TECHNICAL REPORT #31:

Teacher Use Study: Reading Aloud vs. Maze Selection

Christine Espin, Stan Deno, Kristen McMaster, Rebecca Pierce, and Seungsoo Yeo

RIPM Year 5: 2007 – 2008

Dates of Study: October 2007 – May 2008

September 2009

Produced by the Research Institute on Progress Monitoring (RIPM) (Grant # H324H30003) awarded to the Institute on Community Integration (UCEDD) in collaboration with the Department of Educational Psychology, College of Education and Human Development, at the University of Minnesota, by the Office of Special Education Programs. See progressmonitoring.net.

Purpose

This research focused on factors that might influence teachers' use of progress monitoring data in their design of instructional programs. Specifically, the study investigated the effects of using reading aloud vs. maze selection progress measures on teachers' instruction and student learning. We hypothesized that the type of measure would affect the focus of teachers' instruction, with maze selection measures leading to a greater focus on reading comprehension and reading aloud measures leading to a greater focus on phonics and fluency. We further hypothesized that these differences in instruction would affect student performance, with students monitored with maze selection performing better in reading comprehension, and students monitored with reading aloud performing better in phonics and fluency. The research question addressed in the study was, "What are the effects of using different types of progress measures (reading aloud vs. maze selection) on teacher instruction and student learning?"

Method

Setting and Participants

The study took place in a large urban district in a midwestern state. The district had 35,757 students in grades K-12: 39.7% African American, 29.6% White American, 17.2% Hispanic American, 8.9% Asian American, and 4.6% Native American. For 23.2% of the students, English was spoken as a second language. The percentage of students on free and reduced lunch was 72% for K-8 schools and 60% for K-6. Approximately half of the students (51%) were male and 15.9% received special education services.

Teacher participants. Participants in the study were 18 special education teachers and 66 special education students from the classes of participating teachers. Participating teachers had an average of 18.86 years (range 2.5 to 27 years) teaching experience, with an average of 11.52

years in special education (range 2.5 to 20 years). All teachers were licensed in special education. Licensures included E/BD only (*n* = 2), LD only (*n* = 1), E/BD and LD (*n* = 2), LD and elementary education (*n* = 2), E/BD and elementary education (*n* = 2), E/BD and secondary math (*n* = 1), E/BD and DCD (*n* = 1), E/BD, LD, and elementary education (*n* = 2), LD and DCD (*n* = 1), DCD and elementary education (*n* = 1), E/BD, LD and DCD (*n* = 1), and E/BD, LD, DCD, and elementary education (*n* = 2). Nine teachers held master's degree and one a Reading Specialist certification. Teachers had used CBM progress monitoring for an average of 12.83 years (range 2 to 25). All had received CBM training through district training sessions; in addition 10 participants had received additional training through a university class, 10 had attended conferences or workshops, and 6 had used self-study approaches. Self-study was an option included for teachers who may have had no specific training in CBM, but who had done some independent reading to learn about the measurement procedure. Of the six teachers who marked this option, five had also marked one or all of the more direct training choices (i.e., district training, classes, or workshops). The remaining teacher had also marked learning while attending a conference.

Ages of the teachers were 20 to 29 (*n* = 1), 30 to 39 (*n* = 1), 40 to 49 (*n* = 8), 50 to 59 (*n* = 6), and over 60 (*n* = 2). Fifteen teachers were White, two were Hispanic American and one was African American. Seventeen teachers were female. Nine of the teacher participants were recruited from a district professional development class that incorporated sharing CBM data via an online website. The other nine were recruited through special education lead teacher meetings and one-on-one discussions.

Student participants. The teachers invited all eligible children in their caseload to be participants. To be eligible for this study, a student needed to be receiving special assistance in reading, be able to read at least 10 words of connected text in 1 min, and be likely to remain on the teacher's caseload for the year. Permission for participation was obtained for 99 students.

The study made use of a within-teacher design; thus, students were first paired based on performance levels, and then one of each pair was randomly assigned to a condition. Students were ranked ordered within teacher according to mean score on the pretest reading aloud probes. Students with similar mean scores were then paired. If more than two students' scores were close, we looked at the difference between correct maze choices (CMC) and incorrect maze choices (IMC); the two participants with the closest difference score became the matched pair. This occurred twice during the pairing process. Each teacher monitored one, two, or three matched pairs of students. All students who were not paired were placed in a no-treatment group. This larger sample was used for some descriptive analyses, for example, calculating correlations between CBM and achievement test measures.

A total of 41 pairs were created (n = 82 students). After matched pairs were created, a list of random numbers was used to assign the partners to a condition. If the number was even the first student was assigned to the maze selection condition and the partner to the reading aloud condition. If the number was odd, the first student was assigned to the reading aloud condition and the partner to the maze selection condition. Once the paired students had been assigned to the conditions, a t-test was used to analyze the mean reading aloud pretest score for both conditions. There was no statistically significant difference between the two groups (t = .70, df = 80, p = .48).

