Abstract Title Page

Title: The development of a classroom observational system for studying the quality of reading comprehension and vocabulary instruction in first grade classrooms.

Author(s): Russell Gersten, PhD, Instructional Research Group, Joe Dimino, PhD, Instructional Research Group, Madhavi Jayanthi, EdD, Instructional Research group, Rebecca Newman-Gonchar, PhD, Instructional Research Group.

Abstract Body

Background/context:

Intuitively, classroom observation is an appealing way to study the quality of instruction, and hopefully to link observed practices with students' growth in learning. Yet, reliable measurement of important qualities of instruction in a classroom has proved to be an intricate, arduous task.

Researchers over the years have developed a variety of observational systems for measuring classroom events and behaviors. At one end are the high-inference, open-ended qualitative naturalistic observations. In this measurement system, the observer takes detailed notes of the instructional events and behaviors observed in the classroom. The observer later makes an inference regarding the quality of instruction based on the field notes gathered. These subjective interpretations of quality, however, are often colored by the observer's own experiences and beliefs (McIntosh, Vaughn, Schumm, Haager, & Lee, 1994). High inference systems are thus ideal, when the research focus is still exploratory in nature.

At the other end of the continuum are low-inference, quantitative observation systems like those used by Anderson, Evertson, and Brophy (1979), Foorman, Francis, Beeler, Winikates, and Fletcher (1997), and Stallings and Kaskowitz (1974). Low inference systems are useful in recording discrete events and behaviors that can be easily defined, observed, and measured. The advantage of these systems is that low inference in needed on the observer's part, as the recording of instructional events and behaviors in a classroom has an objective basis. Low inference systems are therefore typically associated with high inter-rater reliabilities. However, low inference systems do not lend themselves well to observing and recording complex instructional events and behavior in a classroom. They are more applicable in answering research questions such as, "How much time was spent on teaching vocabulary, word-level reading, or reading connected text?", "How many students were engaged during the phonics segment of the lesson?", and "How much time was devoted to small group instruction?"

At the middle of the spectrum of measurement systems are the moderate inference measures such as inferential rating scales that are part qualitative and part quantitative in nature. This middle of the road measure has allowed researchers to combine some of the advantages of both high inference and low inference observation systems. They can produce inter-rater reliabilities (unlike qualitative field notes), but the reliabilities tend to be lower that those of the aforementioned systems (Gersten, Baker, Haager, & Graves, 2005). They are also not nearly as limiting as the low inference measures in the type of behaviors and events that can be measured.

Moderate inferential rating scales are useful in assessing more complex aspects of teaching such as quality of teacher modeling, clarity of explanations, and how vocabulary concepts are taught. Gersten, his colleagues, and other researchers (e.g., Edmonds & Briggs, 2003; Gersten et al., 2005; Haager, Gersten, Baker, & Graves, 2003) used rating scales to evaluate quality of reading instruction in classrooms in California and Texas. Rating scales have also played a role in the complex observation system developed by Foorman and colleagues (Foorman & Schatschneider, 2003).

Typically, scores on rating scales often have much stronger correlations with growth in outcomes than variables generated from low inference systems (Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004; Stoolmiller, Eddy, & Reid, 2000; Gersten, Carnine, Zoref, & Cronin, 1986). However, even with these advantages, the use of moderate inference measures in large-scale studies is of concern. Despite their high face-validity, difficulties in development of standardization procedures and clear-cut definitions have been of issue (Kennedy, 1999). In addition, rating scales are prone to halo effects; consequently, a high internal consistency reliability coefficient could mean either that an observational scale measures a construct reliably or that the observational scale is marred by halo effects. Such problems of interpretation do not arise with scales that require fewer global judgments and fewer inferences.

The construct of quality instruction is complex and multi-dimensional, open to numerous interpretations. How can quality of instruction be measured in a reliable and valid manner? This was a question we attempted to answer in our multiple-site randomized controlled study of professional development. Our goal was to assess the extent to which teachers implemented the approaches for teaching comprehension and vocabulary that experimental research consistently supported as being sophisticated. We needed a classroom observation system that was suitable for assessing the quality of reading comprehension and vocabulary instruction for students in Reading First classrooms across the country. Our measure therefore had to be sensitive to the various nuances and fine distinctions that epitomize classroom teaching. As we embarked on our study, the limitations of the various classroom observation systems were foremost in our minds.

