

A Framework for Analysis of Case Studies of Reading Lessons

Joanne F. Carlisle¹, Ben Kelcey², Cheryl Rosaen³, Geoffrey Phelps⁴ & Anita Vereb¹

¹School of Education, University of Michigan, Ann Arbor, Michigan

²College of Education, Criminal Justice, and Human Services, University of Cincinnati, Cincinnati, Ohio

³College of Education, Michigan State University, Lansing, Michigan

⁴Educational Testing Services, Princeton, New Jersey

Correspondence: Joanne F Carlisle, School of Education, University of Michigan, Ann Arbor, Michigan. E-mail: jfcarl@umich.edu

Received: June 26, 2013 Accepted: July 16, 2013 Online Published: July 29, 2013

doi:10.11114/jets.v1i2.161

URL: <http://dx.doi.org/10.11114/jets.v1i2.161>

Abstract

This paper focuses on the development and study of a framework to provide direction and guidance for practicing teachers in using a web-based case studies program for professional development in early reading; the program is called Case Studies Reading Lessons (CSRL). The framework directs and guides teachers' analysis of reading instruction by focusing their attention to three critical dimensions of the process of teaching; in theory, analysis of a wide variety of reading lessons, using this framework, should contribute to teachers' expertise. We report on a study of the Thinking Questions, which scaffold teachers' analysis of the reading lessons, to determine the extent to which their responses meet theoretical expectations. Results suggest that teachers' ratings of lessons tap their overall expertise in analysis of reading instruction, such that the three dimensions and features that represent these do not constitute separate factors. However, performance on the Thinking Questions differentiated more and less experienced teachers. As expected, less experienced teachers wrote longer and more specific comments about the instruction than more experienced teachers, who tended to highlight effective principles. The results suggest that an analytic framework of the kind used in CSRL holds promise as an effective component of a case-based professional development program. However, they also point to the need for further study of the framework and its influence on teachers' own teaching practices.

Keywords: case studies, professional development, practicing teachers, reading

1. A Framework for Analysis of Case Studies of Reading Lessons

In recent years, there have been repeated calls for improvement in the nature of opportunities for practicing teachers to develop their professional knowledge and expertise (e.g., Borko, 2004; Putnam & Borko, 2000). Case studies offer one promising approach to engage teachers in analysis of teaching and effective practice. When they offer teachers opportunities to grapple with problems of teaching they encounter in their own work, studying cases may lead them to improve their own teaching. Advances in video and web technologies have made case study programs accessible and appealing to teachers (e.g., Borko, Whitcomb, & Liston, 2009; Harrington, 1995).

Case studies have been widely used as a resource for learning in teacher preparation programs but rarely as a form of professional development for practicing teachers. In preservice settings, course instructors set the purpose and provide direction and guidance (Merseth, 1996; Schrader, Leu, Kinzer, et al., 2003), but this form of support for teachers' learning is not feasible for practicing teachers. Several web-based programs have incorporated systems to guide analysis of instruction usually in mathematics (Santagata, Zannoni, & Stigler, 2007; van Es & Sherin, 2002, 2006). However, there aren't well-established guidelines for designing an effective framework for analyzing instruction; an important question is how to encourage teachers working in a web-based professional development environment to analyze instruction deeply and systematically in order to learn from their study of cases. With this challenge in mind, we developed a professional development program called Case Studies of Reading Lessons (CSRL) that provides a framework for analysis of elementary reading lessons; central to this framework is the Thinking Questions. The study reported herein was designed to investigate whether the Thinking Questions framework serves as a valid and reliable way for teachers to analyze the quality of reading instruction, reflecting the theoretical framework on which this program was based.

1.1 Components of Case Studies for Practicing Teachers

A case study is “a descriptive research document based on a real-life situation or event” (Merseeth, 1996). Case studies have three major components—the case, information about the context and background of the case, and directions and guidance for studying the case. With regard to the nature of cases, Merseeth (1996) and others (e.g., Putnam & Borko, 2000) argued that the case should focus on actual lessons from real classrooms so that teachers can reflect on and learn from authentic teaching. With regard to context, teachers need background information so that it is possible to understand, interpret, and evaluate lesson events and dynamics. Directions for working on a case and the system for analyzing instruction should ensure that the teachers are actively engaged in analysis of salient issues—making what Harrington called “reasoned decisions” (Harrington, 1995). To encourage the development of teachers’ reasoning, they are often asked to support their evaluation of lessons with evidence and explanation in written format (e.g., Copeland & Decker, 1996; Santagata & Angelici, 2010; van Es & Sherin, 2002).

CSRL conforms to these standards. There are 17 case studies, each focused on one teacher’s classroom. Each teacher contributed 2 to 4 lessons on a given topic (e.g., features of nonfiction texts). For example, one lesson might introduce the strategy of summarizing; a second lesson might provide guided application. A video provides the primary basis for analyzing instruction; it is accompanied by the following resources: (a) *Context* (about the school and classroom), (b) *About the lesson* (teacher’s explanation of the purpose and design of the lesson), (c) *Materials* (e.g., photocopies of books used in the lesson), (d) *Questions* (the framework for analysis of the lesson called the *Thinking Questions*), (e) *Reflection* (the classroom teacher’s reflections after the lesson). After completing analysis of a lesson, teachers can read *Comments* about the lesson from two experts in early literacy. The screen shot in Figure 1 gives a sense of the interactive nature of the program. Here the program user is examining the book while watching the video of the lesson.



Figure 1. Screenshot of CSRL lesson

1.3 Development of the Thinking Questions Framework

To develop a framework to direct and guide teachers’ analysis of reading lessons, we first needed a theoretical framework for analysis of instruction in reading. Research on effective pedagogy in general and its application to the teaching of reading suggested the importance of analyzing reading lessons as unfolding in the process of teaching. Shulman (1987) presented pedagogical reasoning as the core of the process of teaching; he described a cyclical process that begins with comprehension (e.g., determining purposes) and then moves through transformation (e.g., preparation, selection of methods of teaching), instruction (e.g., management, presentations), evaluation (e.g., checking for students’ understanding and evaluation of one’s own performance), and reflection (e.g., reviewing, critically analyzing); the end of the process is a new comprehension (e.g., consolidation of new understandings). Hiebert, Morris, Berk, and Jansen (2007) presented a similar albeit simpler view of the major

components of the teaching process that effective teachers examine in each lesson they teach. These include planning of lessons, choice of methods for students to learn and practice reading skills, and evaluation of students' response to the lesson (e.g., engagement). We adapted Hiebert et al.'s proposal of the major components (i.e., dimensions) of the teaching process to guide teachers' analysis of the quality of instruction in CSRL case studies. We identified specific features that represented each dimension by drawing on studies of effective instruction (e.g., Putnam & Borko, 2000; Roehler & Duffy, 1991; Shulman, 1987; Shulman, 1992; Rosenshine & Stevens, 1984; Snow, Griffin & Burns, 2005).

