiple instructional contexts that form the school day.

Teachers provide explicit instruction that helps students understand letters, sounds, and words as well as how to apply this knowledge to reading and writing. These building blocks of written language are basic to students' ability to read and write texts with fluency and must be directly taught. The faster and more automatic students are with these operations, the more attention is freed for thinking about the meaning of texts (NICHD, 2000). In addition, the skilled teacher both provides specific lessons and also demonstrates, prompts, and support students as they apply their knowledge in various reading and writing contexts (Pinnell & Fountas, 1998).

A Perspective on Teacher Development of Expertise

The DLLT is also informed by core principles about the nature of expertise development from studies in diverse fields. We have sought to marry to these core principles rich clinical details about comprehensive literacy teaching. These clinical accounts are rooted in professional development activities, extending over many years,

aimed at improving literacy instruction in many different and diverse classroom and school contexts.

A significant body of cognitive theory and research in the learning sciences across multiple domains suggests that expertise emerges through guided practice over time (Ericsson, Krampe, & Tesch-Romer, 1993). At its core, it involves not only acquiring more knowledge but also “a change in the nature and use of knowledge” (Glaser & Baxter, 2002, p. 181; see also Bransford, Brown & Cocking, 1999). Experts are able to recognize quickly patterns in complex phenomena and use this as a basis for subsequent actions. (On the latter point see, for example, Flyvbjerg, 2001, pp. 9-24). In contrast novices "may not be able to recognize the specific features of a problem to which experts readily attend" (van Es & Sherin, 2002, p. 576).

These principles ring true in clinical observations about comprehensive literacy instruction. As noted above, such instruction involves a complex mix of processes where: teachers analyze students’ current strengths and needs; select specific teaching approaches and establish strategic emphases based on these observations; continually reassess the results of their instruction to inform subsequent teaching moves; all the while maintaining warm and trusting relationships with children and engaging their interests (see for example Russell & Munby, 1991). More expert teachers are able to integrate quickly into their subsequent actions evidence about where children are in their development as readers and writers, with their theoretical knowledge about the reading and writing processes, and how instruction is or is not advancing immediate learning goals.

Research on expertise also documents that such expertise is highly situated depending on substantive context (Bransford et al., 1999). This insight that expertise is not generic but rather operates in specific activity domains, where individuals must integrate detailed knowledge with prior experience in that domain, has important implications for our efforts to develop and test an assessment system for teachers' literacy practice. Rather than assuming that teacher expertise in literacy instruction is similar across all instructional systems, we sought to anchor our assessment efforts within the instructional system of comprehensive literacy and the materials, tools, routines and pedagogical practices that it affords. The DLLT seeks to assess teachers' capabilities to bring their accumulated knowledge, skill, and experience to bear on the work activity that occurs day-to-day in such classrooms.

In this context, more expert teachers have an understanding of the core instructional components of comprehensive literacy, are able to enact each efficiently, and most importantly, discern how to leverage each to advance the learning of each child, group, and class of students. To be sure, this kind of pedagogical skill entails significant procedural knowledge about particular instructional techniques but also moves well beyond it. In a real sense, it is teacher's expertise that makes the program, material, or instructional technique "work" (Bond & Dykstra, 1967).

In comparison, extensive clinical observations on novice teachers within comprehensive literacy instruction (see for example Neufeld & Roper, 2003) suggest that teachers relatively new to these practices tend to rely on their newfound procedural knowledge and consequently focus attention on the more procedural aspects of their instruction (e.g., selecting an appropriately-leveled text for guided reading or offering an

introduction to the book). They are less likely to responsively differentiate activity based on the needs and cues of individual learners.

A Context of Comprehensive Literacy Instruction

Comprehensive literacy instruction typically consists of several interrelated instructional activities, each of which has particular purposes and each of which must be examined as a distinct instructional setting. Small-group reading instruction, for example, aims to ensure that students learn to comprehend written texts (Pearson & Fielding, 1991; Pressley, 1998), as well as learn to use phonics skills to take words apart while reading for meaning (Pressley, 1998; Snow, et al., 1998). Through such activity, teachers aim to address comprehension and vocabulary development while simultaneously providing explicit instruction in reading fluency (NICHD, 2000; Pinnell et al., 1995). Many teachers of comprehensive literacy also provide daily mini-lessons on conventions, skills, and the craft of writing. Instruction in writing, in fact, contributes substantially to children's understanding about words (Clay, 1991; NICHD, 2000) as they learn to hear the sounds in words (phonemic awareness) and learn to look at letters and words (Liberman & Shankweiler, 1985; Vellutino & Scanlon, 1987; Lundberg, Frost, & Petersen, 1988) in ways that support both reading and writing achievement.

Primary instructional activities. Although many variations exist in practice, comprehensive literacy instruction is typically organized around some or all of the following literacy activities:

1. *Interactive read-aloud (usually whole-class instruction).* Teachers read aloud to students an array of texts that are carefully selected to help students think in various ways about texts. The teacher uses intentional conversation (conversational moves directed

toward a goal of instruction) and also promotes routines such as “turn and talk” to help children learn how to talk with each other about texts. The opportunity to engage in “text talk” is rich (Beck & McKeown, 2001). The teacher is decoding the words of the text by reading it aloud, but in every other way, young students are processing it and expanding their understanding through talk that is grounded in texts (Fountas & Pinnell, 2006).

2. *Shared reading (whole-class or small-group instruction)*. In shared reading, the teacher and children read from a common text. Usually, the text is read several times. Group support helps students to process more difficult texts than they could read independently, although it is still important to match the complexity of the text to the group. Using this familiar text, the teacher makes appropriate teaching points that extend children’s understanding of the reading process (McCarrier, Pinnell, & Fountas, 2000).

3. *Guided Reading (small group instruction)*. Students are grouped because they are similar in their reading development at a point in time. The teacher selects a text that is appropriate for the group, introduces it in a way that will help students read it effectively, supports individuals during reading as needed, and invites students to discuss it afterward. The teacher makes strategic teaching points directed toward key aspects of the reading process (Fountas and Pinnell, 1996).

4. *Interactive writing (whole-class or small-group instruction)*. In this highly supportive context, children can fully participate in the writing process (McCarrier, Pinnell, & Fountas, 2000). The teacher and children collaboratively compose a text and then write it word by word on a large chart. At several carefully selected points, the teacher invites individual children to come up to the chart and make contributions by adding letters or words. These occasions have high instructional value in helping children

learn the construction of words (phonics) as well as important aspects of the writing process.

5. *Writing workshop (whole-class and individual)*. The teacher provides a mini-lesson on some aspect of writing; then students write independently as individuals confer with the teacher. At the end of the workshop, there is a brief sharing period during which the teacher can reinforce the mini-lesson principle and invite writer-to-writer feedback. Students write daily, applying critical principles to their own production of writing in a range of genres.

6. *Word study (whole class)*. The teacher provides a mini-lesson on phonics and students apply the principle independently. While phonics and word study are embedded in all the previously described contexts, here the instruction is preplanned, direct, and explicit. The emphasis is to teach directly important principles related to how words work and/or the rules of English spelling (Pinnell & Fountas, 1998).

A matrix organization. The reading and writing activities described above constitute a repertoire of practices that good teachers weave together based on their pedagogical knowledge and their observation of children. Rather than breaking instruction down into highly discrete components, each of which may be intended to advance a specific sub-skill, multiple learning objectives can simultaneously be advanced through any of the six specific activities described above. As a consequence, effective instruction demands facility in recognizing and being able to act on the complex interconnections, opportunities, and constraints that may exist among multiple goals being pursued for students, and the instructional activities at a teacher's command to use.

A considerable demand is placed on the basic social organization of the classroom as a great deal of this activity occurs simultaneously during comprehensive literacy instruction. Unless these general aspects of teaching are solidly in place, the routines of instruction may easily break down and the intended student learning is unlikely to occur. Building on and extending out from this, strong teachers strategically “echo” key ideas across activities to make instruction more powerful within each. In its most expert rendition, comprehensive literacy instruction entails a thoughtful orchestration of activity across the multiple instructional activities described above. Deeper processing, such as comprehending, is not learned in discrete lessons that direct children to practice a “strategy” as if it is one skill to be learned from one book. Rather, teachers continually remind children of aspects of processing across the day and in connection with many texts that they read or write. This teaching for strategies (Fountas & Pinnell, 2006) is key to helping children build effective systems for reading and writing. It represents a meta-principle for the instructional system that provides the context for the teacher expertise development that we sought to measure in this study.

The Developing Language and Literacy Teaching Rubrics

The development of the DLLT built on insights from an earlier small-scale study (Lyons & Pinnell, 2001) that sought to examine the relationship among literacy coaches’ practice, the nature of the use of comprehensive literacy instruction in classrooms, and students’ annual learning gains. Based on a long-term observational study of ten teachers, Lyons and Pinnell conceptualized three scales for selected elements in comprehensive literacy instruction and found that student learning gains were greatest in classrooms where teachers scored high on these scales. Subsequently, Literacy Collaborative staff

crafted a larger set of clinical protocols to guide literacy coordinators in their efforts to provide professional development for teachers. These prior activities encouraged us that more general and valid assessments could be designed.

Specifically, we sought to use the conceptual frameworks introduced above to detail and extend these clinical tools. Considerable attention was given to specifying the language of the response for each rubric element in each scale based on our conceptual frameworks. Following on the idea that expertise is context- and task-specific, we chose to organize the DLLT around the six specific instructional activities commonly found in comprehensive literacy instruction. Thus, the core of the DLLT consists of six separate rubrics, one for each activity, designed to represent the range of teaching practices likely to be seen in classrooms where teachers are operating at varying levels of expertise within that activity. In addition, we added two more holistic assessments—one focusing on teachers' efforts to organize their classrooms; and a second, on teachers' efforts to orchestrate the enactment of literacy practices across the multiple instructional components of comprehensive literacy in ways that maximize student learning. (See Figure 1 for the list of eight rubrics). Some additional comments on the last two lesson-holistic rubrics are offered below.

[Insert Figure 1, Eight Areas for Rubric Development, about here.]

General aspects of teaching. This rubric focuses on the basic social organization of the classroom to support quality instruction. These attributes are important in any classroom but are especially so in a comprehensive literacy framework because a great deal of activity is occurring simultaneously at any given time. The observer uses this

rubric to evaluate classroom materials and organization, student engagement, the quality of teacher-student interactions, and sense of community among students.

Teaching for strategies. This was designed to further differentiate truly expert practice from more mid-level efforts. The goal of reading instruction is to have students learn ways of thinking—literal, inferential, and analytic while reading fiction and nonfiction texts; word solving (including phonics and word analysis); and fluency (including phrasing and intonation). Likewise, instruction seeks to help students become fluid, expressive writers who understand how texts are organized, who can logically advance an argument, and who have developed their own voice. Accomplishing these goals entails coordinating students’ learning experiences across all literacy activities so that they “add up” to produce strong, self-extending readers and writers.

An Elaborated Example

Rather than attempting a brief description here of all eight rubrics, we have chosen to illustrate the interweaving of the ideas introduced above in the context of one rubric, Guided Reading. Some seven separate elements comprise this rubric. (See Figure 2.) Each element (or row in a rubric) focuses on a specific key aspect of this instructional activity. The ratings in each row move from simple descriptions of base practice at level one to a cumulative description of expert practice at level four. The narrative descriptions are intended to provide a sense of what change in practice might look like for each of these aspects of instruction (rather than just giving a numeric rating). To deepen our illustration, we offer an explanation for the scoring of element six that focuses on the selection of teaching points after small group reading.

[Insert Figure 2, Guided Reading Rubric, about here]

A teacher who is just beginning to provide small group instruction (guided reading) might begin by simply going through the procedural steps. The lesson might *look* like guided reading, but the teacher is operating on basically the same theory as previously used while implementing other approaches. The lesson has the surface characteristics of a reading group rather than the intensive teaching for strategies characteristics of guided reading when well executed. For example, after the reading, the teacher may not respond explicitly to what she has observed as students read. The score would be a 1—“Makes no teaching points, even though there were opportunities to do so.”