We attempted to measure *quality* of instruction by considering *quantity* as a reasonable surrogate. We hypothesized that the quality of reading comprehension and vocabulary instruction can be estimated by the number of times an evidence-based instructional behavior is seen. We were aware that in some areas – for the number of literal questions asked per lesson or the number of interactions centered on activating background knowledge – quantity would not necessarily serve as an estimate of quality. In fact we thought, based on earlier experiences in observational research, that too much time spent on activation activities might have a dampening effect on comprehension. However, for the majority of variables of interest we thought that quantity might serve as a reasonable means to assess quality of comprehension and vocabulary instruction in an objective fashion. These include variables such as number of teacher models of compare-contrast or story grammar elements, providing definitions with multiple examples, and the amount of practice in "finding the gist" of a section of a story. It is therefore logical to assume that the higher frequency of highly valued' research-supported teaching behaviors, the greater will be the quality of reading instruction in that classroom. With this framework in mind we made concerted efforts towards the development of a valid and reliable classroom observation measure that can be used in our large scale randomized control trial.

Purpose/objective/research question/focus of study:

The purpose of our professional development study was to examine the impact of Teacher Study Groups on the teacher practice, teacher knowledge, and student outcomes, specifically in the areas of reading comprehension and vocabulary.

In this presentation we will be discussing the findings of classroom observations that were conducted to assess the *quality* of teaching practice in Reading First classrooms. We will

describe the *Reading Comprehension and Vocabulary (RCV) Observational Measure* (Gersten, Dimino, & Jayanthi, 2007), a moderate inference classroom observational measure that we used to assess the *quality* of teacher instruction, and discuss the dilemmas faced in developing the measure. We will also present validity, reliability, and teacher performance data, and discuss implications for teacher practice, specifically drawing attention to areas where effective instruction was minimal.

Setting:

The multi-site study was conducted in three large urban school districts from three states: California (CA), Pennsylvania (PA), and Virginia (VA). A total of 19 Reading First schools were involved in the study (10 treatment, 9 control).

Population/Participants/Subjects:

Our initial teacher sample included 84 first grade teachers (40 treatment, 44 control); however, three teachers (1 treatment, 2 control) dropped out of the study for a variety of reasons: family problems, illness, and leaving the school district. Our final analytic teacher sample consisted of 81 teachers (39 treatment, 42 control). Our initial student sample included 575 students (273 TSG, 302 control), with mobility issues resulting in a final analytic sample of 468 students (217 TSG, 251 control).

Intervention/Program/Practice:

The TSG intervention was comprised of 16 interactive sessions held at the school site twice a month from October to mid-June. The first eight sessions focused on vocabulary instruction. The remainder of the sessions addressed explicit reading comprehension instruction. Each session lasted approximately 75 minutes. Sessions were conducted at the discretion of the school principal, either before or after school to maximize instructional time during the school day. A 4-step recursive process was instituted during each TSG session: (a) Debrief Previous Application of the Research, (b) Walk Through the Research, (c) Walk Through the Lesson, and (d) Collaborative Planning. This 4-step recursive process provided a common format for the TSG sessions across facilitators and sites, while leaving room for flexibility to respond to issues or concerns specific to the site or individual teacher. Teachers in the control condition participated in scheduled school and district professional development activities. During the study, control teachers did not participate in our TSG sessions or have access to the materials.

We developed the *Reading Comprehension and Vocabulary (RCV) Observational Measure* (Gersten et al., 2007), a moderate-inference measure, to assess the quality of classroom reading instruction. The measure is well aligned with the extant literature on effective reading instruction (e.g., Anderson et al., 1979; Baumann & Kameenui, 1991; Beck, McKeown, & Kucan, 2002; Graves, 2006). The items reflect two major pedagogical aspects of effective instruction: explicitness of instruction and nature of the interactive instruction (i.e., the amount of scaffolding practice and feedback provided) (Ball, 1990; Beck, McKeown, Sandora, Kucan, & Worthy, 1996).

The measure development encompassed a 12-month iterative process of extensive field-testing and on-going refinement of the measure. The items in the measure came from experimental research (as opposed to direct observations of classroom practices), and researchers sometimes use different terms to refer to similar teaching strategies. For both these reasons, creating operational definitions so that each item was mutually exclusive was an arduous and time-intensive task. The measure is presented in Appendix B (Figure 1).