The three dimensions as well as specific features that represent each of these are shown in Appendix A. The first dimension, *Purpose and design*, focuses on the relation between the teacher's purpose or lesson plan and the events shown in the lesson. Porter and Brophy (1988) proposed that effective teachers know exactly what they intend to accomplish through their instruction; they explain the goal of the lesson to the students and keep the goal in mind as they teach. As Cameron, Connor, and Morrison (2005) found, clear goals and classroom management are critical to ensure that instruction is relevant and ensure optimal opportunities for students to learn. The second dimension is *Instruction*, reflecting the choice of practices used to foster student learning, given the content (e.g., explaining, coaching). Time spent providing direct instruction is critical, but so is teachers' guidance of activities that provide review, practice, and application (Foorman & Torgesen, 2001; Roehler & Duffy, 1991; Taylor, Pearson, Peterson, & Rodriguez, 2005). The third dimension is *Student engagement and participation*. The quality of students' learning is dependent on teachers' monitoring of students' response to a lesson and ability to motivate and engage students in the topic and activities of a lesson (Guthrie, Wigfield, Humenick, et al., 2006). Effective teachers use instructional actions to promote students' interest in their own literacy development (e.g., Pressley, Wharton McDonald, Raphael, et al., 2002). Porter and Brophy (1988, p. 82) stated that, "effective teachers continuously monitor their students' understanding of presentations and responses to assignments. They routinely provide timely and detailed feedback, but not necessarily in the same ways for all students."

In theory, teachers who seek to acquire expertise in reading instruction benefit when they are invested in the cognitively challenging work of analyzing instruction (Bransford, Derry, Berliner, et al., 2005; Harrison, Pead & Sheard, 2006; Rosaen, Lundeberg, Cooper, et al., 2008). To develop the quality of teachers' reasoning about issues of instruction in CSRL lessons, they were required to rate lesson features and explain their views in writing. Written analyses of case studies holds teachers responsible for explaining their perceptions of features of that affect lesson quality (e.g., Harrington, Quinn-Leering & Hodson, 1996; Lai & Calandra, 2010; van Es & Sherin, 2002).

The screenshot in Figure 2 shows that the teacher can work on the Thinking Questions while watching the video of the lesson

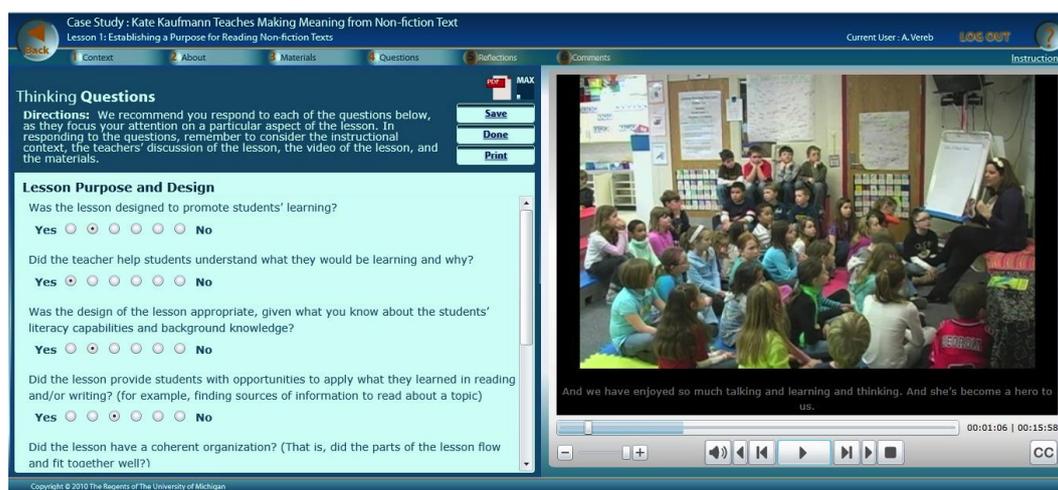


Figure 2. Responding to Thinking Questions while watching the lesson

1.4 Examining Premises of the Thinking Questions Framework

In the study reported herein, we examine the soundness of the theoretical framework by examining the technical characteristics of the Thinking Questions measure and by asking whether responses to the Thinking Questions varied with the expertise of the teacher. Such a study is necessary before we carry out further studies to determine the efficacy of the program for teachers (Desimone, 2009).

To accomplish this, we needed to recruit participants who would respond to all three dimensions of Thinking Questions for one or two case studies (ordinarily, they respond to one set in each lesson in order to avoid cognitive overload). Responses of each teacher to all three dimensions made it possible to examine the validity of the design of the Thinking Questions—for example, whether the three dimensions were independent components and whether the features that were chosen to represent each dimension contributed to that dimension. Because in theory teachers' pedagogical reasoning affects all aspects of their teaching, we also expected to find an overarching level of expertise to which the three dimensions would be related. Teachers' responses to particular questions (e.g., was the pace of the lesson appropriate?) might also be related to their general expertise.

In addition, we sought to determine whether groups of teachers who varied in their professional background differed in performance on the Thinking Questions, given the expectation that analytic reasoning about instruction was a characteristic associated with expertise (e.g., Bransford et al., 2005). On the basis of previous study results (Ainley & Luntley, 2007; Krull, Oras, & Sisask, 2007), we hypothesized that more experienced teachers would evaluate the features of reading lessons somewhat differently than less experienced teachers. Experts perceive "patterns" within the fast-moving dynamics of reading lessons that novices often do not notice; novices' analysis of instruction has been described as highly contextualized and focused on concrete and familiar elements (Bransford et al., 2005; Krull et al., 2007).

As the focus in CSRL is on effective pedagogy that could be applied to lessons with different literacy content, we wanted to be certain that the less experienced teachers were sufficiently knowledgeable to apply the Thinking Questions framework to different literacy lessons. It would be unrealistic to think that pedagogical reasoning (Shulman, 1987) can be effectively applied without adequate knowledge of the content area, in this case reading. Thus, we recruited less experienced teachers who were completing an advanced graduate course in elementary reading instruction and more experienced teachers, as determined by years' teaching, graduate degree attainment, and specialized endorsements. We reasoned that the less experienced teachers would have studied the content of early reading instruction (e.g., how to teach vocabulary) but would not have depth and breadth of experience evaluating pedagogical features – for example, would not be sensitive to "telling" features within the dynamics of lessons, as others have found (van Es & Sherin, 2002).