After learning more (possibly through coaching and professional development), the same teacher might begin to try to make teaching points and the score could be a 2—“Makes teaching points, but they do not help students to engage in effective processing of texts.” Here, the teacher is taking on important aspects of guided reading but not yet connecting with where students are as developing readers and providing instruction based on observations of student reading behaviors. Over time, this teacher might eventually score a 3—“Makes teaching points, but not all of the teaching points help students engage in effective processing of text.” Here, the teacher is doing a good approximation of the approach and having a positive impact, but is not always responsive to students’ learning.

Finally, at the highest level on the rubric at a score of 4, the teacher “makes superbly chosen, specific teaching points that help students engage in effective processing of texts.” At this high level of expertise, the teacher draws on precise observations of patterns in students’ reading behavior to make teaching points that have

strong generative value, enabling the readers to gain processing power that can be applied to other texts. For example, a teacher who observed a pattern of children's lack of attention to errors might revisit places in the text to show children how to think about whether what they read made sense and "looked right," a teaching decision that would help readers develop their ability to check on themselves as they read any new text.

The full rubric for guided reading requires the rating of each aspect of instruction represented by the seven rows. The teacher may be more or less proficient on the seven different elements that comprise this rubric. That is, the teacher could be strong at supporting individual students during reading but relatively weak in selecting and engaging students in explicit teaching points after reading. The end result is a guided reading profile, consisting of multiple scores, for any observed lesson. A similar logic played out in conceptualizing the rubrics for the remaining five instructional components.

In sum, the overall goal for each rubric is to distinguish procedural teaching (or "going through the motions") from more expert practice. Taken together as a set of eight rubrics, the DLLT assesses teachers' development within the separate components of the instructional framework, as well as teachers' capability to interweave these elements into effective overall instruction. Embedded within each rubric is an explicit theory of student development as a reader and writer, and a corresponding language and literacy framework aimed at advancing this. This in turn is melded to principles about expertise development as this might apply in the context of comprehensive literacy instruction. Overall, 51 different elements of comprehensive literacy instruction are assessed in this fashion by the DLLT.

Field Testing, Reliability, and Validity Analyses

Next, we describe the design for field testing the rubrics, the training protocol for their use, the actual data collected and the analyses undertaken.

Data Collection Design

A field trial of the rubrics was carried out by 22 literacy coordinators in 17 schools in 4 districts. Nine of the schools were in one district where 50-89% of the students qualified for Free or Reduced Price Lunch (FRPL). Six schools in two other districts had 47-74% FRPL. Approximately 40 percent of the students in each of these 15 schools were minority. In contrast, less than 20% of the students in the last two schools in the fourth district qualified for free and reduced price lunches and about 20 percent were minority.

Each school had an experienced literacy coordinator, trained by the Literacy Collaborative at The Ohio State University, who agreed to use the rubrics to observe and document literacy activities in selected classrooms. The basic design plan called for each coordinator to observe four teachers, one at each grade level K–3, on five occasions (once a month) from January through May of 2005. Coordinators within each school were asked to select two novice teachers (two years or less experience teaching within a comprehensive literacy framework) and two more experienced teachers (three or more years of teaching with the framework) to include in the study. Novice teachers were not necessarily ones with limited teaching experience. The criterion used to determine novice versus more experienced teachers focused only on the number of years that each teacher had been implementing a comprehensive literacy approach in their classroom. We asked

the literacy coordinators to observe a full literacy block (between 90 and 150 minutes) and then rate each lesson according to the draft rubrics.

In order to obtain data on inter-rater reliability, a subset of paired observations was designed for both the initial and final time points. On these occasions, a participating coordinator from one school traveled to another school in the same district so that the two coordinators could jointly observe but independently rate the same set of lessons. In order to minimize the overall data collection burden on the participating literacy coordinators, the paired observations conducted at time points 1 and 5 counted toward their overall commitment to conduct up to twenty observations apiece for our instrumentation study. This resulted in a modification to the basic design plan with some teachers observed on five occasions, and others on only three occasions (time points 2, 3, and 4) in order to make the gathering of the inter-rater reliability data feasible.

Training Protocol for Using the DLLT Rubrics

The literacy coordinators in the field study were involved in a two-day intensive training session on the use of the rubric. This training involved a structured introduction to each part of the rubric.

1. Researchers described the overall rubric, for example, for an interactive read-aloud, including why particular aspects were being highlighted for observation while other details may be omitted. Discussion followed about the wording and developmental distinctions that might be observed for the particular practice or rubric element.

2. The group observed a complete lesson on video and each participant independently rated the lesson using the rubric. This was followed by a discussion in the large group to clarify what they had seen and to reach consensus about the rating.

3. A second video of the same instructional component (enacted by another teacher) was then viewed and independently scored by each literacy coordinator. This time the literacy coordinators moved to small groups to share their ratings and to discuss the rubric. These small-group discussions were audio and videotaped for further analysis by the researchers (Rodgers & Hung 2006).

Actual Data Collected

We could not fully implement the design as planned in every school for a variety of reasons including the school structure (e.g., only K–2 enrollments), the demography of the school staff (e.g., no novice teachers), the role assignments of the participating literacy coordinators (not all coordinators work across the K–3 grades), and the refusal of a few individual teachers to participate. In addition, in a small number of cases, teachers who originally agreed to participate subsequently had to drop out, and the full observation protocol could not be completed for them.

In the end, we obtained useable rubrics from 275 literacy block observations. These observations came from 78 classroom teachers in the 17 schools. Of this group, 37 teachers were novices within a comprehensive literacy framework, and 41 were more experienced. The average literacy coordinator observed 3.5 teachers and the average teacher was observed on 3.5 occasions. In addition, paired observations were conducted on 33 lessons at time points 1 and 5, respectively. Each coordinator was involved in at least one joint observation at both of these time points.

Using an Item Response Theory Model to Examine Scale Properties

Each rubric, consisting of 3 to 10 separate elements, is independently rated on a four-point descriptive scale. A primary analytic task for this study was to examine

whether these multiple elements combined together into a theoretically interpretable and empirically defensible scale given the conceptual framework detailed earlier. To accomplish this, we analyzed our data using Rasch analysis (Wright & Masters, 1982). A similar approach was taken by Rowan, Camburn, and Correnti (2004), who used Rasch analysis to analyze the difficulty levels of items on their teacher logs.

The Rasch Rating Scale model is an extension of an item response latent trait model that is now commonly employed in standardized test construction. Instead of the dichotomous (correct/incorrect) responses in an IRT model, the Rating Scale model is designed for ordered categorical data as in our four-point rubrics. The Rating Scale analysis empirically defines a measurement scale that is anchored in the relative probability of a teacher's practice being scored as corresponding to the categorical description associated with each element that comprises a rubric. In a properly fitting scale, the rubric elements and subcategories arrange in a hierarchical order. Teachers are subsequently "measured" on this scale based on the practice ratings recorded for each of them. The scale units are measured in logits (i.e., the log odds of being in a particular score category on a given element relative to an overall base category.)

A Rating Scale analysis produces a diverse array of statistics for examining the quality of the underlying measurement system. First, there are element difficulty statistics, which estimate the likelihood that a teacher will score highly on each rubric element. Pedagogical skills that are relatively easy for teachers to acquire, such as selecting engaging and appropriate books in the Interactive Read Aloud rubric, tend to have lower difficulty estimates than more advanced skills, such as facilitating strong discussion during the reading of the book. Thus, how the element difficulties locate

themselves within their respective scales provides critical evidence as to whether the rubric elements are empirically sequencing in a theoretically reasonable order.

Second, and also important, are a set of element infit statistics. These estimates provide information about the degree to which the teacher ratings on a particular element are consistent with the element's placement in a hierarchically ordered scale. Individuals who are rated high on a particular element should be more likely to be rated high on the "easier to acquire skills" that are below it in the scale and less likely to demonstrate competence on the "more difficult" or more expert elements that appear above it in the scale. In essence, the infit statistic measures the degree to which the element's behavior deviates from what we would expect in a perfect developmental scale. Under the Rasch Rating Scale model, a probability value can be attached to the estimated infit statistic associated with each element within a rubric. The underlying null hypothesis associated with this test statistic is of the form "assuming that this element is part of a hierarchically ordered scale, how likely is it that we could get an infit statistic this large by chance alone?" Thus, this aspect of the Rasch analysis provides information for judging whether it is appropriate to view the combination of elements in a rubric measure as forming a coherently ordered scale.

Third, the Rating Scale analysis provides a standard set of statistics about the internal consistency of each scale. The "person reliability statistic" generated in the Rating Scale analysis is a reliability measure similar to Cronbach's alpha.

Evidence on the Reliability of the Rubrics

Below we present inter-rater reliabilities for the 51 separate rubric elements as well as internal consistency reliability and inter-rater reliability for the eight rubric measures.

Inter-rater reliabilities of the rubric elements. Obtaining good inter-rater reliability with observational rubrics can be a nettlesome problem, especially when the rubrics are being used by school-based staff (rather than trained researchers) under ordinary field conditions. In response to this concern, we focused each element within each rubric on an observable and distinct instructional activity. We aimed for clear descriptive language in defining each of the steps within each observation element. We then used the paired observations at time point 1 to help us discern how well our first draft descriptions actually worked when employed by different literacy coordinators. Table 1 presents the results from these observations. We report here on both the exact agreement and adjacent category agreement between raters (i.e., ratings within one category of each other). We initially performed this analysis separately for each element in each rubric. We then averaged across the elements that comprise each rubric to create the rubric-level agreement statistics reported in the table.

The inter-rater reliabilities from time point 1 helped us to improve the draft rubric specifications. While most of the initial rubric elements appeared to work well, we made nine changes in the pilot instrument based on results from the time point 1 analysis. Specifically, we dropped one element each from Shared Reading and Guided Reading where inter-rater reliability seemed problematic. We clarified the descriptions on five other elements (one in Guided Reading, two in Word Study, and two in Teaching for Strategies), and we added one entirely new element to the Teaching for Strategies rubric. This revised set of scales was then used in the observations at time points 2 through 5. We also added at this point another component to our training protocol where an experienced staff member from the Literacy Collaborative at The Ohio State University

jointly observed a lesson with each literacy coordinator at their own school. Both individuals independently rated the lesson and then discussed their differences, if any.

Taken together, these changes resulted in a substantial improvement in the inter-rater agreements element by element from time points 1 to 5. On every rubric, the average adjacent category agreement now exceeded 90%. The exact agreements, while lower, improved substantially to around 60% for all of the scales.

[Insert Table 1, Inter-rater Agreement for the DLLT Rubrics, about here.]

Finally, we also computed a weighted Kappa statistic (Fleiss 1981; Landis & Koch, 1977) on the time point 5 data for each element in each rubric (see Table 2). This statistic provides a direct assessment of agreement in ordered categorical data, which is the type used in our rubrics. The statistic combines information on both categorical agreement and the degree of disagreements to produce an overall index of inter-rater reliability. A Kappa ratio of 0.80 indicates outstanding agreement while values from 0.60 to 0.79 represent substantial agreement. Values from 0.40 to 0.59 are considered moderate inter-rater reliability.

In terms of the DLLT, 20 individual rubric elements had weighted Kappas of 0.70 or higher. The weighted Kappas were less than 0.50 on only six elements and between 0.50 and 0.60 on another 13 elements. Overall, this suggests reasonably good field agreement element by element within each rubric. We undertook one final review of the six elements with Kappas less than 0.50, but found no reason to introduce any further changes in any of these.

[Insert Table 2, Weighted Kappa Statistics for Each DLLT Rubric Element
(Time Point 5), about here.]