We used a cadre of recently retired teachers or program facilitators as our key observational staff. We fine-tuned the measure on the basis of the input we received from these veterans, and the observational notes collected by our research staff. Through the use of tapes and debriefing on issues raised during live observations, we were able to develop a codebook that defines each variable and provides examples and coding rules. As will be demonstrated, we were able to develop an observation system of adequate reliability with at least some of the richness and nuance that was required to seriously study comprehension and vocabulary instruction.

Research Design:

Randomized field trials were used to examine the impact of the TSG intervention. In Year 1 (2004-2005) the study was conducted in a school district in CA only. In year 2 (2005-2006) the study was replicated in school districts in CA, PA, and VA. Participating schools from each district (for both Years 1 and 2) were randomly assigned to either the TSG condition or the control condition. In the CA school district and PA school district, schools were matched prior to random assignment. In CA, 10 schools (6 schools in Year 1 and 4 schools in Year 2) were matched on AYP scores, ethnic composition (percentage Hispanic), and achievement scores. In the PA school district, 6 schools were matched on the basis of having similar percentage of free/reduced lunch, percentage of students Proficient in reading on the 3rd grade Pennsylvania System of Student Assessment (the statewide test). Schools in VA school district were not matched due to feasibility constraints. The sample in the VA school district included three schools. Two of these were small-sized schools, which were combined into one set, and the set was treated as one school for purposes of random assignment.

Data Collection and Analysis:

To measure teaching practice in reading comprehension and vocabulary, we used the 34-item comprehension scale, and the 12-item vocabulary scale from the *Reading Comprehension and Vocabulary (RCV) Observational Measure* (Gersten et al., 2007). Classroom observations were conducted in each classroom ($n = 81$) during April and early May of Years 1 and 2. All teachers were observed once; 30% of the teachers were observed twice, and 1/8th of the teachers were observed by two observers to collect data for inter-observer reliability. Student assessments were administered over a three-week period in Fall and Spring of Years 1 and 2. All measures were administered individually to the randomly selected students from each class.

Given that our data were of a nested nature (i.e., students and teachers nested within schools), we used hierarchical linear modeling (HLM) to perform the main impact analyses. Since our study of the TSG involved the random assignment of schools to TSG and control conditions, we employed a two-level model to estimate treatment effects on teacher and student

outcomes. Since students and teachers within a school share common experiences, their outcomes are likely to be correlated. Thus, our multi-level models included individual- and group-level error terms to account for the clustering of teachers within schools and students within schools.

Findings/Results:

Data from the *RCV observational measure* (Gersten et al., 2007) indicate that TSGs had a significant impact on teaching practice in vocabulary (E.S. = .58, $p < .01$) and comprehension (E.S. = .86, $p < .01$). Descriptive data on the teaching behaviors measured by the *RCV observational measure* (Gersten et al.) are presented in Tables 1 and 2 in Appendix B. These data (especially from the comparison group) provide a vivid picture of current comprehension and vocabulary instruction in Reading First classrooms. We found some use of evidence-based practices, but consistent use was rare.

Inter-observer reliability was on average 84.49% for the vocabulary scale and 90.89% for the comprehension scale. This is noteworthy given that the *RCV observational measure* (Gersten et al., 2007) is a moderate inference measure. The vocabulary scale has a reliability (Cronbach's alpha) of .70. Internal consistency coefficient for comprehension scale was .69. In many cases, the low base rates for the items in comprehension likely caused several low item to total correlations; yet the composite score demonstrates adequate internal consistency measure (Shadish, Cook, & Campbell, 2002). Criterion related validity *for the vocabulary scale* at the teacher level was .23 with *WDRB Reading Vocabulary* and .20 with *WDRB Oral Vocabulary.* For the comprehension scale it was .24 with *WDRB Reading Vocabulary,* .21 with *WDRB Oral Vocabulary, and .08 for WDRB Passage Comprehension.*

Conclusions:

In the field of educational research and evaluation, advances in measurement have not kept pace with the advances in design and data analysis. One area that is in need of much desired attention is the issue – 'How do we measure quality of teaching practice?" The need for a viable and reliable classroom observational system that is sensitive to the various nuances and fine distinctions that epitomize classroom teaching and that includes clear-cut definitions and objective and standardized procedures cannot be underscored. Our findings indicate that the *RCV observational measure* (Gersten et al., 2007) is sensitive to the intervention, a key component of construct validity. This aspect of validity was particularly important to us in that we designed this measure to be one of several outcomes in a study of professional development for first grade teachers. The measure also has acceptable levels of inter-observer reliability and internal consistency. The *RCV observational measure* (Gersten et al.) presents a viable alternative to existing observational methodology that can be used to assess the quality of teacher practice in large scale up projects to yield consistently data that is objective and reliable.