Attrition (from conditions) of 11 students occurred during the study; 10 students moved and 1 was chronically absent from the resource room. If a student dropped from the study, his or her partner was also dropped (5 additional students). The final sample, thus, consisted of 66 students (33 in the maze selection group and 33 in the reading aloud group). Of this final sample, 64 students received special education services and 2 received small group supplemental instruction in reading (tier 2 in a response-to-intervention model). The special education services varied in federal settings (resource room settings to day treatment settings) and type (learning disabilites, developmental cognitive disabilities, and emotional/behavioral disabilities), but all students were considered high-incidence. (Note that if no-treatment students are considered, sample size was 87). See Tables 1 and 2 for specific demographic information.

Table 1
Student Demographics

	Total with permission	Participating students before attrition	Participating students after attrition
Gender			
Male	68	56	44
Female	31	26	22
Grade			
1st	1	1	1
2nd	9	7	5
3rd	28	19	15
4th	19	17	12
5th	22	20	18
6th	7	7	5
7th	8	6	5
8th	5	5	5
SES			
Free lunch	78	64	51
Reduced lunch	11	10	10
Normal rate	5	2	3
No information	5	5	2
Ethnicity			
American Indian	3	3	3
Asian American	7	5	4
African American	52	44	34
Hispanic American	19	17	15
White American	18	13	10

Disability			
Special education	91	78	64
Tier 2	8	4	2
ELL	24	20	19
Attrition	12	11	
Paired Partner Attrition		5	
Total Participants	99	82	66

Note. There were 82 students initially: 66 plus 11 who moved and 5 matched partners. One of the students who moved was in the 'no treatment' groups.

Table 2
Student Demographics by Condition (n = 66)

	Maze Selection	Reading Aloud
Gender		
Male	22	22
Female	11	11
Grade		
1st	1	0
2nd	3	2
3rd	8	7
4th	5	7
5th	9	9
6th	2	3
7th	1	4
8th	4	1
SES		
Free lunch	25	26
Reduced lunch	5	5
Normal rate	1	2
No information	2	0
Ethnicity		
American Indian	1	2
Asian American	2	2
African American	15	19
Hispanic American	9	6
White American	6	4
Disability		
Special education	32	32
Tier 2	1	1
ELL	12	7
Total Participants	33	33

Independent Variables

The independent variable in the study was the type of measure used to monitor student progress in reading: reading aloud vs. maze selection.

Dependent Variables

Dependent variables were teacher instructional focus and student reading performance. The focus of teacher instruction was determined via analysis of case discussions between pairs of teachers. Student reading performance was assessed using CBM pre- and posttest reading aloud and maze selection measures and a standardized achievement test administered at posttest only.

Focus of Teacher Instruction

Online discussions. The focus of teacher instruction was ascertained via analysis of weekly online discussions that teachers held with each other. Once a week, teachers communicated with a partner on a password-protected Moodle (a course management system) website furnished by the professional development department of the local school district. See Appendix A for sample illustrations of a Moodle forum. After logging into this site, teachers entered a forum in which they posted a student's progress-monitoring graph and an accompanying description of a student's progress. Teachers were provided with the following suggestions for the type of information to include in their postings: grade level, classwide screening CBM data (results of tri-annual school-wide screening at grade level), level of material in which the student was monitored and the reasons for choosing that level, the goal selected for the student in terms of expected growth per week, an explanation of the instruction provided to the student that included the focus of the intervention, the number of students in the group, the amount of instructional time, a description of the student's response to the intervention, changes made in the instructional program or in the goal, and a rationale for those changes. After posting,

the teachers would reply to their partner's post. Teachers were provided with the following reflection questions to guide their responses:

- Is there any additional information that would be helpful? Any graphing information that is needed?
- How do you see this student responding to the intervention?
- What do you see going on from this graph?
- If a change in instruction is needed, please provide specific instructional changes that may help this student achieve his/her goal, by using the following categories: change in materials, change in motivation, change in time or setting, change in teacher/student ratio, and change in instructional approach.

Coding the online discussions. Two major categories of codes were used to capture teachers' focus of instruction: reading instruction and other instruction. Reading instruction codes addressed teachers' comments related to interventions focused in improving students' reading, and were the major focus of the study. However, in the process of coding the data, it became apparent that teachers made many "other" types of instructional comments that were not directly focused on reading components. These comments focused on general instructional strategies, for example, ways to improve student motivation, using specific types of materials, or changing the measurement procedures.