Internal consistency reliability for the final rubric measures. Table 3 provides reliability estimates from the Rasch Rating Scale analysis on the eight separate rubrics that compose the DLLT. Six of the eight rubrics achieved internal reliability of 0.75 or higher. The reliability for the Interactive Read Aloud and Shared Reading rubrics were weaker, at 0.63. This is not surprising, given that these two rubrics consist of only three rating elements apiece. As expected, the estimated reliability was higher for those rubrics with a larger number of rating elements. For example, Teaching for Strategies, which consists of ten rating elements, has an internal consistency reliability of 0.91.

[Insert Table 3, Internal Reliability of the DLLT Rubrics, about here.]

Our Rating Scale analysis suggests that each of the rubrics has good internal consistency and that most of these scales, with the exception of Interactive Read Aloud and Shared Reading, can be used separately in subsequent analyses.

Inter-rater reliability for the final rubric measures. Finally, we computed correlations for the eight rubric measures across the thirty-three paired observations conducted at time point 5. These are reported in Table 4. All of these correlations are 0.80 or higher, except for Shared Reading, which was 0.78. These results indicate quite good inter-rater agreement at the rubric measure level.

[Insert Table 4, Correlations among Measures for Paired Observations at
Time Point 5, about here.]

Evidence for Examining Construct Validity

As noted earlier, the Rasch analysis also provides evidence for examining the construct validity of our rubrics as measures of teachers' expertise development in comprehensive literacy instruction. Specifically, scrutiny of the item-difficulty statistics

allows us to investigate whether the empirical ordering of the items follows a theoretically expected pattern within each rubric. Practices that teachers find easy to use in their classrooms should have low item difficulty. Those items that are at the higher end of the difficulty scale should represent more expert practices that are more difficult to instantiate in classrooms and as a result are less frequently observed. Thus, by examining the ordering of the estimated difficulties within each rubric we have a face validity test of each scale. Quite simply, do the difficulty estimates form an interpretable development of expertise sequence?

Also of value are the misfit statistics that evaluate the consistency of the item map across teachers. If the scale represents a general profile that most teachers follow for expertise development within comprehensive literacy instruction, we would expect small item-misfit statistics. In contrast, the presence of large misfit statistics would indicate that the elements of the rubric do not cohere as an overall development of expertise measure. Taken together with the item difficulties, the associated misfit statistics provide empirical evidence to examine the tenability of our assumptions about what movement from less to more expert practice means substantively. We detail below our analysis of the results for Guided Reading, Writing Workshop, and Word Study. The remaining difficulty maps can be found in Appendix A.

Guided Reading. Figure 3 displays this item difficulty map. Each bullet-pointed description represents an element (i.e., a row) of the rubric. The easiest item, at the bottom of the map, is “selection of an appropriate text for the reading level of the group.” Engaging students around such a text with appropriate teacher guidance has potential for advancing significant learning. The selection of an appropriately leveled text is the most

basic routine to establish in a classroom where guided reading is being introduced. If the text is too easy, and children can already read it independently, there may be little opportunity afforded for the strategic teaching necessary to move students to the next level in their reading development. In contrast, if the text is too difficult and students are unable to process the text independently, they are unlikely to experience success as a reader or grow in their ability to read increasingly more difficult texts.

Considerably more difficult for a teacher is the preparation of a book introduction that engages students in a conversation about the text and its meaning. This instructional activity is key in preparing students to work with ideas they will encounter in the text that are challenging or new to them. A good book introduction draws out students' prior knowledge and facilitates their access to new information. It can also involve talk about potentially demanding characteristics of the book and sets the stage for successful processing of vocabulary, figurative language, or other aspects particular to the text. To accomplish this well requires integrating understandings about where students are in their development as readers and a clear articulation of the next, immediate goals for instruction.

[Insert Figure 3, Guided Reading Item Map, about here.]

In the same region of the item difficulty map, we see observer reports about teachers prompting for effective reading strategies as they listen to students read. The clustering of three item difficulties within 0.5 logits of one another suggests that teachers who display strategic thinking in their book introduction are also likely to manifest such behavior in response to students' reading.

Anchoring the top of the scale are a set of items that focus on what teachers do as a guided reading lesson concludes to pull together what they have observed in student reading behaviors in an effort to solidify and reinforce their students' learning. As evidenced by the item difficulties, these aspects of guided reading—both extending meaning through student talk and making explicit teaching points—are less frequently observed. This is not surprising, in that such instruction entails the moment-by-moment decision making of a teacher as she seeks to support students' development as independent readers. Only more expert teachers, who are keen observers and interpreters of student reading behaviors and capable of quick strategic response, are likely to demonstrate this.

Overall, the Guided Reading item map presents a coherent progression from basic instructional routines toward more expert practice that requires teachers to process a diverse array of information in terms of the demands of the text and observed reading behaviors of students, and then integrate and act on this evidence within the larger context of a theory of students' literacy learning and the pedagogical affordances of a comprehensive literacy instructional system. Of note here, we found no evidence of large misfit statistics associated this scale. These results are consistent with an argument that a general progression of expertise development may apply with most teachers.¹ Taken together, these analyses support a claim that this rubric details a map of instructional practices that develop as teachers acquire increased expertise in the enactment of guided reading.

¹ A table of the complete item difficulties and fit statistics appears in the appendix.

Writing Workshop. One of the first aspects for teachers to take on within this instructional activity involves organizing the classroom so there is time for individual conferences with students about their work. Not surprisingly, then, the rubric elements that focuses on this activity anchors the bottom of the Writing Workshop scale. These conferences may be readily introduced into classrooms as students are provided time to write independently and teachers begin to confer individually with students during this time. The productivity of these activities for student learning, however depends on what actually occurs in each conference as well as the rigor of the other instructional supports and activities that a teacher organizes for students while they are waiting for their individual conferences.

The next cluster of items on the difficulty map focuses on the mini-lessons conducted as part of the writing workshop. These mini-lessons provide opportunities to teach specific writing principles. The most basic aspect of the mini-lesson is the selection and “demonstration of an example of the writer’s craft, conventions or process.” This explicit teaching of core writing principles aims to support students in analyzing what they read to inform what and how they communicate in their own writing.

Moving a bit higher in the scale, we find reports that teachers “carefully check students’ understandings” of the connections of the principles articulated in the mini-lesson to their own work. For example, the teacher may ask students to discuss what they are writing with a friend while she circulates around the class to participate in these conversations.

Finally, each mini-lesson sits within a broader instructional agenda in which teachers seek to help students come to understand the processes of good writing. Relating

the mini-lessons to these broader writing principles provides multiple occasions for helping students to develop the cognitive schema necessary to scaffold their increasingly independent work as writers. It is this orchestration of the mini-lesson as a whole that constitutes powerful instruction.

[Insert Figure 4, Writing Workshop Item Map, about here.]

The most difficult items within the Writing Workshop rubric scale involve explicit teaching in the context of students' sharing of their writing. This activity provides spontaneous "teaching moments" where instruction can make the essence of good writing concrete, both for the students who are sharing and for others in the class. Such instruction requires teachers to interweave explicit points about writing principles mentioned earlier in the lesson with the work that students have been doing in class—in essence, close the loop on the entire activity. It makes demands on teachers to integrate what they may see in a specific piece of a student's writing, what they know more generally about students' writing behaviors, and more basic understandings about literacy development and pedagogical content knowledge. Finally, at the most expert level (i.e., at the top of the item difficulty map), teachers are skillful in facilitating student comments on each other's writing. This level of expertise is evidenced in classrooms where students, in talking about their own work, display an understanding of the writing process and use these ideas to scaffold their responses to other students' writing as well. Such talk does not occur spontaneously. It is modeled and facilitated by a skillful teacher.

As was true for the Guided Reading rubric, we found no evidence of item misfit in this Writing Workshop scale either. (See Appendix A for further details.) Taken overall, the item map for Writing Workshop explicates a set of instructional practices that

appear to arrange in a reasonable developmental order. The scale moves from implementing a structure of activities including conferencing, mini-lessons, and sharing to a sophisticated crafting of explicit instructional points in response to what the teachers observe as students engage in authentic writing. As with guided reading, the most demanding aspects for teachers during writing workshop are the moment-by-moment decisions that attempt to connect and integrate students' writing experiences with the core principles of the writing process. As noted earlier, this ability to quickly recognize, analyze, and act on complex patterns of information is the hallmark of expertise. This scale operationally describes the movement toward such expertise within the context of the writing workshop activity.

Word Study. This item map is displayed in Figure 5. The easiest items involve selecting an application task that students can engage in independently that has potential for advancing student learning around how words work. This teaching routine involves choosing appropriate materials to set the context for learning. Somewhat more difficult than selecting an appropriate word study task is using examples in the mini-lesson so that students clearly understand the link between the task and the word study principle that is the focus of the lesson. Of course, examples can be used simply to show students how “to do” the activity that will follow. Elaborating the word study principle as a clear and explicit part of the mini-lesson (see for example the item “Uses good examples; teachers check for understanding and helps students...”) is a more sophisticated aspect of teaching within word study, and this is reflected in its placement on the item difficulty map.

Demonstration of how the principle relates to the application task is still another dimension for teachers to attend to in the mini-lesson. The teacher orchestrates links so

that students understand their work, for example, not simply as sorting words that have the same ending pattern but as a task that helps them understand how letter-sound relationships work and how they can use this in their reading and writing.

[Insert Figure 5, Word Study Item Map, about here.]

The most expert aspects of word study are reflected in the rubric elements that inquire as to whether and how teachers use these sharing opportunities to advance student learning. Similar to writing workshop, this sharing activity provides a context to reinforce core principles about how words work. Likewise, it requires a capacity to analyze and act quickly on a complex pattern of evidence and theory about student learning. The teacher does this by both restating the word study principle, making connections to the work students have just done, and facilitating conversations in which students reflect on what they have learned. These aspects of word study instruction are represented in the two most difficult items in this scale.

Again we found no evidence here, either, of item misfit. Overall, the results for the Word Study scale rank order as we would expect given the expertise development framework articulated above.

Evidence on the Usefulness of the Rubrics for Assessing Change in Teacher Practice

The purpose of this instrumentation study was to develop a reliable and valid tool for assessing changes in teachers' comprehensive literacy instructional activities. We have conceptualized this task as a problem in expertise development. We now evaluate the evidence assembled so far as to whether the DLLT meets this goal.

An empirically grounded model for teacher expertise development within comprehensive literacy instruction. We completed the same conceptual-empirical

analysis illustrated above for each or the other three activity components in comprehensive literacy instruction—Interactive Read Alouds, Shared Reading, and Interactive Writing, and for the two holistic rubrics—General Aspects of Teaching and Teaching for Strategies. In general, the item difficulties mapped along very similar lines. In addition, we found little evidence of item misfit on any of the eight rubrics. (See Appendix A for further details.) These results support a general interpretation that these scales describe a pathway toward developing teacher expertise within a comprehensive literacy instructional framework.

In summarizing across the separate item maps for the eight rubrics that comprise the DLLT, we see a general pattern with regard to teachers' activities within this instructional system. At the most basic level, the scales reflect how teachers begin to implement these activities. The movement toward higher levels on each scale delineate a pathway that teachers traverse as they become more expert in supporting student learning through each individual activity and in orchestrating across them. Figure 6 presents an overall model that summarizes these empirical results.

[Insert Figure 6, Developmental Model of Teacher Practice, about here.]

Initially, teachers focus on the materials and routine activities necessary to implement comprehensive literacy and on the organization of their classrooms to support such student work. They subsequently move toward embedding the principles of the reading and writing processes within this structure to advance student learning. This progression occurs in mini-lessons and other aspects of instruction and increasingly becomes a guide to teachers' selection of materials and texts to engage the whole class and small groups. As teachers become more adept at demonstrating instructional

principles through examples, their observation of students becomes even more critical. It is their combined understanding of the reading and writing processes and observations of how their students engage in problem solving and thinking about texts that allows teachers to identify and create more explicit teaching points in direct response to student learning.