Finally, we wanted to examine more and less experienced teachers' explanations of their lesson ratings. Based on the results of previous studies (e.g., Krull et al., 2007), we expected more experienced teachers to be more prone to making generalizations about the quality of instruction, whereas the less experienced teachers would be likely to write longer responses, showing a tendency to describe instruction in detail. To analyze the written responses, we adapted qualitative methods developed by van Es and Sherin (2002); in particular, we rated the specificity of their explanations of effective features (ranging from general to very specific). We also examined the productivity of their responses (i.e., number of words they wrote).

The three research questions we addressed are as follows: (1) To what extent does teachers' performance on the rating scale suggest (a) support for a general level of expertise in analyzing instruction, (b) support for the three dimensions, and (c) support for the individual features as representative of the specific dimensions? (2) Do responses on the Thinking Questions scale differentiate teachers with more and less professional experience? (3) To what extent do more and less experienced teachers differ in analysis of effective features of lessons in their written comments?

2. Method

2.1 Participants

We recruited two groups of participants. One group was made up of experienced teachers, identified as having expertise in early reading through degree attainment and endorsements or certifications (e.g., National Broad Certification) ($n = 21$). The second group was made up of graduate students in an advanced literacy course ($n = 30$). Both groups of teachers were asked to respond to the Thinking Questions in Diana Richard's (DR) case study ("Using Text Features to Support Comprehension," made up of three lessons) and Kate Kaufmann's (KK) case study, "Making Meaning from Non-fiction Text" (made up of 3 lessons). (Case study teachers' names are pseudonyms.) However, because of time limitations, participating graduate students completed just the DR study as part of their coursework; they were invited to evaluate KK's case study at the completion of their course, and those that did so ($n = 4$) received an honorarium and certificate of participation documenting professional development hours. The more experienced teacher participants ($n = 21$) completed both case studies and received an honorarium and a certificate of participation documenting professional development hours for their personal portfolio.

All but two of the more experienced teachers had 4 years or more in teaching K-3 and a masters' degree or above. (The two teachers who taught less than 4 years in grades K-3 had a master's degree with a reading specialist

endorsement.) Sixty-two percent of the more experienced group had 4-14 years of teaching experience in grades K-3 while 29% had 15 years or greater experience in teaching grades K-3. They were National Board certified (NBC) or had one or more endorsements in Reading/Literacy, ESL, or Early Childhood. The less experienced teachers had fewer than 4 years teaching in kindergarten through grade 3; they had a bachelors' degree. None were National Board certified or had an endorsement in reading.

2.2 Data Sources

Study participants used the CSRL website program to study one or two case studies (<http://csrl.isr.umich.edu/csrl.aspx>). As the program is under study, the site is accessible, but the case studies are password protected.

2.2.1 Responses to Thinking Questions

We asked each participating teacher to complete the entire set of Thinking Questions (i.e., all three dimensions) for the DR case study (2 lessons) and the KK case study (3 lessons). Following their rating of items on the scale, teacher-users provided an overall rating for that dimension. Then they wrote responses to the two open-ended questions, as described earlier (see Appendix A). For this study, we analyzed teachers' written responses to the question focused on effective features for the lessons in the DR case study: "With [the purpose and design of the lesson; effective instruction; students' engagement and participation] in mind, please comment on a few *effective features* of the lesson?"

2.2.2 Survey

Teachers completed an electronic survey to provide information on their professional background (i.e., years teaching K-3 literacy, degree attainment, professional certifications and licensure).

2.3 Procedures for Analysis of Likert Scale Responses to the Thinking Questions

While the Thinking Questions focused on three primary dimensions that represent pedagogical reasoning, we hypothesized that the response patterns would also support an overall factor that describes teachers' general expertise in analyzing instruction, as previously discussed. To determine whether this hypothesis was supported by teachers' responses, we planned to examine several models. A unidimensional model would suggest that responses to the Thinking Questions scale are largely guided by only a single underlying factor, whereas a three-dimensional model would suggest that responses are directly dependent on only the theoretical dimension to which they belong (e.g., instruction). The third model was a bifactor model; the bifactor approach has been used in a wide array of studies in which researchers are interested in both general and specific factors because it offers two key utilities (e.g., Chen, West, & Sousa, 2006; Reise, 2012). First, the bifactor model helps to disentangle the relative strength with which a general or overall factor contributes to all item responses from the strength with which specific factors inform different sets of item responses. Second, the bifactor model affords easier interpretations of the relationships between each of the factors and external covariates; this is because it constructs the general and specific factors orthogonally by consolidating the parts of the specific factors that are common across all factors (Holzinger & Swineford, 1937; Chen, West, & Sousa, 2006). In a bifactor model, each response to a Thinking Question is guided both by general expertise in evaluating teaching and by a secondary dimension specific to the theoretical dimension to which each item belongs.

We conducted a series of item factor analyses to determine whether a one-dimensional, three-dimensional or bifactor structure best fit the observed data. Statistically, we can express the bifactor model as

$$P(Y_{it} = m | \theta) = P(Y_{it} \geq m | \theta) - P(Y_{it} \geq m+1 | \theta) = \frac{1}{1 + \exp\{-[a_i^g \theta^g + a_i^j \theta^j - d_i^m]\}} - \frac{1}{1 + \exp\{-[a_i^g \theta^g + a_i^j \theta^j - d_i^{m+1}]\}} \quad (1)$$

where Y_{it} is the score assigned to Thinking Question i for lesson l by teacher t , m is a specific score, a_i^g is the loading parameter for Thinking Question i for the general dimension, θ^g is teacher t 's general or overall judgment of the lesson. Further, a_i^j is the domain-specific loading parameter for Thinking Question i onto dimension j to which the Thinking Question belongs (i.e., j =lesson purpose and design, instruction, or students' engagement and learning), θ^j is teacher t 's judgment of the lesson with regard to specific dimension j , d_i^m is the threshold of category m in Thinking Question i and M represents the number of categories with m as a specific category. For the alternative models considered, the three-factor model drops the general dimension, θ^g , in equation (1) whereas the one factor model drops the three specific dimensions, θ^j . For each latent variable the scale was set to have a standard normal distribution with mean zero and unit variance. Covariances among the dimensions were estimated in the three-dimensional model whereas the bifactor model collected the shared covariance among factors to form the general factor (Cai, Yang, & Hansen, 2011).