The initial set of codes used for the ***reading instruction*** category were based on the five components of reading instruction identified by the National Reading Panel (NRP, 2000): phonemic awareness, phonics, fluency, vocabulary, and comprehension. Modifications were made as coding progressed. Due to the small number of comments about phonemic awareness

and the fact that teachers did not seem to discriminate between phonemic awareness and phonics, the two categories were collapsed into a single phonics category. In addition, a separate category for word recognition was added. Codes were defined as follows:

Phonics was defined as the use of letter-sound correspondence (in isolation or in combination) to read a word, and included references to phonemic awareness, using phonetic rules to decode, reading word parts, and chunking word components together (e.g., “He is working on long vowel sounds.”).

Word recognition was defined as building automaticity in word reading or word list practice (e.g., “I work on sight words in two places: in their own writing and in stories that they are reading.”), and included references to common lists of sight words (e.g., Dolch), repeated exposure to words, or proper names.

Fluency was defined as reading text in a smooth and effortless manner, and included references to repeated readings and chunking groups of words in phrases (e.g., “Have you considered taping him and then letting him hear himself read at the fast pace and then the slow pace?”). Practice reading connected text was also included under fluency, and included references to reading at home, reading with a partner, and reading silently.

Vocabulary was defined as understanding word meanings, and included references to word study, finding definitions, and interpreting the meaning of a word (e.g., “She has vocabulary issues that seem to interfere with her comprehension.”).

Comprehension was defined as understanding of meaning of text, and included references to the use of context clues, answering and generating questions, summarizing, and recalling ideas (e.g., “He can make inferences most of the time.”).

Codes for the ***other instruction*** category included: materials, motivation, and measurement procedure change. Codes were defined as follows:

Materials was defined the use of general materials or resources to teach reading (e.g., "I am thinking of changing to a 4th grade book."), and included general references to the use of books, workbooks, etc. Reference to specific curricula were listed separately and not included in this category (see Table 5).

Motivation was defined as student interest (or noninterest) in reading or in the material used for teaching reading (e.g., "Wouldn't it be wonderful if they didn't get mad and were motivated to read?").

Changing measurement procedures was defined as changing the method used for CBM progress monitoring as a way to effect improvements in student performance (e.g., "Perhaps you can show the student on the passage where you want the student to end up as an intervention. It seems to motivate my students.").

Once statements were coded into one of the instructional categories listed, the statements were classified as describing either a student characteristics (e.g., "He's been putting more effort into his work.") or an instructional intervention (e.g., "For motivation I have placed him in reading groups with people he likes."). *Student characteristics* were statements that provided static information about a child, such as what a child could or could not do and what a child did or did not know. *Instructional intervention* were statements referring to what the teacher did or was planning to do to improve the child's reading skills. Comments that directly linked student characteristics to instruction were coded as instructional interventions.

Coding procedures. A research assistant organized the archived online discussions by teacher. Each protocol contained all postings and related responses for a teacher's students –

from both conditions, reading aloud vs. maze selection; however, coders were blind to the condition during the coding process. Within each protocol, each student's postings and responses were chronicled by date. The protocols were imported as sources into NVivo (QSR International, 2008). "Source" is the term NVivo uses to refer to imported documents.

For each protocol, a primary researcher and a research assistant independently identified units of information or *statements* and then compared results until agreement was reached on parsing of statements. A statement was identified as a single thought unit related to reading instruction or instruction in general. Statements that referred to specific reading curricula or typical weekly progress monitoring behaviors or measures were not identified for coding.

After protocols were parsed into statements, statements were assigned a code. The researcher and research assistant first coded six teacher protocols together. They then independently coded two randomly selected protocols and calculated point-by-point agreement (total agreement/total number of statements coded). Each statement was coded first using one of the reading / other categories, then as a student characteristic or an instructional intervention. To be considered an agreement, both sets of codes had to be the same. Agreement was 93% and 98%. Coders then worked independently to code remaining protocols.

Surveys. To gather information about instructional changes teachers made as a result of progress monitoring, we asked all teacher participants to complete a researcher-created survey at the end of the study. For detailed information and findings about the surveys, see Tech Report #34—Urban Survey report.

Student Reading Performance

CBM probes. Two CBM reading aloud and two maze selection probes were administered to students at pre- and posttest. All participating students completed the same pre- and posttest

passages. The reading aloud probes were selected from Peabody CBM Reading Passages & Word Counts (Vanderbilt University). The maze selection passages were selected from Project PROACT MAZE Reading Passages (Vanderbilt University). To select passages, a research assistant reviewed all passages and selected five passages that were not culturally-dependent (e.g., based on a holiday, tradition, or folktale), would not likely be familiar to students (e.g., based on a fairy tale), and were longer than 350 words. The primary investigators made the final choices. Selected passages were at a second-grade reading level.