Effectively linking knowledge and observation represents a significant jump in teacher practice as displayed by the difficulty gap in Figure 6. Teachers operating at this level are much more strategic and targeted in their selection of the principles they introduce and which they repeatedly return to. Teachers are now also building opportunities for students to articulate and share what they are learning. This facilitation of student talk during sharing and discussion reinforces student learning and creates a self-extending quality to the activity.

Taken together, the rubrics represent a developmental pathway from novice toward more expert practice as teachers acquire the necessary skills and understandings to be able to respond effectively to where students are as learners, using the resources afforded by a comprehensive literacy framework. At the base level, the rubrics assess whether appropriate classroom structures are in place and teachers can efficiently execute basic instructional routines. These are essential building blocks for good teaching. In moving toward expertise in comprehensive literacy instruction, however, teachers must infuse these structures and routines with research-based understandings of the reading and writing processes, and with their specific knowledge of the developing literacy skills of each child. They must do this while aiming to accelerate the progress of all children toward the ultimate goal of becoming effective, independent readers and writers. This

capability to engage in such complex analysis and to act deftly in response marks the high end of the rubric scales.

Comparing results from novice versus more experienced teachers. The DLLT rubrics have been specifically designed to assess teacher development over time within the pedagogical practices of a comprehensive literacy framework. To validate fully such a claim, we need a longitudinal study that tracks the development of teachers new to comprehensive literacy and how their instruction changes over time as they are exposed to professional development and coaching in comprehensive literacy instruction. A useful first test based on cross-sectional data, is whether we observe any differences on the rubric scores for teachers relatively new to comprehensive literacy versus those who have been engaged in it for several years.

As part of the basic design for the instrumentation study, we had sampled participants according to the number of years that each teacher had been exposed to Literacy Collaborative professional development. Approximately half of the teachers were *new* to Literacy Collaborative (less than two years of professional development) and the other half more *experienced* (two or more years in the program). According to the basic program description offered by the Literacy Collaborative, it typically takes about three years of professional development for a teacher to learn to become proficient within this instructional framework. This suggests that we should see significantly different reports on the DLLT rubrics if we divide the sample into these two groups—novices with less than two years experience with comprehensive literacy instruction, and those with two year more experience with this form of instruction. Figure 7 presents a set of box

plots comparing these two groups on the eight separate measures. Note that all of the mean differences presented here are statistically significant beyond the 0.001 level.

[Insert Figure 7, Comparing Novice and Experienced Teachers on the DLLT, about here.]

In general, we found large differences between these two groups on each of the six core instructional activities that are typically found in comprehensive literacy instruction. For example, in Guided Reading and Writing Workshop, the median score for the group of teachers *experienced* with comprehensive literacy is equivalent to roughly the 75th percentile in the distribution for the teachers who are *novices*, i.e., new to comprehensive literacy. Similar size differences exist for the holistic rubric, General Aspects of Teaching.

An even larger difference appeared when we compare *novice* and *experienced* teachers on the Teaching for Strategies rubric. On this rubric, the 25th percentile *experienced* teacher was scoring at about the same level as a 75th percentile *novice* teacher. Recall that this rubric draws on evidence from the entire literacy block rather than just when a particular component is being taught, and seeks to evaluate how well teachers integrate their efforts across all elements of the framework toward strategic teaching to advance students' literacy learning. Conceptually, this is our most direct measure of expert teaching, and not surprisingly, it discriminates the most between teachers who are novices versus experienced in comprehensive literacy instruction.

Summary, Conclusions and Limitations

In developing the rubrics detailed in this paper, we drew on theory, empirical research in elementary literacy learning, and clinical expertise regarding literacy instruction and professional development to advance this instruction. Our research team

combined a diverse collegueship of expertise, including extensive experience observing teacher practice, coaching for teacher development, literacy research, psychometrics, and research methodology. Each of these fields played an essential role in building, refining, and field testing the rubrics for teacher development presented in this article.

The DLLT rubrics direct attention to teachers' instructional behaviors and language use within the context of a comprehensive literacy classroom. The extensive clinical observations that have now accumulated on this instructional approach guided us in conceptualizing the variations in practice that were likely to occur and how we might relate these to a theory of expertise development within this instructional system. As a result, we were able to design a set of rubrics where both high levels of both inter-rater and over-time reliability were relatively easy to achieve. In this latter regard, we were mindful of recent findings from Rowan, Camburn, and Correnti (2004), who reported that 15 or more observations were needed in their study to distinguish reliably among teachers in their instruction. We were able to accomplish this objective in the DLLT with substantially fewer observations.

Research Connecting Professional Development to Student Learning

The work on measuring change in teachers' literacy practice detailed in this article sits within a larger research context of efforts to examine the linkages between professional education initiatives and improvements in student learning. Any integrated research program of this sort must examine a set of interrelated questions that begins with the intentional design of a professional development program. First, "What exactly do we expect teachers to learn and be able to do and how would we know if they have acquired such capabilities?" Second, does professional education actually translate into

observable changes in classroom practice? Third, do these developing classroom practices in turn culminate in improvements in student learning? Probing this inter-related set of questions provides a basic organizing paradigm for research on the effects of professional education.

Unfortunately, in many efforts to promote changes in teaching, clear specification of the changes expected, as well as a system for their reliable and valid measurement, is often lacking. In essence, this creates a “black box” at the heart of the intended change process. Enormous energy and resources may be placed into the design and implementation of professional education without the ability to actually track the intended changes in classrooms that in turn aim to improve student learning. Absent this, we don’t know how to interpret fully the observed student outcomes and we lack empirical evidence for formative evaluation and subsequent fine-tuning of the professional education initiatives.

Advancing Literacy Practice Improvement at Scale

The data presented here were collected by reading coaches as a part of their professional activities. The results indicate that this tool can be used reliably by coaches in the context of their everyday work in schools. This is significant because the “common language” embedded in a tool such as the DLLT can be a valuable resource for promoting individual teacher learning and for organizing communities of practice around such improvement. These rubrics could be used by coaches, for example, as a guide for differentiating staff development activities among teachers. Coaches might also use these rubrics as a lens to focus their observations in classrooms, making it possible to provide more targeted individual feedback to teachers, and for identifying larger, school-wide

professional development priorities. The rubrics can also function as a useful tool for teachers to guide their own self-reflection.

Perhaps most significant, the DLLT rubric represents a distinct perspective on how we think about the changes in teaching that might result from intentionally designed professional development. Many large-scale instructional reform initiatives tend to view teacher change through a lens that stresses “fidelity of implementation.” This view directs attention to whether teachers are using the materials and activities in the ways intended by the program designers. This is, of course, not irrelevant to a comprehensive literacy program. In fact, the efficient execution of the basic routines of comprehensive literacy instruction anchors the bottom to middle categories on many elements of the DLLT. However, the deeper goal in such professional education is to develop teachers as reflective practitioners. From this view, expert teaching entails attention to where their students are as readers and writers, cognizance of the next “goals in view” for student development and the skillful organizing and execution of appropriate instruction in response. It requires a capability to engage in complex analysis that blends theories of literacy learning, critical observations of students’ actual reading and writing activities, and in-depth knowledge of how various aspect of the instructional system that teachers are working within can be best deployed to advance subsequent learning. Moreover, this analysis must occur quickly, within the moments of instruction. The capability for such complex pattern recognition and expedient action is the hallmark of expertise. The upper categories on the DLLT rubric elements that focus on the six specific comprehensive literacy activities, along with the *Teaching for Strategies* rubric, aim to assess this expert capability.

In sum, the DLLT seeks to assess teachers' development of expertise within a comprehensive literacy instructional system. We conceive of this development as beginning with the introduction and experimentation with new methods in classroom practice to achieving efficient use of routines to finally evidencing genuine expertise in the orchestration of instructional activities to advance student learning. We believe that this framework affords a useful heuristic for thinking about the problems of improving complex instruction at scale, and one that teachers will likely embrace as a valid and respectful accounting of their own continuous improvement efforts.

Nudging the Methods for Rubric Development Forward

The DLLT also introduce a new measurement model that shows promise for improving future rubric construction for charting teacher development. Clinical rubrics tend to strive for a rich description of instruction across domains, such as affective and academic, and across curriculum subjects, such as math, reading, and social studies. It can be difficult using such descriptive rubrics, however, to achieve traditional standards of reliability and validity. Individual teachers may develop unevenly on the different components that comprise these composite descriptions, thereby forcing observers to make somewhat arbitrary choices as to which category values to assign. By treating each row of the rubric as, in essence, a distinct rating scale element, we sought to reduce this difficulty. Rich composite descriptions of practice still result, but how the elements combine in an account of practice for a particular teacher at a specific time is now allowed to vary. In addition, the Rasch Rating Scale permits an analysis of the hierarchy of difficulty of these descriptive elements. This affords valuable information for assessing construct validity while simultaneously strengthening the instrument's reliability.

Limitations

The DLLT rubrics were designed to assess the development of teaching expertise within the context of comprehensive literacy instruction. The latter is widely used in schools today and as a result we anticipate that these rubrics may have broad utility. To date, however, they have only been applied within Literacy Collaborative schools, which represent just one instantiation of comprehensive literacy instruction. While the rubrics are intended for broader use, further reliability and validity assessments in other contexts are needed.

In addition, the rubrics were designed for assessing changes in teachers' practice over time, mapping from novice activity toward expertise. We have presented a first set of empirical analyses to examine the validity of such use. While the results presented here are encouraging, longitudinal studies are still needed. Specifically, we need to directly examine the link between teachers' exposure to professional education and changes in instructional practice as measured on these rubrics. We also will need to evaluate how measured changes on the DLLT relate to changes in student learning over time. If these empirical links are validated, this would support future use of these rubrics as indicators both for judging the efficacy of professional development initiatives in comprehensive literacy instruction and for predicting improvements in student learning that might follow. In our view, developing reliable and valid tools of this sort is an important step toward more ambitious instructional efforts, such as providing students with comprehensive literacy instruction at scale.

References

- Adams, M.J. (1990). *Beginning to read: thinking and learning about print*. Cambridge, MA: MIT Press.
- Anderson, R.C., Wilson, P.T., & Fielding, L.C. (1998). Growth in reading and how children spend their time outside of school. *Reading Research Quarterly*, 23(3), 285-303.
- Beck, I. L., & McKeown, M. G. (2001). Text talk: Capturing the benefits of read-aloud experience for young children. *Reading Teacher*, 55, 10–20.
- Bissex, G. (1980). *GNYS at work: A child learns to read and write*. Cambridge, MA: Harvard University Press.
- Bond, G.L., & Dykstra, R. (1967). The cooperative research program in first-grade-reading instruction. *Reading Research Quarterly*, 2(4), 5-142.
- Bransford, J.D., Brown, A.L., & Cocking, R.R. (1999). *How people learn: Brain, mind, experience and school*. Washington D.C.: National Academy Press.
- Calkins, L.M. (1994). *The art of teaching writing*. Portsmouth, NH: Heinemann.
- Clay, M.M. (1991). *Becoming literate: The construction of inner control*. Portsmouth, NH: Heinemann.
- Clay, M.M. (1998). *From different paths to common outcomes*. York, ME: Stenhouse.
- Clay, M.M. (2001). *Change over time in children's literacy development*. Portsmouth, NH: Heinemann.
- Clay, M.M. (2005). *Literacy lessons designed for individuals*. Portsmouth, NH: Heinemann.