Analyses to answer our second research question involved examination of the extent to which teachers' scores on the underlying factors correlated with their level of professional experience. To do so, we concurrently linked the factors in equation (1) with an explanatory component such that

$$Q = \beta_0 + \beta_1 X_t + u_t \quad (2)$$

where β_0 is the intercept, X_t represents the level of experience for teacher t with coefficient β_1 , and u_t as the normally distributed residual error. This analysis constitutes a preliminary test of external validity of the Thinking Questions scale.

2.4 Procedures for Analysis of Written Responses

Codes for analyzing the open-ended questions were developed for each of the three dimensions. In the qualitative component of our study, we used two measures. One measure is the productivity of their responses (i.e., word count). The second, adapted from van Es and Sherin (2002), is the specificity of teachers' analysis of effective features. Appendix B provides a table with levels of specificity, a definition of each level, and an example from the written responses to the question on effective features. The coding process was developed and refined by the research team until the codebook contained clear definitions and examples of teacher-user responses. Qualitative analyses of the written responses were carried out with coders blind to group membership. Once team members coded and rated responses individually, they discussed any differences in ratings until a consensus was reached.

3 Results

3.1 Exploring the Factor Structure of the Thinking Questions

Our first research question tested the hypothesis that teachers' responses to Thinking Questions would be guided by both general and specific factors (i.e., dimensions). We expected that the results would support a bifactor structure, such that teachers' responses reflected their general expertise in analyzing the effectiveness of features of reading instruction and also factors specific to the theoretical dimensions.

As a preliminary evaluation of this hypothesis, we compared the fits of one-dimensional, three-dimensional, and bifactor models. The likelihood ratio test indicated the improved fit offered by the bifactor model (as compared to the three- and one-dimensional structure). (See Table 1.) The information criteria were split. Whereas the Akaike information criterion (AIC) preferred the bifactor model, the more conservative Bayesian Information Criterion (BIC) preferred the three-dimensional model. However, the three-dimensional model revealed high correlations among the three factors (see Table 2), again suggesting a general factor.

Table 1. Comparison of the One-Dimensional, Three-Dimensional and Bifactor Models

Model	Deviance	Deviance test p -value	AIC	BIC
Bifactor	6281	--	6509	6870
Three dimensions	6325	<0.001	6520	6831
One dimension	6363	<0.001	6549	6843

Note. Deviance test compared the bifactor model against the alternative model.

Table 2. Correlations among Dimensions in Three-Dimensional Model

Dimension	Lesson purpose and design	Instruction	Students' engagement and participation
Lesson purpose and design	--		
Instruction	0.94	--	
Students' engagement and participation	0.79	0.85	--

We examined the relational strength of the individual Thinking Questions to their dimensions. As can be seen in Table 3, the results show the strength of the relationships between the questions and the general expertise in analyzing instruction, as well as the strength of the relationships between the questions and their specific dimensions (above and beyond that of the general dimension). Most questions showed very strong relationships to the general dimension. One notable exception was question 4 under "Students' engagement and participation" (i.e., "Were students given opportunities to work with one another and to share their ideas?"). By comparison, the majority of questions loaded weakly onto the specific dimensions, especially for the "Instruction" dimension. With the exception of the fourth question under "Students' engagement and participation," all questions show greater dependence on the general dimension than their respective specific dimensions.

Table 3. Loadings for the Bifactor Item Factor Analysis of Thinking Questions

Item	General	Lesson Purpose & Design	Instruction	Students' Engagement & Participation
A1	1.991	0.605	--	--
A2	1.149	0.441	--	--
A3	1.96	0.568	--	--
A4	1.263	0.726	--	--
A5	1.516	0.455	--	--
A6	1.936	0.560	--	--
B1	1.299	--	0.281	--
B2	1.242	--	0.032	--
B3	1.016	--	0.120	--
B4	0.719	--	0.132	--
B5	1.551	--	0.348	--
B6	1.274	--	0.332	--
C1	1.114	--	--	0.869
C2	0.596	--	--	0.500
C3	1.012	--	--	0.279
C4	0.133	--	--	0.600
C5	1.120	--	--	0.471
C6	1.107	--	--	0.991

Note. A = Lesson purpose and design; B = Instructional methods; C = Student engagement and participation. The figures are conceptually standard regression coefficients; they are presented on a log-odds scale.

Because of the strong relationship between the specific questions and general dimension, the ability of our scale to reliably describe teachers' expertise in analyzing instruction was quite strong. However, the weak relationship between the questions within each dimension and the dimensions themselves constrained the ability of our scale to describe differences among teachers in terms of the three specific dimensions. Figure 3 displays the scale information function provided for each of the dimensions. The scale information function is an approach to describing the amount of measurement error or uncertainty present in scales and is similar to the reliability of a scale but varies as a function of the underlying ability, θ .

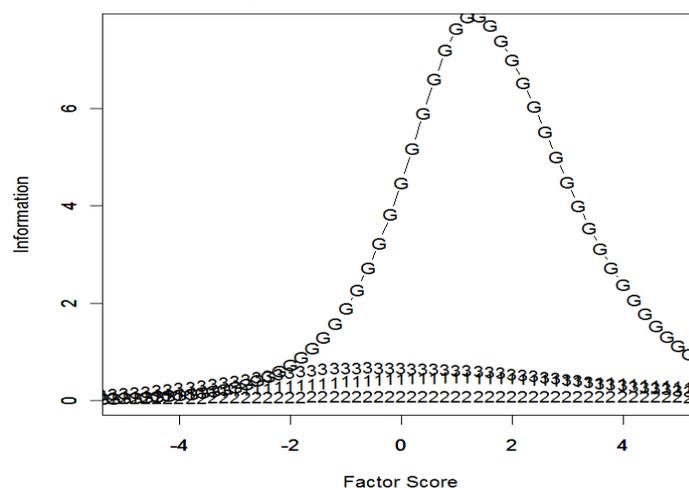


Figure 3. Information Provided by the Scale for the Specific and General Dimensions/factors

As is evident from Figure 3, scores on the general dimension (labeled G) representing expertise in analysis of lessons are highly reliable. The scale provides the most information just above the mean (zero) where its reliability is approximately 0.89. Even at three standard deviations above the mean and about a half a deviation below the mean the corresponding reliability was still about 0.80. On the other hand, the specific dimensions (labeled 1, 2, and 3) contained considerable measurement error. The reliability of each specific dimension remained below 0.50 and likely contained substantial measurement error.

3.2 Differences in the Responses of More and Less Experienced Teachers

Our second question was designed to explore the predictive validity of the Thinking Question factor structure; we expected differences on the factors between more and less experienced teachers. As shown in Table 4, our analyses

indicated that the response patterns of more and less experienced teachers differed significantly on both the general dimension and on each of the three specific dimensions, suggesting that the bifactor structure is capturing important general and dimension-specific differences in teachers' responses. Although more and less experienced teachers differed, we are unable to comment on the authority of these differences (whether more experienced teachers response were more 'correct') because there are no 'true' answers to each question, based on the video of the lesson.