Standardized reading test. Two subtests from the Kaufman Tests of Educational Achievement – Revision II (KTEA-II; Kaufman & Kaufman, 2004) were administered at posttest: Letter-Word Identification and Reading Comprehension. The Letter-Word Identification subtest requires students to read lists of isolated letters and words, starting at an item corresponding to the student's grade-level and continuing until four consecutive errors have been made. The first section assesses the student's knowledge of letter names and sounds. The word list begins with high-frequency sight words and phonetically regular words, but progressively introduces more difficult words contain unpredictable pronunciations. This subtest is intended to measure a student's reading vocabulary (i.e., words familiar enough for automaticity (Kaufman & Kaufman, 2004).

The Reading Comprehension reflects the student's ability to gain meaning from text (Kaufman & Kaufman, 2004). The questions become progressively more difficult, beginning with items that require students to match a word with a picture. The next level of difficulty requires students to read and perform a simple command. Most of the items entail asking a student to read a passage and answer its accompanying questions orally. Although a few questions provide choices, most questions are open-ended and are meant to elicit single-word or

short-phrase answers. Students begin at a level determined by their raw score on the Letter-Word Identification subtest and continue until they make four errors in a set or 5 consecutive errors. The published split-half reliability coefficient of the KTEA is .97 for the Letter-Word Identification subtest, .93 for the Reading Comprehension subtest, and .97 for the Reading Composite. Correlations between the two subtests are reported to be .68, and correlations with other standardized reading tests (WIAT-II, WJ-III, PIAT-R/NU) range from .76 to .85 (Kaufman & Kaufman).

Table 3 presents intercorrelations between the reading aloud, maze selection, and Kaufman subtests for the student participants in our study. (Note that the correlations are calculated from the larger sample of students who had permission to participate.)

Table 3
Intercorrelations Among Outcome Variables

	Reading aloud (pretest mean)	Reading aloud (post mean)	Maze (pretest mean)	Maze (post mean)	KTEA-II Letter-Word Identification*	KTEA-II Reading Comprehension*
Reading aloud (pretest mean)	1/00	.94	.72	.72	.86	.76
Reading aloud (post mean)		1.00	.72	.76	.86	.78
Maze (pretest mean)			1.00	.77	.76	.74
Maze(post mean)				1.00	.71	.76
KTEA-II Letter-Word Identification					1.00	.85
KTEA-II reading comprehension						1.00

Note. $N = 87$.
*Raw scores used for correlations with KTEA subtests.

Procedures

We first discuss components of the study that examined the focus of teacher instruction, including the progress monitoring passages used by teachers, the monitoring procedures teachers followed, the fidelity of those procedures, and the online setting in which teachers discussed

progress monitoring results. Next, we discuss the administration and scoring of measures used to determine student reading performance.

Focus of Teacher Instruction

Weekly Progress monitoring passages. For weekly progress monitoring, teachers used district-provided reading aloud probes. Within each grade level (grades 2 to 6), 30 passages were selected from the *Houghton Mifflin Invitations to Literacy* reading curriculum published in 1998, and were adopted by the district to be used for progress monitoring. Thirty grade 1 passages were selected from *Houghton Mifflin Legacy of Literacy*, 2001. The passages reflect a variety of subjects, authors, and styles found in each grade level anthology. The lengths of the passages varied by grade level, ranging from around 100 words in the first grade set to around 300 words in the sixth grade set.

Maze selection measures were generated by the research team from the Houghton-Mifflin reading aloud passages. The maze selection probes were created using the rules outlined by Fuchs and Fuchs (1992). The first sentence of each passage was left intact, after which every seventh word was deleted and replaced with a 3-word choice: the correct choice and two distracters. Distracters were within one letter in length of the correct choice, did not make contextual sense, did not rhyme with and were not close in sound or letter configuration to the correct choice, were real words, did not require the students to read more than 1.5 lines ahead in the passage to eliminate a choice, and were not so high in vocabulary that the student might mistake the distracter for a nonsense word. If a proper noun or a hyphenated word was the seventh word, the eighth word was replaced with the 3-word choice. If the eighth word was also a proper noun, the sixth word was replaced. After being formatted, each passage was independently edited by two other team members to ensure compliance with rules.

We asked teachers to give feedback about the technical format of the maze selection probes created for this study. Specifically they were asked about the format of the probes and if they observed any problem that resulted from converting the district's reading aloud passages into maze selection probes. See Appendix B for resulting teacher comments.

Weekly progress monitoring. All teachers administered CBM measures once per week to participating students and used graphs of student progress to make instructional decisions. Students in the reading aloud group were monitored weekly using reading aloud measures; students in the maze selection group were monitored weekly using maze selection measures. After scoring the measures (see description below), teachers entered the results into a district-owned website which contains progress monitoring data for all students. This website is only accessible by school district personnel; researchers could not independently access the information. After the data were entered, they were displayed in a graph format.