- Crosson, A.C., Boston, M., Levison, A., Matsumura, L.C., Resnick, L.B., & Junker, B.W. (2006). Beyond summative evaluation: The Instructional Quality Assessment as a professional development tool (CSE Technical Report 691). University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing, Center for the Study of Evaluation.
- Dahl, K., Scharer, P., Lawson, L., & Grogan, P. (1999). Phonics teaching and student achievement in whole language first grades. *Reading Research Quarterly*, 34(3), 312-341.
- DeStefano, J. (1978). *Language, the learner and the school*. New York: John Wiley and Sons.
- Dewey, J. (1997). *Experience and Education*. NY: Simon and Schuster.
- Doherty, R.W., Hilberg, R.S., Epaloose, G., & Tharp, R.G. (2002). Standards performance continuum: Development and validation of a measure of effective pedagogy. *Journal of Educational Research*, 26, 78-89.
- Edmunds, M. & Briggs, K. (2003). The instructional content emphasis instrument: Observations of reading instruction. In S. Vaughn & K.L. Briggs, *Reading in the classroom: Systems for the observation of teaching and learning* (pp. 31-52). Baltimore: Paul H. Brooks.
- Ehri, L.C. (1991). Development of the ability to read words. In R. Barr, M.L. Kamil, P. Mosenthal, and P.D. Pearson (Eds.), *Handbook of reading research*, (Vol. 2, pp. 383-417). White Plains, NY: Longman.

- Ehri, L.C. (1998). Grapheme-phoneme knowledge is necessary for learning to read words in English. In *Word recognition in beginning literacy*, J. Metsala and L.C. Ehri, (Eds.), Hillsdale, NJ: Lawrence Erlbaum.
- Elkonin, D.B. (1971). The development of speech. In A.V. Zaporozhets & D.B. Elkonin (Eds), *The psychology of preschool children* (pp. 111-185). Cambridge, MA: MIT Press.
- Elkonin, D.B. (1973). Reading in the USSR. In J. Downing (Ed.), *Comparative Reading* (pp. 551-580). NY: Macmillan.
- Ericsson, K.A., Krampe, R., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*(3), 363-406.
- Ericsson, K. A. (1996). The acquisition of expert performance: An introduction to some of the issues. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games* (pp. 1-50). Mahwah, NJ: Erlbaum.
- Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. Feltovich, and R. R. Hoffman, R. R. (Eds.). *Cambridge handbook of expertise and expert performance* (pp. 685-706). Cambridge, UK: Cambridge University Press.
- Estrada, P. (2004). Patterns of language arts instructional activity and excellence in first- and fourth-grade culturally and linguistically diverse classrooms. In H.C. Waxman, R.G. Tharp, & R.S. Hilberg (Eds.), *Observational research in U.S. classrooms: New approaches for understanding cultural and linguistic diversity* pp. 122-143. Cambridge, U.K.: Cambridge University Press.

- Ferriero, E., & Teberosky, A. (1982). *Literacy before schooling*. Portsmouth, NH: Heinemann.
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York: Wiley.
- Flybjerg, B. (2001). *Making social science matter: Why social inquiry fails and how it can succeed again* (S. Sampson, Trans.). Cambridge, UK: Cambridge University Press.
- Foorman, B.R. & Schnatschneider, C. (2003). Measurement of teaching practices during reading/language arts instruction and its relationship to student achievement. In S. Vaughn & K.L. Briggs, *Reading in the classroom: Systems for the observation of teaching and learning* (pp. 1-30). Baltimore: Paul H. Brooks.
- Fountas, I. C., & Pinnell, G. S. (1996). *Guided reading: Good first teaching for all children*. Portsmouth, NH: Heinemann.
- Fountas, I. C., & Pinnell, G. S. (2006). *Teaching for comprehending and fluency: Thinking, talking, and writing about reading, K-8*. Portsmouth, NH: Heinemann.
- Gersten, R., Baker, S. K., Haager, D., & Graves, A. W. (2005). Exploring the role of teacher quality in predicting reading outcomes for first-grade English learners: An observational study. *Remedial and Special Education, 26*, 197-206.
- Glaser, R., & Baxter, G.P. (2002). Cognition and construct validity: Evidence for the nature of cognitive performance in assessing student performance. In H. Braun, D. Jackson, & D. Wiley (Eds). *The Role of Constructs and Psychological and Educational Measurement* (pp. 179-191). Mahwah, N J: Erlbaum.

- Goswami, U., & Bryant, P. (1990) *Phonological skills and learning to read*. Hillsdale, NJ: Lawrence Erlbaum.
- Graves, D.H. (1983, 1994). *Writing: teachers and children at work*. Portsmouth, NH: Heinemann.
- Graves, A. W., Gersten, R., & Haager, D. (2004). Literacy instruction in multiple-language first-grade classrooms: Linking student outcomes to observed instructional practice. *Learning Disabilities Research & Practice, 19*, 262–272.
- Greenwood, C.R., Abbott, M., & Tapia, Y. (2003). Ecobehavioral strategies: Observing, measuring, and analyzing behavior and reading interventions. In S. Vaughn & K.L. Briggs, *Reading in the classroom: Systems for the observation of teaching and learning* (pp. 53-82). Baltimore: Paul H. Brooks.
- Haager, D., Gersten, R., Baker, S., & Graves, A. (2003). The English language learner observation instrument for beginning readers. In S. Vaughn & K. Briggs (Eds.), *Reading in the classroom: Systems for the observation of teaching and learning*. Baltimore, MD: Brookes.
- Johnston, P. (1997). *Knowing literacy: Constructive literacy assessment*. York, ME: Stenhouse Publishers.
- Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*, 437-447.
- Juel, C. & Minden-Cupp, C. (2000). Learning to read words: Linguistic units and instructional strategies. *Reading Research Quarterly, 35*, 458-492.
- Kamil, C., & Manning M. (2002). Phonemic awareness and beginning reading and writing, *Journal of Research in childhood Education, 17*, 1: 38-47.

- Landis, J. R., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Liberman, I. Y. and Shankweiler, D. (1985) Phonology and the problems of learning to read and write. In Topical Issue (I. Y. Liberman, Guest Editor), *Remedial and Special Education*, 6(6), 8-17.
- Lindfors, J.W. (1987). *Children's language and learning* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Lindfors, J.W. (1999). *Children's inquiry: Using language to make sense of the world*. Urbana, IL: National Council of Teachers of English.
- Lundberg, I., Frost, J., & Peterson, O. P. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly*, 23, 264–84.
- Lyons, C. A., & Pinnell, G. S. (2001). *Systems for change in literacy education: A guide to professional development*. Portsmouth, NH: Heinemann.
- Lyons, C.A. (2003). *Teaching struggling readers: How to use brain-based research to maximize learning*. Portsmouth, NH: Heinemann.
- McCarrier, M. C., Pinnell, G. S., & Fountas, I. C. (2000). *Interactive writing: How language and literacy come together*. Portsmouth, NH: Heinemann.
- Meek, M. (1988). *How texts teach what readers learn*. South Woodchester, Stroud: The Thimble Press.

- National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. Washington, DC: U.S. Department of Health and Human Services, NIH Pub. No 00-4754.
- Neufeld, B., & Roper, D. (2003). *Coaching: A Strategy for Developing Instructional Capacity, Promises and Practicalities*. Washington, DC: Aspen Institute.
- Pearson, P. D., & Fielding, L. (1991). Comprehension Instruction. In R. Barr, M. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (vol. 2, pp. 815–60). New York: Longman.
- Pinker, S. (2000). *Words and rules: The ingredients of language*. London: Phoenix.
- Pinnell, G. S., Pikulski, J. J., Wixson, K. K., Campbell, J. R., Gough, R. B., & Beatty, A. S. (1995). *Listening to children read aloud: Data from NAWP's Integrated Reading Performance record at grade 4* (Report No. 23-FR-04). Prepared by Educational Testing Service under contract with the National Center for Education Statistics, Office of Educational research and Improvement, U.S. Department of Education.
- Pinnell, G. S., & Fountas, I. C. (1998). *Word matters: Teaching phonics and spelling in the reading/writing classroom*. Portsmouth, NH: Heinemann.
- Pressley, M. (1998). *Reading instruction that works: The case for balanced teaching*. New York: Guilford Press.
- Ray, K. W. (2004). *About the authors*. Portsmouth, NH: Heinemann.

- Read, C. (1971). Pre-school children's knowledge of English phonology. *Harvard Educational Review* 41: 1-34.
- Resnick, L., Matsumura, L., & Junker, B. (2006). Measuring reading comprehension and mathematics instruction in urban middle schools: A pilot study of the instructional quality assessment (CSE Technical Report 681). University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing, Center for the Study of Evaluation.
- Rodgers, E., & Hung, C. (2006). Understanding the work of coaching: A lens for viewing classroom practice. Paper presented at the American Education Research Association, San Francisco.
- Rodgers, E. (2007). Write now! Don't wait to teach struggling readers about writing. In P.L. Scharer & G.S. Pinnell (Eds.) *Guiding K-3 writer to independence: The new essentials* (pp.183-198). Columbus, OH: The Literacy Collaborative at The Ohio State University.
- Rogoff, B. (1997). *Apprenticeship in thinking: Cognitive development in social context*. NY: Oxford.
- Rowan, B., Camburn, E., Correnti, R. (2004). Using teacher logs to measure the enacted curriculum: A study of literacy teaching in third-grade classrooms. *The Elementary School Journal*, 105, 75-101.
- Ruddell, R.B., Ruddell, M.R., & Singer, H. (Eds.). (1994). *Theoretical models and processes of reading* (4th edition). Newark, DE: International Reading Association.

- Rumelhart, D. (1994) Toward an interactive model of reading. In R.B. Ruddell, M. R. Ruddell and H. Singer (Eds.), *Theoretical models and processes of reading* (4th ed, pp. 864-894). Newark, DE: International Reading Association.
- Russell, T., & Munby, H. (1991). Reframing : The Role of experience in developing teachers' professional knowledge. In D.A. Schon (Ed). *The reflective turn: Case studies in and on educational practice* (pp. 164-187). New York: Teachers College Press
- Schon, D. (1993). *The reflective practitioner*. New York: Basic Books.
- Short, K.G., Harste, J.C., & Burke, C. (1996). *Creating classrooms for authors and inquirers* (2nd ed.). Portsmouth, NH: Heinemann.
- Singer, H. (1994). The substrata-factor theory of reading. In R.B. Ruddell, M. R. Ruddell and H. Singer (Eds.), *Theoretical models and processes of reading* (4th ed, pp. 894-927). Newark, DE: International Reading Association.
- Snow, C., Burns, M., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Tharp, R.G., & Gallimore, R. (1988) *Rousing minds to life: Teaching learning and schooling in social context*. Cambridge, UK: Cambridge University Press.
- van Es, E.A., & Sherin, M.G. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Technology and Teacher Education*, 10(4), 571-596.
- Vellutino, F. R., & Scanlon, D. B. (1987). Phonological coding, phonological awareness, and reading ability: Evidence from a longitudinal and experimental study. *Merrill Palmer Quarterly*, 33, 321–63.

- Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wolf, M.K, Crosson, A.C., & Resnick, L.B. (2006). *Accountable talk in reading comprehension instruction (CSE Technical Report 670)*. University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing, Center for the Study of Evaluation.
- Wood, D., Bruner, J.S., & Ross. G.. (1976). The role of tutoring in problem-solving. *Journal of Child Psychology and Child Psychiatry, 17, 2: 89-100.*
- Wright, B. D., & Master, G. N. (1982). *Rating scale analysis*. Chicago: Mesa Press.