Table 4. Relations between Teachers' Responses to Specific Dimensions and Their Professional Experience

Dimension	Coefficient	Standard error	p-value
General	0.50	(0.12)	<0.01
Lesson purpose & design	0.85	(0.07)	<0.01
Instruction	0.86	(0.10)	<0.01
Student Engagement	0.69	(0.11)	<0.01

3.3 Differences in Written Responses

The third question focused on comparison of the two groups on two measures: word count (a count of all words written in the text box) and specificity (coding criteria identified 4 levels, ranging from non-specific to very specific) (see Appendix B). With regard to word count, the less experienced teachers wrote significantly longer responses than the more experienced teachers. Table 5 shows the means and standard deviations by lesson and dimension for the two teacher groups; Table 6 gives the results of t-tests.

Table 5. Comparison of Word Count in Written Comments on Effective Features of DR Lessons by More and Less Experienced Teachers

	Lesson and Dimension	Word Count
Less experienced	L1, A	96.17 (48.4)
	L1, B	89.5 (66.5)
	L1, C	67.0 (44.0)
	L2, A	76.0 (41.0)
	L2, B	90.2 (45.4)
	L2, C	81.1 (56.4)
More experienced	L1, A	61.4 (37.5)
	L1, B	47.1 (31.0)
	L1, C	41.9 (35.5)
	L2, A	54.0 (30.8)
	L2, B	53.3 (44.0)
	L2, C	46.0 (33.8)

Note. L = Lesson; A = Lesson Purpose and Design; B = Instruction; C = Student Engagement and Participation

Table 6. Comparison of Word Count in Comments Written by More and Less Experienced Teachers on Effective Features of DR Lessons

Lesson	Dimension	t-value	Probability (p) value
Lesson 1	Lesson purpose & design	2.757	.008
	Instruction	3.047	.004
	Students' participation & engagement	2.167	.035
Lesson 2	Lesson purpose & design	2.084	.042
	Instruction	2.892	.006
	Students' participation & engagement	2.549	.014

To compare the two groups on specificity, we used nonparametric analyses, examining the number of teachers in each group whose response was coded at each level (1 through 4 for each measure). Results showed that the less experienced teachers were rated as providing explanations of effective features with higher levels of specificity than the experienced teachers in two of the six written responses; Table 7 shows the percent of teachers who received ratings on the scale of 1 to 4. We note that only for the Lesson 2, C Student Engagement, did the any of teachers provide very general comments (receiving a rating of 1). Overall, experienced teachers tended to give moderately specific comments, whereas less experienced teachers tended to focus more on details in the lesson. Significant differences by group were found for Lesson 1, A Purpose and Design, X^2 (df 2) = 9.772, p = .008, and for Lesson 1, B Instruction, X^2 (df 2) = 7.713, p = .021.

Table 7. Ratings of Specificity of More and Less Experienced Teachers' Written Comments on Effective Features

Lesson and Dimension	Ratings	1	2	3	4
Lesson 1, A*	Less Exper		3.3	50	46.7
	More exper		33.3	47.6	19
Lesson 1, B*	Less exper		3.3	53.3	43.3
	More exper		31.6	42.1	26.3
Lesson 1, C	Less Exper		26.7	50	23.3
	More exper		31.6	52.6	15.8
Lesson 2, A	Less exper		6.7	56.7	36.7
	More exper		21.1	57.9	21.1
Lesson 2, B	Less exper		6.7	56.7	36.7
	More exper		15.8	47.4	36.8
Lesson 2, C	Less exper	6.7	10	43.3	40
	More exper	5.3	21.1	52.6	21.1

Note. Less exper (less experienced) = 30; More exper (more experienced) = 21.

* Significant Pearson Chi-square for Lesson 1, A: X^2 (df2) = 9.772, $p = .008$; for Lesson 1, B: X^2 (df 2) = 7.713, $p = .021$.

4. Discussion

CSRL was developed for use as a program for practicing teachers to improve their analysis of the quality of early reading instruction. Built into the web-based program was systematic guidance for teachers in the analysis of the case studies (Merseeth, 1992; van Es & Sherin, 2002). The Thinking Questions framework both directs teachers' work on the case studies and offers a theoretical basis for evaluating dimensions and features of reading instruction. It was necessary, therefore, to investigate the extent to which teachers' responses to the Thinking Questions met the expectations of the theoretical design and differentiated more and less experienced teachers.

The results largely supported our hypotheses, although they uncovered issues that need further investigation. First, they provide evidence that the responses to the Thinking Questions reveal variation in teachers' expertise in analysis of reading lessons. Second, more and less experienced teachers performed differently on the rating scale, and their written explanations showed that more experienced teachers were more likely to offer generalizations about effective features. In what follows, we reflect on what the results tell us about the design and promise of CSRL.

4.1 General and Specific Dimensions That Contribute to Analysis of Lessons

The Thinking Questions focus teachers' attention on three dimensions central to the process of teaching, as proposed by Hiebert et al. (2007) and Shulman (1987); these focus on how teachers plan, teach, and evaluate their lessons. We organized the Thinking Questions framework so that teachers concentrated on evaluation of the three dimensions separately. However, because these dimensions are interrelated parts of a single process, we anticipated that each would tap into teachers' general expertise in analyzing reading lessons. The results supported this expectation. Specifically, the bifactor model showed that a general dimension representing expertise in analyzing reading instruction characterized teachers' responses across the three dimensions.

The three specific dimensions were largely subsumed by this general dimension, so that individually they contributed little additional information. The three specific dimensions were significantly related to one another (as shown in Table 2). Furthermore, ratings of the particular features under each dimension actually loaded more strongly on the general dimension. We are not convinced that the three specific dimensions do not in some way distinguish teachers' expertise in analysis reading instruction, above and beyond that of the general dimension. Possibly the relatively small number of teachers in the study (51 overall) constrained our ability to measure multiple dimensions. However, it makes sense that general expertise in analysis of instruction influenced teachers' views of the instruction across the specific dimensions.

We note that the loadings in the dimension called Instruction are generally quite low. This may be because of the variety of teachers' actions and activities that were included in this dimension. Similarly, in his discussion of aspects of pedagogical reasoning, Shulman (1987) discussed the construct of instruction as including the wide variety of teaching actions used during lessons, including managing the classroom, providing explanations, and assigning activities (to name a few). The choice of methods depends on judgments about appropriate ways the content of lessons might be delivered. The diversity of the features that represent Instruction in the Thinking Questions framework might make this dimension appear to lack coherence, as reflected in the low reliability of items. It would

be valuable to carry out a study of a revised version of the Instruction dimension that included additional features of instruction. Moreover, such a study might involve a larger group of teachers to determine whether there is any empirical basis for grouping the instructional features as we have done in the Thinking Questions.