For both conditions, teachers followed a guideline for administering and scoring each probe. When administering the 1 min reading aloud probe the teachers were instructed to supply the word for the student after 3 seconds, supply an unusual name if a student struggled with it and count the word as correct, count a self-correct as correct, and not to penalize a child for dialect. Both total words (TW) and errors (IW) were counted. A word was counted as an error if it was mispronounced, omitted, or provided for the student after 3 seconds. Errors were subtracted from total words to determine words read correctly (WRC: TW – IW). Both WRC and IW were recorded on the graphs.

When scoring a 2 min maze selection probe, the teacher marked an X over any incorrect choice the student made. When two adjacent errors occurred, the teacher stopped scoring, made a vertical line after the last correct answer, and scored correct (CMC) and incorrect maze choices

(IMC) up to that point. A maze selection choice was counted as incorrect if the wrong item was circled, no choice was made, more than one answer was circled, or if the student's choice could not be determined. Recorded scores, CMC and IMC, were entered into the district's protected progress-monitoring website, which produced a graph of the student's data. For this study, teachers monitored student progress for 22 weeks.

Observations. We divided the study duration into three rotations of six to seven weeks. The first rotation was from November 12 to December 21, the second from January 7 to February 24, and the third from February 25 to April 18. At the end of the first rotation and the beginning of the third rotation, teachers were observed as they administered weekly CBM measures to at least one student from each condition. They were observed by research assistants who used a procedural checklist that contained the required administration and scoring actions. Step completion for all teachers averaged 95% with a range of 80% to 100%. Additionally, scoring agreement between the teacher and the observer was calculated. Scoring agreement averaged 99% with a range of 83% to 100%.

After the second observation, teachers were asked what intervention changes had been made since the beginning of the study. Before they answered, the teachers were asked to consider the following factors: instructional activities, amount of instructional time, the setting, group size, motivational strategies, and materials. Fourteen teachers responded to this question. Four either stated no comment or made no changes, five made group changes, and five made individual programming changes. Within the individual changes, there were two setting changes, five additional instructional activities added, and one increase of time spent on an instructional activity.

Online discussions. To examine how the different progress monitoring measures influenced the teachers' perspectives about instruction, we asked teachers to discuss students' current and future instructional plans with a peer using a discussion format within a secure, within-district website. The 18 teachers were divided into grade levels (primary, intermediate, and upper) and were assigned a partner randomly for the first rotation. New dyads were formed for the second and third rotations. Teacher pairing for these rotations was not random to avoid duplicating dyads; however, during the third rotation, one dyad was the same as during the first rotation.

Once a week, each participating teacher was asked to post information (including the current CBM graph) about one student and respond to the posted information about the partner's student. The teachers selected the order of postings so they could give priority to students about whom they desired advice. The paired students per teacher varied from two to six students. During each rotation, teachers with six students were asked to post information about each student; those with four students were asked to post each student once, and then select two students to post a second time; and those with two students were asked to post each student three times.

Frequency of online interaction. Although we asked participants to post and respond at least six times during each rotation (total of 18 possible for all rotations), the number of actual postings per teacher ranged from 4 to 56 and the number of responses ranged from 1 to 37. Throughout the study, there were a total of 271 postings and 183 responses (145 replies and 38 responses to replies). At least one response was given for 124 of the postings. Figure 1 shows the postings and postings responded to by condition within each of the three rotations. Finally, a total of 35,769 words were written (range per teacher = 350 to 5,294); 17,639 for students in the

reading aloud group (range per teacher = 113 to 2,731), 18,007 for students in the maze selection group (range per teacher = 101 to 2,808), 123 not related to any student.

Figure 1. Postings and postings responded to by rotations.

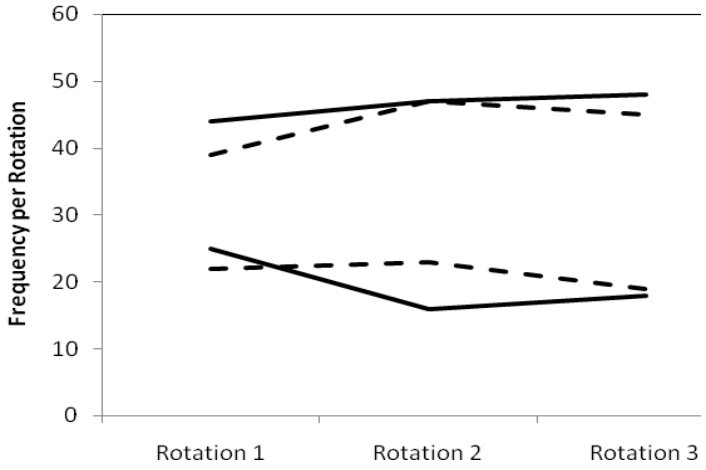


Figure 1. Maze selection = solid line; Reading aloud = dotted line. The top two lines represent postings. The bottom two lines represent postings to which there was a response.