Figure 1

Eight Curriculum Areas for Rubric Development	
<i>Rubrics Related to Specific Instructional Contexts</i>	
1. Interactive Read Aloud	Usually whole class instruction.
2. Shared Reading	Whole class or small group instruction.
3. Guided Reading	Small heterogeneous groups.
4. Interactive Writing	Whole class or small group instruction.
5. Writing Workshop	Whole class lessons and individual conferences
6. Word Study	Usually whole class instruction, followed by individual application.
<i>Rubrics Used Across Instructional Contexts</i>	
7. General Aspects of Teaching	
8. Teaching for Strategies across Contexts	

Figure 2

Guided Reading — Group 1		Time began: _____	Time ended: _____	<input type="checkbox"/> <i>Element not present during the observation</i>
<i>Text Selection:</i> The teacher:				
1 __ Selects a text that is not the appropriate level for the group.	2 __ Selects a text that is the appropriate level for the group, but provides few opportunities for students to learn.	3 __ Selects a text that is the appropriate level for the group and provides some opportunities for students to learn.	4 __ Selects a text that is the appropriate level and is very well matched to the group and provides many opportunities to learn.	
<i>Text Introduction:</i> The teacher:				
1 __ Provides for some introductory activities may be present, but does not attend to the central elements of an introduction (meaning of whole text, language, aspects of print)	2 __ Provides an introduction that includes some or even all elements (meaning of whole text, language, aspects of print), but is fragmented and not cohesive.	3 __ Provides an introduction that includes some or all elements (meaning of whole text, language, aspects of print), but is somewhat uneven.	4 __ Provides an introduction that includes some or all elements (meaning of whole text, language, aspects of print) in a highly integrated, engaging, and cohesive way.	
__ Does not engage children with the text or in interaction with the teacher or other students.	__ May engage children in some conversation, but talk is unfocused and does not help them engage with meaning of the text.	__ Engages children in conversation; some of the talk helps them engage with the meaning of text.	__ Engages students in a conversation that brings them into the text and supports thinking about the meaning of the text.	
<i>During Reading:</i> The teacher: <input type="checkbox"/> If teacher has appropriate reasons for simply listening to oral reading or letting children read on their own, check this box and record no rating for this row.				
1 __ Either does not sample oral reading or interrupts too much with interactions that take the reader “off track.”	2 __ Samples oral reading; interactions give children “clues” for guessing or tells words, but provides little help in engaging in effective reading behaviors.	3 __ Samples oral reading and provides some demonstrations and sometimes prompts for (as needed) effective reading behaviors.	4 __ Samples oral reading and demonstrates, reinforces, and consistently prompts (as needed) for effective reading behaviors and problem solving actions.	

After Reading: The teacher:

1 ___ Does not engage children in discussion of the meaning of the text.	2 ___ Engages children in discussion after reading but talk is unfocused or sometimes off topic.	3 ___ Engages in some discussion of the meaning of the text. Students make comments that indicate they are thinking about the meaning of the text.	4 ___ Engages children in a rich discussion of the meaning of the text that is evident in students' comments about their thinking.
___ Makes no teaching points, even though there were opportunities to do so.	___ Makes teaching points, but they do not help students to engage in effective processing of texts.	___ Makes teaching points, but not all of the teaching points help students engage in effective processing of text.	___ Makes superbly chosen, specific teaching points that help students engage in effective processing of text.

Word Work: The teacher:
the observation.

Optional: Check and do not record rating if not present during

1 ___ Shows something about words, but the work is either too easy or too hard for students and may interfere with learning. Word work may involve teaching words "from the book."	2 ___ Shows something about words, but the teaching is not specific and clear, and there is no evidence that students understand the task.	3 ___ Shows children something about words. Students participate and perform the task with some understanding.	4 ___ Shows children something explicit and strategic about how words work. Students are engaged and there is evidence that they are learning more about word solving.
---	---	---	---

Table 1. Inter-Rater Agreement for the DLLT Rubrics

	<u>Read Aloud</u>		<u>Shared Reading</u>		<u>Guided Reading</u>	
	<i>Obs 1</i>	<i>Obs 5</i>	<i>Obs 1</i>	<i>Obs 5</i>	<i>Obs 1</i>	<i>Obs 5</i>
Exact Agreement	53%	74%	36%	59%	58%	71%
Adjacent Agreement	90%	98%	84%	95%	92%	92%

	<u>Interactive Writing</u>		<u>Writing Workshop</u>		<u>Word Study</u>	
	<i>Obs 1</i>	<i>Obs 5</i>	<i>Obs 1</i>	<i>Obs 5</i>	<i>Obs 1</i>	<i>Obs 5</i>
Exact Agreement	53%	53%	63%	69%	64%	71%
Adjacent Agreement	87%	93%	90%	93%	91%	92%

	<u>General Aspect of Teaching</u>		<u>Teaching for Strategies</u>	
	<i>Obs 1</i>	<i>Obs 5</i>	<i>Obs 1</i>	<i>Obs 5</i>
Exact Agreement	49%	58%	42%	57%
Adjacent agreement	94%	95%	88%	92%

Table 2. Weighted Kappa Statistics for Each DLLT Rubric Element (time point 5)

<i>Read Aloud</i>		<i>Word Study</i>		<i>General Aspects of Teaching</i>	
RA1	0.77	WS1	0.84	GAT1	0.62
RA2	0.76	WS2	0.67	GAT2	0.74
RA3	0.96	WS3	0.51	GAT3	0.45
<i>Shared Reading</i>		WS4	0.62	GAT4	0.51
SR1	0.64	WS5	0.54	GAT5	0.50
SR2	0.71	WS6	0.85	GAT6	0.72
SR3	0.51	WS7	0.74	GAT7	0.53
<i>Interactive Writing</i>		<i>Guided Reading</i>		GAT8	0.66
IW1	0.58	GR1	0.65	GAT9	0.58
IW2	0.40	GR2	0.59	<i>Teaching for Strategies</i>	
IW3	0.92	GR3	0.75	TS1	0.62
IW4	0.61	GR4	0.68	TS2	0.52
IW5	0.42	GR5	0.78	TS3	0.37
<i>Writing Workshop</i>		GR6	0.68	TS4	0.76
WW1	0.83	GR7	0.74	TS5	0.56
WW2	0.62			TS6	0.56
WW3	0.43			TS7	0.58
WW4	0.73			TS8	0.36
WW5	0.86			TS9	0.72
WW6	0.82			TS10	0.58
WW7	0.78				

Table 3. Internal Reliability of the DLLT Rubrics

Rubric	Person Reliability	Number of Elements
Read Aloud (n=234)*	0.63	3
Shared Reading (n=120)	0.63	3
Guided Reading (n=218)	0.79	7
Interactive Writing (n=104)	0.75	5
Writing Workshop (n=213)	0.80	7
Word Study (n=148)	0.83	7
General Aspects of Teaching (n=325)	0.90	9
Teaching for Strategies (n=323)	0.91	10

*Note: The number of observations varies for each aspect of the framework for several reasons: 1) Some teachers were not using all the aspects of the framework in their classrooms; 2) Some aspects of the framework are targeted to early grades (K, 1) and so are not observed in the other grades; 3) On some occasions, observations were interrupted by the school schedule, and the entire literacy block was reduced for that day.

Table 4. Correlations Among Measures for Paired Observations at Time Point 5

Read Aloud	0.91
Shared Reading	0.78
Guided Reading	0.85
Interactive Writing	0.88
Writing Workshop	0.85
Word Study	0.86
Aspects of Teaching	0.80
Teaching for Strategies	0.81