Appendixes

Appendix A. References

Anderson, L., Evertson, C., & Brophy, J. (1979). An experimental study of effective teaching in first-grade reading groups. *Elementary School Journal, 79*, 193-223.

Ball, D. L. (1990). Reflections and deflections of policy: The case of Carol Turner. *Educational Evaluation and Policy Analysis*, *12*, 247-259.

Baumann, J. F., & Kameenui, E. J. (1991). Research on vocabulary instruction: Ode to Voltaire. In J. Flood, D. Lapp & J. R. Squire (Eds.), *Handbook of research on teaching the English language arts* (pp. 604-632). Upper Saddle River, NJ: Merrill/Prentice Hall.

Beck, I. L., McKeown, M. G., & Kucan, L. (2002). *Brining words to life: Robust vocabulary instruction*. New York, NY: Guilford Press.

Beck, I. L., McKeown, M. G., Sandora, C., Kucan, L., & Worthy, J. (1996). Questioning the author: A yearlong classroom implementation to engage students with text. *Elementary School Journal*, *96*, 385-414.

Edmonds, M., & Briggs, K. L. (2003). *The instructional content emphasis instrument: Observations of reading instruction*. Baltimore, MD: Brookes.

Foorman, B. R., Francis, D. J., Beeler, T., Winikates, D., & Fletcher, J. (1997). Early interventions for children with reading problems: Study designs and preliminary findings. *Learning Disabilities: A Multi-Disciplinary Journal, 8*, 63-71.

Foorman, B., & Schatschneider, C. (2003). Measurement of teaching practices during reading/language arts instruction and its relation to student achievement. In S. R. Vaughn & K. L. Briggs (Eds.), *Reading in the classroom: Systems for the observation of teaching and learning*. Baltimore, MD: Brookes.

Gersten, R., Baker, S., Haager, D., & Graves, A. (2005). Exploring the role of teacher quality in predicting reading outcomes for first grade English learners: An observational study. *Remedial & Special Education*, *26*, 197-206.

Gersten, R., Carnine, D., Zoref, L., & Cronin, D. (1986). A multifaceted study of change in seven inner city schools. *Elementary School Journal, 86*, 257-276.

Gersten, R., Dimino, J., & Jayanthi, M. (2007). Towards the development of a nuanced classroom observational system for studying comprehension and vocabulary instruction. In B. M. Taylor & J. E. Ysseldyke (Eds.), *Effective Instruction for Struggling Readers, K-6* (pp. 196-215). New York, NY: Teachers College Press.

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M. D., Greenwood, C., & Innocenti, M. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, *71*, 149-164.

Graves, M. F. (2006). *The vocabulary book: Learning & instruction*. New York, NY: Teachers College Press.

Haager, D., Gersten, R., Baker, S., & Graves, A. (2003). The English-language learner classroom observation instrument: Observations of beginning reading instruction in urban schools. In S. R. Vaughn & K. L. Briggs (Eds.), *Reading in the classroom: Systems for observing teaching and learning*. Baltimore, MD: Paul Brookes Publishing.

Kennedy, M. M. (1999). Approximations to indicators of student outcomes. *Educational Evaluation and Policy Analysis, 21*, 345-363.

McIntosh, R., Vaughn, S., Schumm, J. S., Haager, D., & Okhee, L. (1994). Observations of students with learning disabilities in general education classes. *Exceptional Children*, *60*, 249-261.

Schatschneider, C., Fletcher, J., Francis, D. J., Carlson, C. D., & Foorman, B. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology*, *96*, 265-282.

Shadish, W., R., Cook, T., D., & Campbell, D., T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton Mifflin Company.

Stallings, J., & Kaskowitz, D. (1974). *Follow through classroom observation evaluation 1972-1973* (SRI Project URU-7370). Stanford, CA: Stanford Research Institute.