Student Reading Performance

Administration of pretests and posttests. All CBM measures were administered by research assistants within the students' schools. Students within a special education classroom were assessed during one time setting. Students who were absent were tested individually within the next week. First, two maze selection probes (2 min each) were group administered to all participants from the classroom. Next, 2 reading aloud probes (each 1 min in duration) were individually administered. During only the posttest sessions, the KTEA-II Letter-Word Identification and the Reading Comprehension subtests (in that order) were individually administered after the CBM probes. All examiners followed the same script and procedure (See Appendix C). There were approximately six months of instruction between testing sessions.

Scoring fidelity. Two research assistants independently scored all maze selection probes. A word selection was scored as incorrect if it was not the correct answer, it was skipped, more than one selection was circled, or the selection could not be determined. Scoring included the number of CMC and IMC preceding two adjacent errors. Any differences were discussed until both parties were in agreement. For data analysis, we used CMC – IMC scores.

Reading aloud probes were scored for total words (TW) and errors (IW) during administration. A word was scored as incorrect if it was mispronounced, if it was omitted, or it was supplied after 3 seconds. Differences in dialect or speech patterns were not counted as errors. Test administrators tape-recorded 12% of their probes. Each passage was scored twice by data collectors to ensure accuracy. In addition, agreement was calculated for 12% of the passages which were audio taped. Another research assistant independently scored the taped passages. Interscorer agreement, which was calculated by dividing the total agreements by the sum of agreements plus disagreements, was 95%. Words read correctly (WRC: TW – IW) was used for data analysis.

Working independently, two research assistants determined the raw scores, standard scores, and percentiles for the KTEA Letter-Word Identification subtest, Reading Comprehension subtest, and reading composite. Any differences were resolved before the data were entered into the Excel file. Data were entered separately by two research assistants, and then the files were compared to ensure accuracy of data entry. Because all students took the same form of the KTEA, raw scores were entered into the analysis. Our sample was a cross-grade sample of poor readers. Standard scores would reduce variability in scores because they would be based on age-level norms. Thus, a both a 3rd-grade and 6th-grade student might have standard scores of 90, but have different raw scores and different levels of reading performance.

Correlations were calculated between reading aloud and scores on the KTEA for a large group of students (students from this study and a parallel study conducted in a different district (see Tech Report #32). Correlations between scores on the reading aloud probes and raw scores for the KTEA were .82 and .78 for letter-identification and comprehension respectively. Note that correlations with standard scores were .40 and .34 for letter-identification and comprehension respectively. Correlations for maze were .74 and .80 for KTEA raw scores, and .24 and .25 with KTEA standard scores.

Results

In this study we examined what effects reading aloud and maze selection measures have on a teacher's instructional focus and on student reading performance. First, we discuss the differential effects on the focus of teacher instruction, and then, the effects of student performance.

Focus of Teacher Instruction

The research team analyzed the online discussions that occurred within teacher dyads to examine what effects the type of CBM reading measure (reading aloud vs. maze selection) had on teacher instruction. Our hypothesis was that when teachers used the reading aloud measure, their attention would focus on word-oriented categories, such as phonics, word recognition, and fluency. Conversely, we hypothesized that when teachers used maze selection measures, their attention would focus on meaning-oriented categories such as vocabulary and comprehension.

Using NVivo, students' postings and accompanying responses were divided by condition, reading aloud or maze selection. Postings and responses were also divided by rotations. Using Matrix Queries, we determined the number of statements made in each category by condition, and the number of statements per rotation by condition. With this information, we determined the

percentage of statements made for each category by condition and across conditions (total number of coded statements).

Overall content of teacher exchanges. In the initial analysis, we examined the overall content of teacher exhanges. Teachers made 378 (55%) statements about students in the reading aloud condition, and 309 (45%) about students in the maze selection condition. For both conditions, the percentage of statements falling in the *reading-related* category was greater (62.7% and 59.55% for reading aloud and maze selection respectively) than the percentage falling in the *other* category (37.3% and 40.45 % for reading aloud and maze selection respectively). For both conditions, the percentage of comments made about student characteristics was approximately the same as those made about instructional interventions (47% vs. 53% respectively for both reading aloud and maze selection. See Appendix D for specifics.)