Figure 3

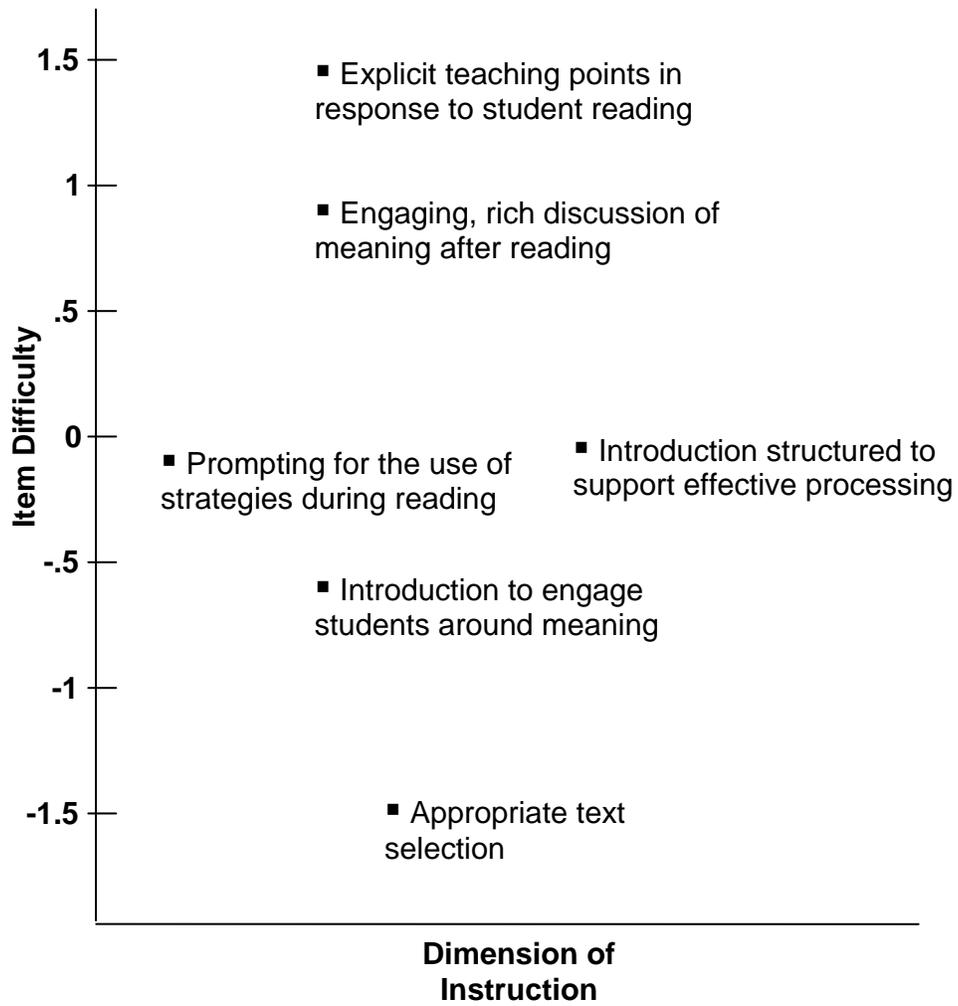


Figure 4

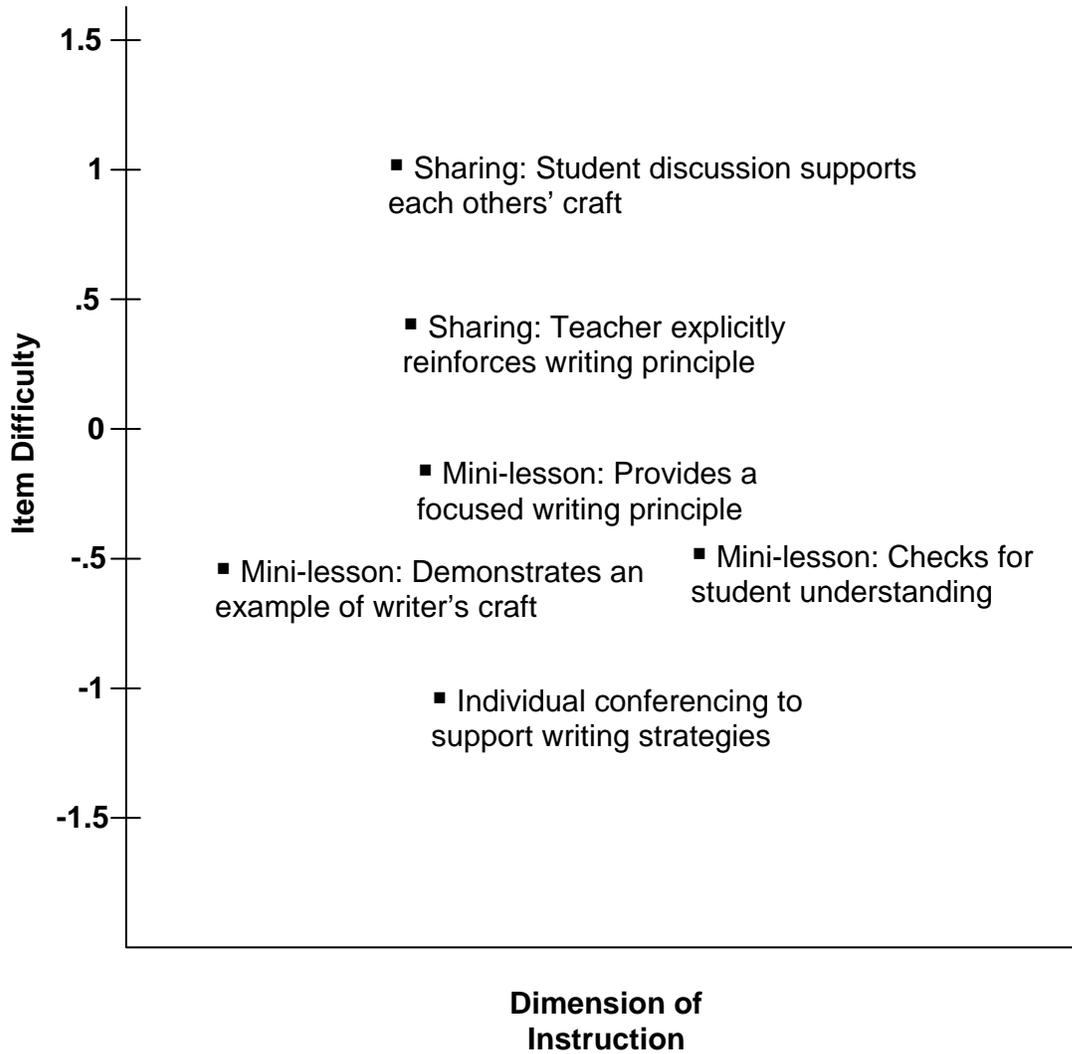


Figure 5

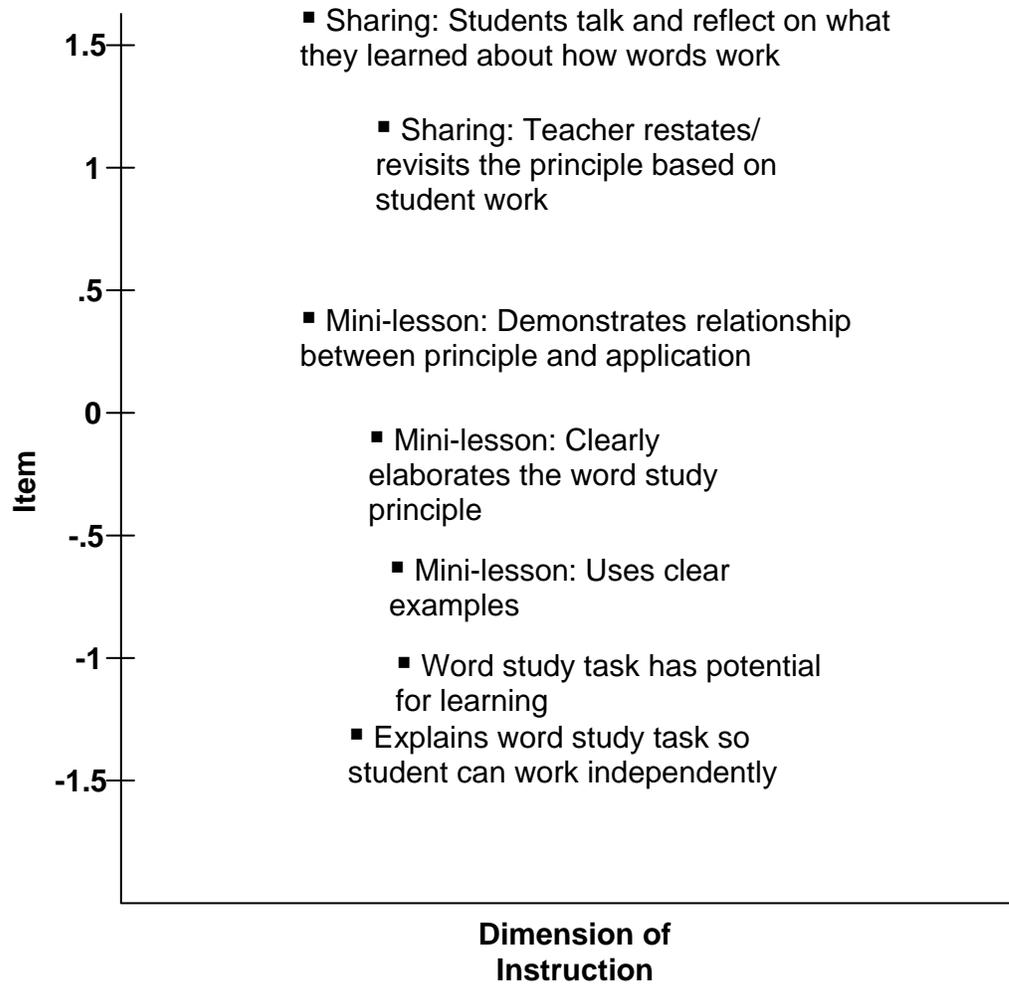
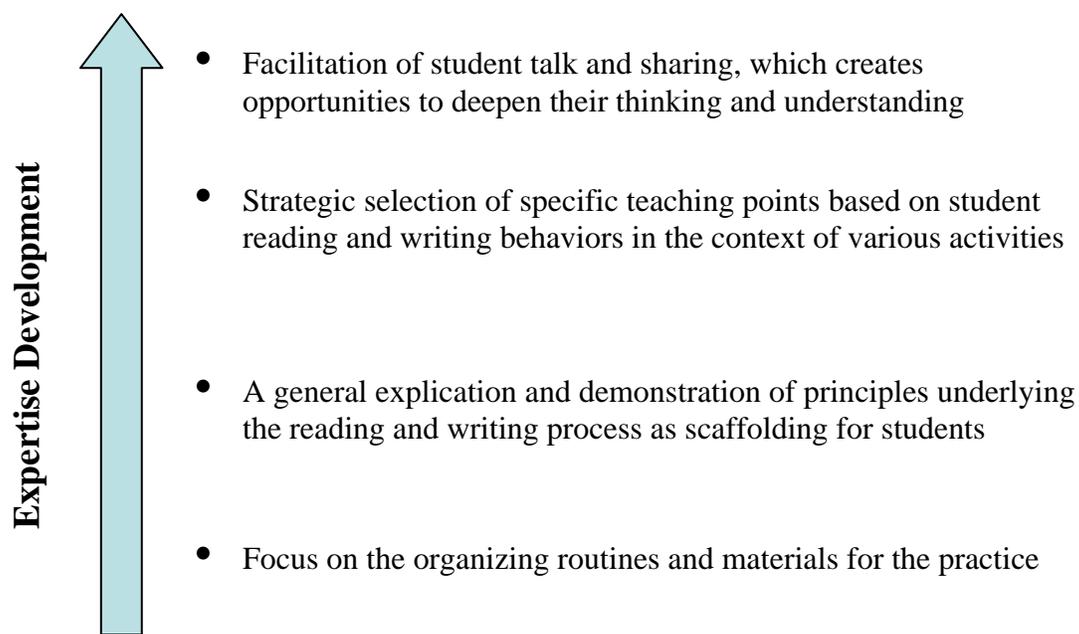


Figure 6



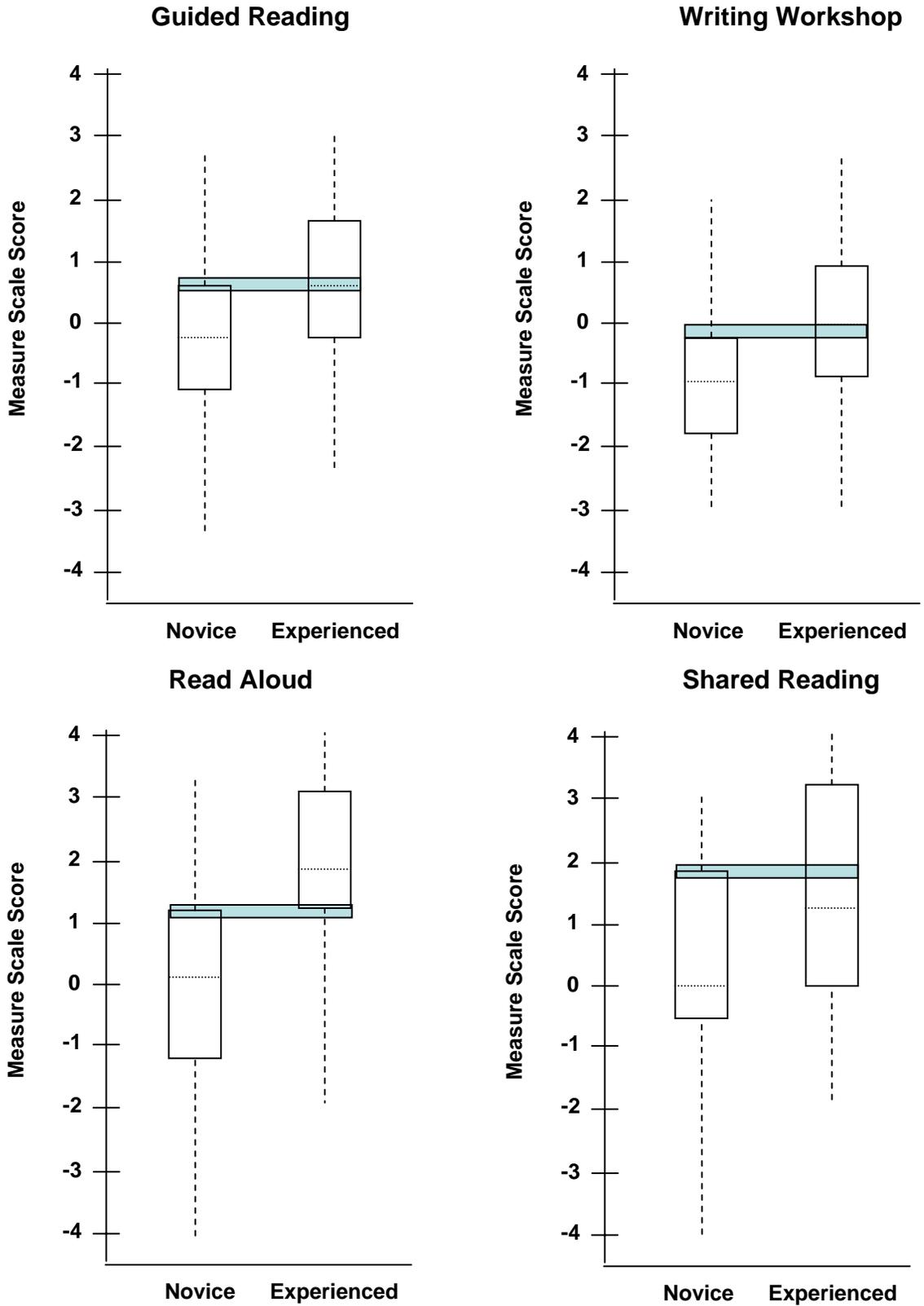
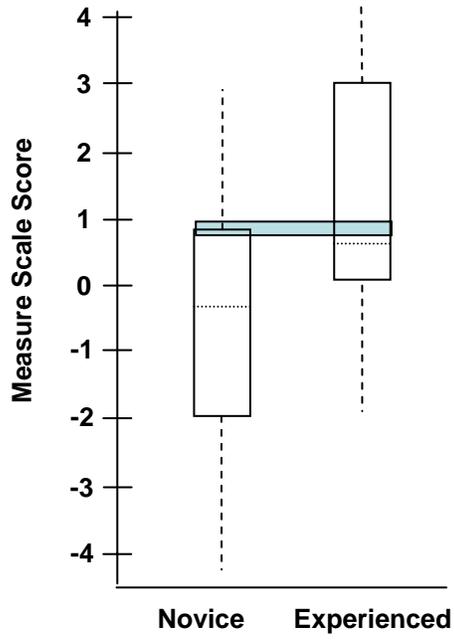
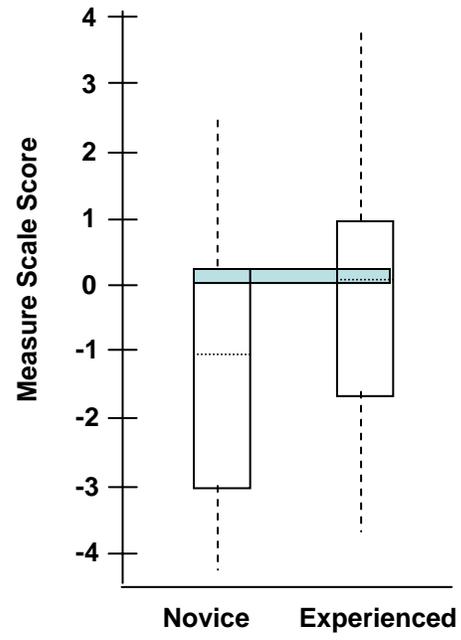