4.2 Responses of More and Less Experienced Teachers

Examination of the factor structure of the responses to the Thinking Questions showed differences in the patterns of more and less experienced teachers. A major finding is that we were able to reliability measure teachers' perceptions of the quality of key pedagogical features of reading lessons. Our results are compatible with those of previous researchers (e.g., Krull et al., 2006) in showing that, compared to relatively inexperienced teachers, teachers categorized as experts tended to make judgments based on previous experience in order to identify effective features of lessons. This might suggest that, with repeated use of the CSRL program, teachers who appeared at the outset to lack the ability to identify "telling" features of lessons might increasingly show expertise in analysis of reading lessons, as researchers have found in mathematics (e.g., van Es & Sherin, 2002; Santagata & Angelici, 2010). As we recruited one group of less experienced teachers who were taking a graduate course in reading, we surmise that the nature and type of their responses to the Thinking Questions were not simply driven by lack of content knowledge about reading and reading instruction.

With regard to teachers' written explanations of effective features of the lessons, we expected that differences in the specificity of the written comments would distinguish more and less knowledgeable and experienced teachers (e.g., Krull et al., 2007; van Es & Sherin, 2002). Analyses of the teachers' written explanations indicated that the more experienced teachers were more prone to making generalizations, based on their knowledge of effective practices in teaching early reading. This tendency might reflect their ability to efficiently identify patterns and relations among lesson features (e.g., van Es & Sherin, 2002). For example, an excerpt from one of the more experienced teachers' comments is as follows: "There was a very clear purpose, which was stated to the students. The parts of the lesson flowed seamlessly, with everything relating to the lesson. During the guided reading time, it was evident that the book was at the instructional level of the group and all students were engaged in reading activities, rather than round robin. This strategy works for all levels of readers."

The less experienced teachers had higher specificity ratings on two of the six written comments about effective features. They were more likely than experienced teachers to focus on particular details in the lessons. Note, for example, the detailed account in the following written comment:

Within this lesson, there were many features of effective instruction. In the beginning, Ms. Kaufmann stressed the purpose for the lesson, strategically connecting the activities and thinking for the day to previously explored lessons. She demonstrated her own thinking, based upon the topic of Harriet Tubman, which the students were all familiar with due to their exploration of the book. Ms. Kaufmann instructed for about eight minutes and then allowed students to get up and relocate to a different part of the room to practice the skill that she had demonstrated. Her follow-up lesson lasted about the same amount of time, showing that she emphasizes actual engagement in the act of reading as the most important part of the reading lesson. Students were able to share with each other, which gave each of them an active role to play in the mini-lesson. This "turn and talk" time allowed all students to share their thinking in a low-pressure activity while learning from their partner's comments.

This tendency might also have contributed to their significantly longer responses. The efforts the less experienced teachers made to explain what they saw as effective features might have been affected by their concurrent study of reading instruction in their graduate course.

Apart from differences we observed in the specificity and productivity of written responses by teachers in the two groups, it is our impression that the process of explaining in writing perceptions of the quality of instruction contributes to an analytic mindset (e.g., Harrington et al., 1996), although further research is needed to explore this possibility. Other researchers have found that providing teachers with guidance in analysis of videotaped lessons contributes to their understanding of effective instruction, particularly as applied to their own teaching (e.g., Santagata & Angelici, 2010).

4.3 Summary, Implications, Limitations, and Future Research

Overall, the results suggest that the Thinking Questions framework is a promising way to provide guidance for analysis of instruction in a case studies program for teachers of early reading. The Thinking Questions framework offers teachers a system that they can use to analyze reading lessons they teach, as well as those taught by others. Applied to different lessons and cases, the framework helps teachers appreciate not only the situative nature of teaching but also the value of practice in analyzing actual instructional events as a way to develop expertise.

While factor analyses might suggest collapsing the three dimensions into one, we suspect that there are good reasons to retain the division of questions into three dimensions. From a theoretical perspective, the separate dimensions provide structure to a teacher's process of planning, carrying out, and evaluating lessons. A worthwhile step would be to get teachers' perceptions of the value of the Thinking Questions framework with and without the three-dimensional structure. Clearly, planning lessons (as captured by the dimension, Lesson purpose and design) affects instruction (the second dimension), and evaluation of lessons is likely to take both the initial plan and the characteristics of instruction into account. This is an important issue for further research not only for CSRL but also for testing the theoretical framework, which posits that specific features of lessons need to be evaluated within the context of the teaching process.

The study has some limitations and areas where further research is needed. First, we were able to recruit a relatively number of participants; with a larger group, we might find greater support for the role of the three dimensions in teachers' analysis of instruction. A second issue involves the question of criteria to validate teachers' "expertise"—a problem that Krull et al. (2007) tried to address. For example, it is not clear that a combination of years of teaching and advanced educational attainment truly distinguishes experts from novices. What may be needed is an independent measure of teachers' knowledge about reading. Third, further study is needed to scale up CSRL in order to determine whether the program supports teachers' learning about effective instruction that transfers to their own instruction (Desimone, 2009). Ultimately, some evidence is needed to support the premise that use of a case studies program such as CSRL affects teachers' practices and their students' reading achievement.

For the present, we have demonstrated that it is possible to provide direction and guidance for practicing teachers' study of cases of reading lessons within a video-based, multimedia program. This preliminary study of the framework within CSRL suggests that it offers practicing teachers an opportunity to deepen their understanding of effective instruction in elementary reading.

Author Note

This study was made possible by a Teacher Quality grant from the Institute for Education Sciences (IES) to the first author (Award #R305A080295-09). IES is not responsible for the design of the website program and the study or interpretations of results. We thank Emily Mihocko, Andrea Melnick, Jodi Johnson, Meredith House, Larry Laferte, and our advisory board for their assistance in the development of CSRL. We also would like to recognize the enormous contributions of the elementary teachers who helped us prepare case studies of their reading instruction.