Stoolmiller, M., Eddy, J. M., & Reid, J. B. (2000). Detecting and describing preventative intervention effects in a universal school-based randomized trial targeting delinquent and violent behavior. *Journal of Consulting and Clinical Psychology*, *68*, 296-306.

Appendix B. Tables and Figures

Table 1

Descriptive Data on the Comprehension Items of the *RCV Observational Measure*

ITEMS	Mean Frequency of Behavior			
	Experimental N = 39		Control N = 42	
	Mean	SD	Mean	SD
Preparatory activities	2.30	2.68	2.33	3.14
Text cues to interpret text	0.70	1.09	0.17	0.38
Visualize events, clarify, re-read	0.63	1.36	0.51	0.97
Evaluate predictions	0.08	0.33	0.00	0.00
Generate questions about text	0.40	1.28	0.11	0.34
Make text-to-text connections	0.09	0.32	0.05	0.18
Make inferences, summarize/find main ideas	0.82	1.41	0.59	1.02
Retell, sequencing	0.60	1.44	0.22	0.51
Story grammar elements	0.18	0.37	0.05	0.21
Compare-contrast or cause-effect text structure	0.04	0.13	0.08	0.24
Reiterates or reinforces concepts that highlight the meaning of text	1.99	2.69	1.43	2.05
Preparatory activities	6.99	6.78	6.49	7.90
Text cues to interpret text	1.55	2.63	1.25	2.23
Visualize events, clarify, re-read	1.49	2.81	0.25	0.73
Evaluate predictions	0.56	1.90	0.19	0.97
Generate questions about text	0.18	0.67	0.17	0.69
Make text-to-text connections	0.22	0.53	0.14	0.40
Summarize/find main ideas	1.02	1.51	0.32	0.56
Retell, Sequencing	1.36	1.44	0.56	0.72
Story grammar elements	0.49	0.84	0.20	0.54
Compare-contrast or cause-effect text structure	0.18	0.49	0.18	0.54
Asks students to answer literal recall questions from the text	16.24	11.03	8.17	8.15
Asks students questions requiring inferences based on text	12.15	8.58	6.49	6.69
Asks students to justify or elaborate their responses	1.84	2.26	0.67	1.16
Teacher keeps students thinking for 2+ seconds before calling on a student for response	1.72	2.97	0.53	0.73
Teacher gives independent practice in answering comprehension questions or applying comprehension strategy(ies) with expected product	0.81	2.25	0.19	0.45
Communicates clearly what student/s did correctly about the strategy	0.11	0.31	0.09	0.48
Reinstructs when student makes a mistake by encouraging child to try again or reminding student about comprehension strategy	0.24	0.74	0.02	0.15
Uses a graphic organizer before, during, or after lesson	0.76	1.01	0.36	0.79

Table 2

Descriptive Data on the Vocabulary Items of the *RCV Observational Measure*

ITEMS	Mean frequency of behavior			
	Experimental N = 39		Control N = 42	
	Mean	SD	Mean	SD
Provides an explanation, a definition, and/or an example.	5.17	3.96	3.62	3.32
Elaborates using multiple examples.	0.88	1.08	0.47	0.78
Elaborates using contrasting example(s) to pinpoint definition.	0.95	2.11	0.13	0.35
Uses visuals, gestures, facial expressions, pictures, or demonstrations to teach word meanings.	4.82	8.31	2.05	2.33
Asks students to answer questions or participate in activities that require knowledge of words.	17.51	15.20	8.45	6.44
Gives students opportunity to apply word learning strategies - using context clues, word parts, root meaning.	0.67	1.42	0.68	1.67
Teacher further pinpoints the definition by extending or elaborating students' responses.	1.85	1.85	0.56	0.80

Figure 1

RCV Observational Measure

Comprehension - 1st Interval

A. Explicitness of Instruction	Tally		Total (Max 15)	Notes
Prior to reading, teacher				
1. Conducts preparatory activities: relating text to student experiences/previous readings, background knowledge, discussing pictures, title/author, browsing (book cover, spine, TOC), predicting story content.				
During or after reading, teacher				
2. Models the use of following (includes think-alouds)				
a. Text cues to interpret text: pictures, sub-headings, captions, graphics				
b. Visualize events, clarify, re-read.				
c. Evaluate predictions				
d. Generate questions about text				
e. Make text-to-text connections				
f. Make inferences, summarize/find main ideas- theme, character analysis				
g. Retell, sequencing – what's happening, what happened first				
h. Story grammar elements - except for theme, character analysis	N	Y		
i. Compare-contrast or cause-effect text structure	N	Y		
3. Reiterates or reinforces concepts that highlight the meaning of text.				