Our main focus were the comments realated to reading instruction; however, it was interesting to note that there were group-related differences in statements made within the materials, motivation, and measurement procedure changes categories. Specifically, in the reading aloud condition, a relatively higher percentage of statements were made related to motivation (19.84%) and materials (5.29%) than in the maze selection condition (11.65% and 2.27%, respectively). In contrast, in the maze selection condition, a large percentage (26.54%) of comments were made related to measurement changes, a pattern not seen in the reading aloud condition (12.17%).

Content of comments focused on reading instruction. Our main focus, and our hypotheses, related to the reading categories. In Table 4, the numbers and percentages of comments broken down by category and condition are reported. Of note, when monitoroing with reading aloud measures, teachers made a higher percentage of statements focused on word

recognition (17.72%) and fluency (32.49%) than when monitoring with maze selection (8.70% and 16.30% for word recognition and fluency respectively). In contrast, when monitoring with the maze selection measures, teachers made a higher percentage of statements focused on comprehension (41.85%) than when monitoring with reading aloud (21.52 %).

Table 4
Only Units of Information About the Reading Instructional Comments

	Maze		Reading Aloud		Across Conditions
	Statements (% of Total)	Percent of Condition	Statements (% of Total)	Percent of Condition	Statements (% of Total)
Phonics	49 (11.64)	26.63	45 (10.69)	18.99	94 (22.33)
Word Recognition	16 (3.80)	8.70	42 (9.98)	17.72	58 (13.78)
Fluency (practice included)	30 (7.13)	16.30	77 (18.29)	32.49	107 (25.42)
Vocabulary	12 (2.85)	6.52	22 (5.23)	9.28	34 (8.08)
Comprehension	77 (18.29)	41.85	51 (12.11)	21.52	128 (30.40)
Total	184 (43.71)	100.00	237 (56.29)	100.00	421 (100.00)

Curriculum. As an aside, we counted the number of times teachers mentioned a specific curriculum. The specific curricula are listed below, broken down by grade level (see Table 5). When discussing students monitored with reading aloud measures, teachers mentioned curriculum 160 times; with maze selection procedures, curriculum was mentioned 133 times.

Table 5
Curriculum by Grade Level

1st	2nd	3rd	4th
	Soar to Success	Corrective Reading	New Practice Readers
	Houghton-Mifflin leveled readers	Read Naturally	Corrective Reading
	Direct Instruction (sight words)	Explode the Code	Read Naturally
	Reading Mastery	SRA Specific Skills series	Explode the Code
	Language!	CCC	Houghton-Mifflin spelling and vocabulary
	Read Naturally	Accelerated Reader	Language!
	Early Success	Houghton-Mifflin	SRA
		Soar to Success	Reading Mastery
		Early Success	Spelling Mastery
		Houghton-Mifflin leveled readers	Houghton-Mifflin
		Reading Mastery	

		Spelling Mastery SRA Direct Instruction	TAP IFL Edmark Merrill Linguistics
5th	6th	7th	8th
Corrective Reading Read Naturally Houghton-Mifflin Explode the Code Steck-Vaughn for vocabulary and comprehension SRA Multiple Skills Series New Practice books Target Spelling Everyday Math Accelerated Reader Language! Edmark Spelling Mastery Merrill Linguistics	SRA Corrective Reading Read 180	Corrective Reading SRA SRA Direct Instruction Patricia Cunningham's Word Building Read Naturally	Corrective Reading SRA Direct Instruction Patricia Cunningham's Word Building

Student Reading Performance

Table 6 presents means and *SD*s for all dependent variables broken down by condition.

Table 6
Descriptive Statistics for Total Sample, Maze Selection, and Reading Aloud

	Total (N = 66)		Maze (n = 33)		Reading Aloud (n = 33)	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Mean Maze (pretest)	2.90	2.91	3.04	2.99	2.77	2.88
Mean Reading Aloud (pretest)	48.72	30.68	50.03	34.38	47.52	27.34
Mean Maze (posttest)	5.19	4.82	5.87	5.03	4.57	4.62
Mean Reading Aloud(posttest)	61.80	31.54	63.40	35.74	60.33	27.63
Kaufman (raw score) Letter-Word Identification	38.06	9.36	38.03	10.15	38.09	8.71
Kaufman (standard score) Letter-Word Identification	75.85	7.55	76.75	8.17	75.00	6.94
Kaufman (percentile) Letter-Word Identification	7.67	9.29	8.75	8.68	6.65	9.85
Kaufman (raw score) Reading Comprehension	18.82	9.47	19.56	10.75	18.12	8.19

Kaufman (percentile) Reading Comprehension	6.69	10.56	8.77	12.35	4.74	8.27
Kaufman (standard score) Reading Comprehension	73.42	8.73	75.13	9.99	71.82	7.15
Kaufman (raw score) Composite	149.27	15.57	151.88	17.30	146.82	13.55
Kaufman (standard score) Composite	72.97	7.50	74.28	8.30	71.74	6.54
Kaufman (percentile) Composite	5.68	8.50	6.84	8.55	4.59	8.44

Two separate ANCOVAs were conducted for reading aloud and maze selection. Pretest was entered as the covariate. No significant group differences were found for either reading aloud (F = .56, df = 1, p = .75) or maze selection (F = 1.43, df = 1, p = .18). Two separate ANOVA's were conducted for the Kaufman letter-word ID and Kaufman comprehension. Pretest was entered as the covariate. No significant group differences were found for either letter-word identification (F = .01, df = 1, p = .98) or comprehension subtests (F = .38, df = 1, p = .54).