References

- Ainley, J., & Luntley, M. (2007). Towards an articulation of expert classroom practice. *Teaching and Teacher Education, 23*, 1127-1138. <http://dx.doi.org/10.1016/j.tate.2006.09.002>
- Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher, 33*, 3-15. <http://dx.doi.org/10.3102/0013189X033008003>
- Borko, H., Whitcomb, J., & Liston, D. (2009). Wicked problems and other thoughts on issues of technology and teacher learning. *Journal of Teacher Education, 60*, 3-7. <http://dx.doi.org/10.1177/0022487108328488>
- Bransford, J., Derry, S., Berliner, D., Hammerness, K., & Beckett, K. L. (2005). Theories of learning and their roles in teaching. In L. Darling-Hammond & J. Bransford, *Preparing teachers for a changing world* (pp. 40-87). San Francisco: Jossey-Bass.
- Cai, L., Yang, J. S., & Hansen, M. (2011, May 2). Generalized full-information item bifactor analysis. *Psychological Methods, 16*, 221-248. <http://dx.doi.org/10.1037/a0023350>
PMid:21534682 PMCID:PMC3150629
- Cameron, C. E., Connor, C. M., & Morrison, F. J. (2005). Effects of variation in teacher organization on classroom functioning. *Journal of School Psychology, 43*, 61-85. <http://dx.doi.org/10.1016/j.jsp.2004.12.002>
- Chen, F., West, S., & Sousa, K. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research, 41*, 189-225. http://dx.doi.org/10.1207/s15327906mbr4102_5
- Copeland, W. D., & Decker, D. L. (1996). Video cases and the development of meaning making in preservice teachers. *Teaching and Teacher Education, 12*, 467-481. [http://dx.doi.org/10.1016/0742-051X\(95\)00058-R](http://dx.doi.org/10.1016/0742-051X(95)00058-R)
- Dede, C., Ketelhut, D. J., Whitehouse, P., Breit, L., & McCloskey, E. M. (2009). A research agenda for online teacher professional development. *Journal of Teacher Education, 60*, 8-19. <http://dx.doi.org/10.1177/0022487108327554>
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better

- conceptualizations and measures. *Educational Researcher*, 38, 181-199.
<http://dx.doi.org/10.3102/0013189X08331140>
- Foorman, B. R., & Torgesen, J. (2001). Critical elements of classroom and small group instruction promote reading success in all children. *Learning Disabilities Research and Practice*, 18, 203-212.
<http://dx.doi.org/10.1111/0938-8982.00020>
- Guthrie, J. T., Wigfield, A., Humenick, N. M., Perencevich, K. C., Taboada, A., & Barbosa, P. (2006). Influences of stimulating tasks on reading motivation and comprehension. *Journal of Educational Research*, 99, 232-245. <http://dx.doi.org/10.3200/JOER.99.4.232-246>
- Harrington, H. (1995). Fostering reasoned decisions: Case-based pedagogy and the professional development of teachers. *Teaching and Teacher Education*, 11, 203-214. [http://dx.doi.org/10.1016/0742-051X\(94\)00027-4](http://dx.doi.org/10.1016/0742-051X(94)00027-4)
- Harrington, H., Quinn-Leering, K., & Hodson, L. (1996). Written case analyses and critical reflection. *Teaching and Teacher Education*, 12, 25-37. [http://dx.doi.org/10.1016/0742-051X\(96\)89078-0](http://dx.doi.org/10.1016/0742-051X(96)89078-0)
- Harrison, C., Pead, D., & Sheard, M. (2006). 'P, not P and possibly Q': Literacy teachers learning from digital representations of the classroom. M. McKenna, L. D. Labbo, R. D. Kieffer, & D. Reinking (Eds.), *International handbook of literacy and technology (Vol. 2)* (pp. 257-272). Mahwah, NJ: Erlbaum.
- Hiebert, J., Morris, A. K., Berk, D., & Jansen, A. (2007). Preparing teachers to learn from teaching. *Journal of Teacher Education*, 58, 47-61. <http://dx.doi.org/10.1177/0022487106295726>
- Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika*, 2, 41-54.
<http://dx.doi.org/10.1007/BF02287965>
- Krull, E., Oras, K., & Sisask, S. (2007). Differences in teachers' comments on classroom events as indicators of their professional development. *Teaching and Teacher Education*, 23, 1038-1050.
<http://dx.doi.org/10.1016/j.tate.2006.02.001>
- Lai, G., & Calandra, B. (2010). Examining the effects of computer-based scaffolds on novice teachers' reflective journal writing. *Educational Technology, Research, and Development*, 58, 421-437.
<http://dx.doi.org/10.1007/s11423-009-9112-2>
- Merseth, K. K. (1996). Cases and case methods in teacher education. In J. Sikula et al., (Eds.), *Handbook of research on teacher education*, 2nd ed. (pp. 722-744).
- Porter, A., & Brophy, J. (1988, May). Synthesis of research on good teaching: Insights from the work of the Institute for Research on Teaching. *Educational Leadership*, 74-85.
- Pressley, M., Wharton McDonald, R., Raphael, L.M., Bogner, L. & Roehrig, A. (2002). Exemplary first grade teaching. In B. M. Taylor and P. D. Pearson (Eds.), *Teaching reading: Effective schools, accomplished teachers* (pp. 73-88). Mahwah, NJ: Erlbaum.
- Putnam, R. T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher*, 29, 4-15. <http://dx.doi.org/10.3102/0013189X029001004>
- Reise, S. P. (2012). The rediscovery of bi-factor measurement models. *Multivariate Behavioral Research*, 47, 667-696. <http://dx.doi.org/10.1080/00273171.2012.715555>
- Roehler, L. R., & Duffy, G. G. (1991). Teachers' instructional actions. In R. Barr, M. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.) *Handbook of reading research* (pp. 861-883). NY: Longman.
- Rosaen, C. L., Lundeberg, M., Cooper, M., Fritzen, A., & Terpstra, M. (2008). Noticing noticing: How does investigation of video records change how teachers reflect on their experiences? *Journal of Teacher Education*, 59, 347-360. <http://dx.doi.org/10.1177/0022487108322128>
- Rosenshine, B., & Stevens, R. (1984). Classroom instruction in reading. In P. D. Pearson, R. Barr., M. L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research (Vol. 1)* (pp. 745-798). NY: Longman.
- Santagata, R., & Angelici, G. (2010). Studying the impact of the Lesson Analysis Framework on preservice teachers' abilities to reflect on videos of classroom teaching. *Journal of Teacher Education*, 61, 339-349.
<http://dx.doi.org/10.1177/0022487110369555>
- Santagata, R., Zannoni, C., & Stigler, J. W. (2007). The role of lesson analysis in preservice teacher education: An empirical investigation of teacher learning from a virtual video-based field experience. *Journal of Mathematics Teacher Education*, 10, 123-140. <http://dx.doi.org/10.1007/s10857-007-9029-9>