Figure 7

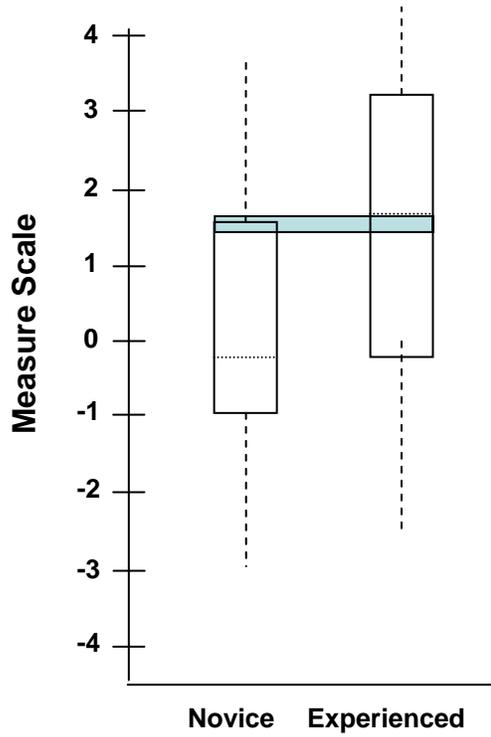
Interactive Writing



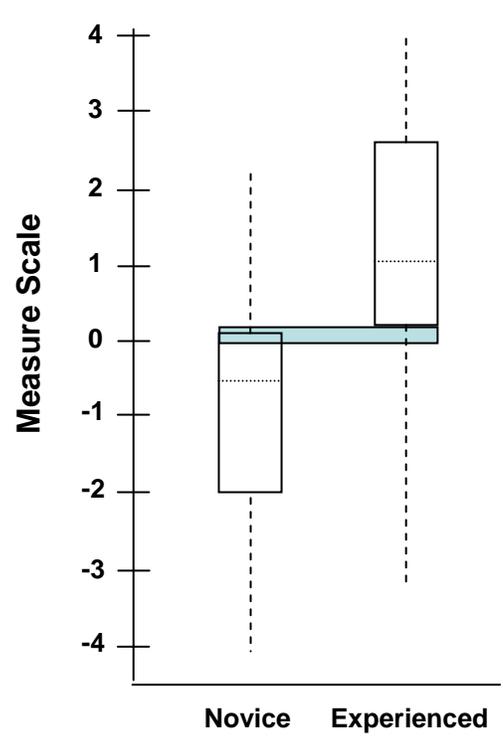
Word Study



Aspects of Teaching



Teaching for Strategies



Appendix A

Part 1—Measurement Statistics

Read Aloud

Person Reliability — 0.63

	Category 4	Item Difficulty	Infit Statistic
RA Row 3	Discussion after reading efficiently builds on overall meaning and extends students' thinking about the text.	0.71	0.95
RA Row 2	Teacher reads aloud and invites interaction; pauses add to the read-aloud session; almost all pauses are very well timed and result in good discussion during reading.	-0.19	0.94
RA Row 1	Teacher engages attention of the students prior to reading with brief comments or questions; prepares students for active listening and response.	-0.52	1.07

Shared Reading

Person Reliability — 0.63

	Category 4	Item Difficulty	Infit Statistic
SR Row 3	Makes appropriate teaching points that extend children's understanding of the reading process. Almost all are clear, specific, and well timed.	1.11	0.98
SR Row 2	Teacher engages almost all children in active shared reading of the text.	-0.33	1.05
SR Row 1	Text is appropriate (language, print, layout, interest) for the age level and the experience of students; text has many learning opportunities.	-0.78	0.92

Guided Reading

Person Reliability — 0.79

	Category 4	Item Difficulty	Infit Statistic
GR Row 6	Makes superbly chosen, specific teaching points that help students engage in effective processing of text.	1.23	1.12
GR Row 5	Engages children in a rich discussion of the meaning of the text that is evident in students' comments about their thinking.	0.83	1.04
GR Row 7	Optional: Shows children something explicit and strategic about how words work. Students are engaged, and there is evidence that they are learning more about word solving.	0.36	0.98
GR Row 4	Samples oral reading and demonstrates, reinforces, and consistently prompts (as needed) for effective reading behaviors and problem solving actions.	-0.08	0.75
GR Row 2	Provides an introduction that includes some or all elements (meaning of whole text, language, aspects of print) in a highly integrated, engaging, and cohesive way.	-0.09	0.86
GR Row 3	Engages students in a conversation that brings them into the text and supports thinking about the meaning of the text.	-0.64	0.99
GR Row 1	Selects a text that is the appropriate level and is very well matched to the group and provides many opportunities to learn.	-1.61	1.16

Interactive Writing

Person Reliability — 0.75

	Category 4	Item Difficulty	Infit Statistic
IW Row 5	Selects a few teaching points that offer new learning without unnecessarily involving children doing what they already know well; children contribute to the writing in ways that have high instructional value.	0.71	0.89
IW Row 3	The teacher engages children in a lively negotiation; options are offered by several children; serious consideration is given to word choice and sequence.	0.34	0.87
IW Row 1	Engages children in interesting experiences and a rich and purposeful discussion before writing;	0.21	0.88
IW Row 4	Keeps the writing moving along at a good pace with superbly selected teaching points; children make contributions that have high instructional value.	0.12	1.02
IW Row 2	Makes writing a highly purposeful and connected activity.	-1.38	1.27

Writing Workshop

Person Reliability — 0.80

	Category 4	Item Difficulty	Infit Statistic
WW Row 5	There is consistent evidence of note taking and continuity from previous conferences.	2.12	2.07
WW Row 6	Provides time for students to share their writing; students comment specifically about other students' writing and show understanding of strategies or craft of writing.	0.54	1.07
WW Row 7	Makes explicit and helpful comments about writing shared by students and clearly reinforces the principle and strategies for writing.	-0.05	1.03
WW Row 1	Provides a mini-lesson that is clearly stated and focused on a writing principle.	-0.50	0.77
WW Row 2	Provides a clear and explicit demonstration or example of what students need to learn as writers (craft, conventions, or process).	-0.58	0.73
WW Row 3	Checks on understanding of principle or application and elicits comments from students that are evidence of understanding.	-0.58	0.83
WW Row 4	Teacher consistently confers with students; interactions prompt for skillful use of strategies or development of writing craft. Most conferences are focused on helping students learn about the <i>writing process</i> .	-0.89	0.97

Word Study

Person Reliability — 0.83

	Category 4	Item Difficulty	Infit Statistic
WS Row 7	Students actively participate in sharing, comment on their work, and show evidence of learning the principle.	1.83	1.20
WS Row 6	Teacher clearly restates the principle and reinforces learning, through examples of students' work.	1.52	1.15
WS Row 4	Provides an application task that is appropriate and has strong potential for helping students develop greater understanding of the principle.	0.79	0.93
WS Row 3	Clearly demonstrates and explains the application task and explicitly relates it to the principle.	-0.06	0.65
WS Row 1	Provides a mini-lesson with a clearly and explicitly stated principle <i>or</i> asks children to derive the principle from examples and to state the principle clearly and explicitly.	-0.38	0.86
WS Row 2	Uses good examples; teacher checks for understanding and helps students understand how the principle is related to reading and writing.	-0.98	0.70
WS Row 5	Explains the application task in a way that enables almost all students to perform the task independently.	-1.15	1.12

General Aspects of Teaching

Person Reliability — 0.90

	Category 4	Item Difficulty	Infit Statistic
GAT Row 8	<i>Quality of Interactions:</i> Student discussion builds on the comments of other students; students provide evidence to support their ideas based on the text.	1.50	0.94
GAT Row 4	<i>Student Engagement:</i> Most students are on task almost all of the time; there is a very high level of engagement and purposeful activity.	0.42	0.78
GAT Row 5	<i>Student Engagement:</i> Transitions are orderly and efficient.	0.39	1.00
GAT Row 7	<i>Quality of Interactions:</i> Students have many opportunities to talk to, and learn from, each other.	0.03	1.17
GAT Row 9	<i>Sense of Community:</i> The teacher helps students to take high degree of responsibility for their own behavior and learning and to show respect for the learning of others. (E.g., students know routines and why they use them; they help and treat others with respect.)	0.01	0.87
GAT Row 3	<i>Classroom Materials and Organization:</i> Student/teacher generated charts are accessible, relevant and routinely used by teacher and students to guide learning.	-0.02	1.20
GAT Row 2	<i>Classroom Materials and Organization:</i> Organization works for maximum student independence; use and placement of materials in the classroom is obvious.	-0.59	0.85
GAT Row 1	<i>Classroom Materials and Organization:</i> Materials are highly organized for efficient use by the teacher and students.	-0.71	1.00
GAT Row 6	<i>Quality of Interactions:</i> The teacher consistently listens and responds to students.	-1.03	0.98

Teaching for Strategies

Person Reliability — 0.91

	Category 4	Item Difficulty	Infit Statistic
TS Row 9	<i>Teaching for Fluency and Phrasing:</i> Across reading instruction, teacher demonstrates, attends to, reinforces, and prompts for reading that is fluent, phrased, and well stressed.	0.96	1.30
TS Row 5	<i>Teaching for Inferential and Analytic Thinking:</i> Teacher models his/her own inference and analysis about texts and supports students in using these strategies; explicitly demonstrates how readers can apply these strategies.	0.79	.93
TS Row 6	<i>Teaching for Word Solving:</i> The teacher consistently helps students learn and apply a wide range of flexible and highly effective word solving strategies—recognize words, use word parts, derive their meaning from context.	0.40	0.97
TS Row 3	<i>Teaching for Inferential and Analytic Thinking:</i> Teacher consistently asks questions that extend the meaning of the text and often bring out multiple perspectives; consistently prompts student for evidence from the text that elaborates and supports their answers.	0.21	0.80
TS Row 7	<i>Teaching for Word Solving:</i> The teacher supports students in learning and expanding their understanding of word meanings in multiple contexts. Words are talked about and revisited often.	0.15	0.86
TS Row 10	<i>Teaching for Fluency and Phrasing:</i> Teacher assists children when there is evidence of dysfluent reading in various contexts; teacher avoids interrupting fluent reading.	0.03	1.51
TS Row 8	<i>Teaching for Word Solving:</i> The teacher actively provides instruction on phonemic awareness and/or letter-sound relationships and students have ample opportunity to practice and apply these skills in multiple contexts.	-0.20	1.21
TS Row 2	<i>Teaching for Literal Thinking:</i> Teacher helps students learn how to search for and use information that is in the text.	-0.38	0.91
TS Row 4	<i>Teaching for Inferential and Analytic Thinking:</i> Teacher helps students access and use relevant prior knowledge to understand meaning beyond the literal text; teacher helps students synthesize new knowledge in support of understanding the text.	-0.83	0.88
TS Row 1	<i>Teaching for Literal Thinking:</i> The teacher helps students notice specific information contained in both fiction and factual texts that is vital to the literal understanding of the text and helps them to have an overall understanding.	-1.12	0.76

Figure Captions

Figure 1. Eight curriculum areas for rubric development

Figure 2. Guided reading rubric

Figure 3. Guided reading item map

Figure 4. Writing workshop item map

Figure 5. Word study item map

Figure 6. Developmental model of teacher practice

Figure 7. Comparing novice and more experienced teachers on the DLLT