B. Student Practice	Tally		Total (Max 15)	Notes
Prior to reading, teacher				
1. Gives students practice in preparatory activities: relating text to student experiences, previous readings, background knowledge, discussing pictures, title/author, browsing (book cover, spine, TOC); predicting story content.				
During or after reading, teacher				
2. Gives students practice in the following:				
a. Text cues to interpret text - pictures, sub-headings, captions, graphics				
b. Visualize events, clarify, re-read				
c. Evaluate predictions				
d. Generate questions about text				
e. Make text-to-text connections	N	Y		
f. Summarize/ find main ideas	N	Y		
g. Retell, Sequencing –what's happening, what happened first	N	Y		
h. Story grammar elements	N	Y		
i. Compare-contrast or cause-effect text structure	N	Y		
3. Asks students to answer literal recall questions from the text. (Specific questions).				
4. Asks students questions requiring inferences based on text.				
5. Asks students to justify or elaborate their responses.				
6. Teacher keeps students thinking for 2+ seconds before calling on a student for response.				
7. Teacher gives independent practice in answering comprehension questions or applying comprehension strategy (ies) with expected product.	N	Y		

C. Corrective feedback: Teacher	Tally		Total (Max 15)	Notes
Communicates clearly what student/s did correctly about the strategy.				
Reinstructs when student makes a mistake by encouraging child to try again or reminding student about comprehension strategy.				
D. Uses a graphic organizer before, during, or after lesson _____	N	Y		

Vocabulary – 1st Interval

A. Explicitness of Instruction	Tally	Total (Max 15)	Notes
The teacher			
Provides an explanation, a definition, and/or an example.			
Elaborates using multiple examples.			
Elaborates using contrasting example(s) to pinpoint definition.			
Uses visuals, gestures, facial expressions, pictures, or demonstrations to teach word meanings. (gestures are related to word meaning)			

B. Student Practice	Tally	Total (Max 15)	Notes
The teacher			
Asks students to answer questions or participate in activities that require knowledge of words. – e.g., define words; make sentences; find words based on clues; show me how you would look if you were cross; raise your hand if I say something that is enormous.			
Gives students opportunity to apply word learning strategies - using context clues, word parts, root meaning.			
Teacher further pinpoints the definition by extending or elaborating students' responses.			

Post-Observation Component

Answer the following questions at the end of your observation:

A. During comprehension instruction,

Teacher gave inaccurate and/or confusing explanations while modeling strategies.	N Y	
Teacher missed opportunity to correct or address error, or provided confusing or inaccurate feedback.	N Y	
Teacher called individually on about half or more of students.	N Y	

B. During vocabulary instruction,

Teacher gave definition, explanation, and/or example that is inaccurate and/or confusing	N Y	
Teacher missed opportunity to correct or address error, or provided confusing or inaccurate feedback.	N Y	
Teacher called individually on about half or more of students.	N Y	

C. Based on your overall judgment, how would you rate the quality of each domain you observed?

	Not Observed	Minimal/Erratic	Partially Effective	Good	Excellent
Comprehension	N/O	1	2	3	4
Vocabulary	N/O	1	2	3	4

D. Please the rate management/responsiveness to students** items on the following 4-point scale.

	Minimal/Poor	Fair	Good	Excellent
The instructional routines appear to be	1	2	3	4
The teacher maximizes the amount of time available for instruction (e.g., brief transitions)	1	2	3	4
The teacher manages student behavior effectively in order to avoid disruptions and to provide productive learning environments	1	2	3	4

** Items are adapted from Teacher Competency Checklist (Foorman & Schatschneider, 2003).

E. How would you rate student engagement today?

				Scale:
Students are engaged during the **first 45 minutes** of the reading block	1	2	3	1=Few students seem engaged during the lesson. 2=Many students seem engaged much of the time. 3=Most students are engaged all of the time.
Students are engaged during the remainder of the reading block.	1	2	3	