- Schrader, P.G., Leu, D. L., Kinzer, C. K., Ataya, R., Teale, W. H., Labbo, L. D., & Cammack, D. (2003). Using internet delivered video cases to support preservice teachers' understanding of effective early literacy instruction: An exploratory study. *Instructional Science*, *31*, 317-340. <http://dx.doi.org/10.1023/A:1024690111227>
- Shulman, L. S. (1987). Knowledge and teaching: Foundations for the new reform. *Harvard Educational Review*, *57*, 1-22.
- Shulman, L. S. (1992). Toward a pedagogy of cases. In J. Shulman (Ed.), *Case methods in teacher education* (pp. 1-30). New York: Teachers College Press.
- Snow, C. E., Griffin, P., & Burns, M. S. (2005). *Knowledge to support the teaching of reading: Preparing teachers for a changing world*. San Francisco: Jossey-Bass.
- Taylor, B. M., Pearson, P. D, Peterson, D., & Rodriguez, M. C. (2005). The CIERA school change framework: An evidence-based approach to professional development and school reading improvement. *Reading Research Quarterly*, *40*, 40-69. <http://dx.doi.org/10.1598/RRQ.40.1.3>
- 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*, 571-596.
- van Es, E. A., & Sherin, M. G. (2006). How different video club designs support teachers in "learning to notice." *Journal of Computing in Teacher Education*, *22*, 125-135.

Appendix A

The Thinking Questions

Note: Question 1-6 in each dimension below is followed by the following scale for teachers' responses:

Yes No

Directions: We recommend you respond to each of the questions below, as they focus your attention on a particular aspect of the lesson. In responding to the questions, remember to consider the instructional context, the teachers' discussion of the lesson, the video of the lesson, and the materials.

Dimension A: Lesson Purpose and Design

1. Was the lesson designed to promote students' learning?
2. Did the teacher help students understand what they would be learning and why?
3. Was the design of the lesson appropriate, given what you know about the students' literacy capabilities and background knowledge?
4. Did the lesson provide students with opportunities to apply what they learned in reading and/or writing? (for example, finding sources of information to read about a topic)
5. Did the lesson have a coherent organization? (That is, did the parts of the lesson flow and fit together well?)
6. Overall, was the lesson effectively designed to achieve a literacy purpose meaningful to the students?

Open-ended questions:

1. With the purpose and design of the lesson in mind, please comment on a few effective features of the lesson.
2. With the purpose and design of the lesson in mind, please offer a few suggestions for ways to improve the lesson.

Dimension B: Instruction

1. Were literacy concepts, skills, and strategies taught effectively?
2. Was the text used effectively in the lesson?
3. Did the teacher provide clear explanations of literacy concepts and processes?
4. Was the pace of the lesson appropriate?
5. Did the activities in the lesson advance students' learning?

6. Overall, was the instruction effective?

Open-ended questions:

1. With effective instruction in mind, please comment on a few effective features of the lesson.
2. With effective instruction in mind, please offer a few suggestions for ways to improve the lesson.

Dimension C: Students' Engagement and Participation

1. Did features of the lesson engage the students' interest and participation? (for example, expressed interest in the topic of the book)
2. Was there sufficient opportunity for the students to discuss texts and/or contribute to reflections on literacy concepts and processes? (for example, the students were asked to explain their ideas)
3. Did the teacher monitor students' understanding of and participation in the lesson? (for example, provided feedback about their work)
4. Were students given opportunities to work with one another and to share their ideas?
5. Were the teachers' activities and discourse responsive to the needs of individual students?
6. Overall, did the lesson foster students' engagement and participation?

Open-ended questions:

1. With students' engagement and participation in mind, please comment on a few effective features of the lesson.
2. With students' engagement and participation in mind, please offer a few suggestions for ways to improve the lesson.

Appendix B

Specificity is coded as a global rating for the entire comment; it reflects how detailed and clear a picture a response evokes. The table below shows the ordering of the codes from very general to very specific, the definition of each level, and an example from analysis of written responses to the question on effective features.

Specificity	Definition	Examples (excerpts from Effective Features responses)
1 General	<i>General</i> is coded for broad statements with no explanations or examples, often in the form of bullets or lists.	"The teacher modeled effectively. She valued their responses. Good scaffolding." (TB1EF)
2 Moderately General	<i>Moderately General</i> is coded for broad statements with limited examples/discussion. These are comments that are too general to form a clear picture of what is being described or suggested.	"The purpose was clearly stated in terms the children could understand. Her referral back to the previous day's story and lesson about the necessity to read for understanding laid that groundwork very effectively. I also liked the pair/share she had students do to elicit a variety of responses because it made the decision as to which word looked right more meaningful." (TB1EF)
3 Moderately Specific	<i>Moderately Specific</i> is coded when there is at least one specific statement, although it has some holes/gaps in explanation.	"There was a very clear purpose, which was stated to the students. The parts of the lesson flowed seamlessly, with everything relating to the lesson. During the guided reading time, it was evident that the book was at the instructional level of the group and all students were engaged in reading activities, rather than round robin. This strategy works for all levels of readers. Showing an actual example of how students in the guided reading group used these strategies to the rest of the class at the debriefing is very powerful. I liked

		the organizational structure of sitting in a circle, with the implicit meaning that we are all equals in this sharing and may all have something to contribute. The teacher was very positive with the students. She called them readers.” (TB1EF)
4 Very Specific	<i>Specific</i> is coded for focused statements with clear, complete, and specific explanations and examples. Should include reasons. Should identify an effective feature.	“The teacher carried through on her mini lesson theme throughout the Reading Workshop time, reinforcing her teaching point after a strong instructional minilesson. The literature used during the small group instruction portion of the Reading Workshop time gave the students a chance to use the strategy that was the focus of the minilesson. One of the students that worked with the teacher in small group was able to give an example in the wrap-up at the very end of the lesson where she had applied the teaching point in her independent reading of text. Initial minilesson was focused, short and to the point. It also involved the students thinking with the teacher. There was spontaneous response during the minilesson and <u>indicated the students were thinking with the teacher</u> . She gave them a chance to practice the strategy being taught on their own and share with a friend <u>ensuring student involvement</u> . She called on the student who yawned and whose attention appeared to be wandering <u>to ensure involvement</u> and he seemed to be applying the strategy being taught. The students were helping each other in small group by discussing what they were doing to problem solve. The teacher kept close track of what everyone was doing throughout the small group lesson. Periodically the teacher pulled the small group together to focus their learning and make generalizations. That gave the students a chance to consolidate their understanding and learning.” (TB1EF)



This work is licensed under a [Creative Commons Attribution 3.0 License](https://creativecommons.org/licenses/by/3.0/).