Intercorrelations among all variables for all variables are presented Table 7. Correlations between pre- and posttest for maze selection and reading aloud were .74 and .94, respectively. Correlations between CBM measures and Kaufman scores ranged from .67 to .85.

Table 7

Intercorrelation among all Variables for all Conditions

Condition	Mean Maze (pretest)	Mean Reading Aloud (pretest)	Mean Maze (posttest)	Mean Reading Aloud (posttest)	KTEA-II * Letter Word Identification	KTEA-II Reading Comprehension*
Mean Maze (pretest)	1.00	0.69	0.74	0.67	0.75	0.73
Mean Reading Aloud (pretest)		1.00	0.68	0.94	0.85	0.76
Mean Maze (posttest)			1.00	0.71	0.67	0.78
Mean Reading Aloud (posttest)				1.00	0.82	0.78

KTEA-II Letter Word Identification*	1.00	0.84
KTEA-II Reading Comprehension*		1.00

Note. $N = 66$.
*Raw scores used.

Intercorrelations among all variables for the maze selection condition are presented Table 8. Correlations between pre- and posttest for maze selection and reading aloud were .74 and .91, respectively. Correlations between CBM measures and Kaufman scores ranged from .63 to .83.

Table 8
Intercorrlation Among all Variables for Maze Selection Condition

Condition (Maze)	Mean Maze (pretest)	Mean Reading Aloud (pretest)	Mean Maze (posttest)	Mean Reading Aloud (posttest)	KTEA-II Letter Word Identification*	KTEA-II Reading Comprehension*
Mean Maze (pretest)	1.00	0.64	0.74	0.56	0.74	0.69
Mean Reading Aloud (pretest)		1.00	0.52	0.91	0.83	0.76
Mean Maze (posttest)			1.00	0.55	0.63	0.70
Mean Reading Aloud (posttest)				1.00	0.79	0.73

KTEA-II Letter Word Identification*	1.00	0.83
KTEA-II Reading Comprehension*		1.00

Note. $N = 33$
*Raw scores used.

Intercorrelations among all variables for the reading aloud condition are presented Table 9. Correlations between pre- and posttest for maze selection and reading aloud were .74 and .95, respectively. Correlations between CBM measures and Kaufman scores ranged from .72 to .86.

Table 9
Intercorrelations Among all Variables for Reading Aloud Condition

Condition (Reading Aloud)	Mean Maze (pretest)	Mean Reading Aloud (pretest)	Mean Maze (posttest)	Mean Reading Aloud (posttest)	KTEA-II * Letter Word Identification	KTEA-II Reading Comprehension*
Mean Maze (pretest)	1.00	0.72	0.74	0.75	0.77	0.75
Mean Reading Aloud (pretest)		1.00	0.77	0.95	0.86	0.76
Mean Maze (posttest)			1.00	0.80	0.72	0.83
Mean Reading Aloud (posttest)				1.00	0.85	0.80
KTEA-II Letter Word					1.00	0.86

Identification*	
KTEA-II Reading Comprehension*	1.00

Note. $N = 33$.
*Raw scores used.

Table 10 includes the no-treatment group, and shows the results of difference in posttest scores and Kaufman scores among the three conditions (maze selection, reading aloud, and no treatment). ANCOVA was used for maze selection and reading aloud; pretest was entered as the covariate. For Kaufman scores, ANOVA was used to examine the difference in Reading Comprehension and Letter-Word Identification among the three groups. No statistically significant between-group differences were found. Because there were no statistically significant differences among the three conditions, a post hoc test was not performed.

Table 10
Difference in all Variables Among Maze Selection, Reading Aloud, and No Treatment

		Pretest		Posttest		Kaufman Test		Three Conditions			
Condition		Maze	Reading Aloud	Maze	Reading Aloud	Reading Comprehension	Letter Word Identification	Difference in Maze posttest (ANCOVA)	Difference in Reading Aloud posttest (ANCOVA)	Kaufman Test	
										Reading Comprehension (ANOVA)	Letter Word Identification (ANOVA)
Maze	Mean	3.04	50.03	5.87	63.40	19.56	38.03				
(n = 33)	SD	2.99	34.38	5.03	35.74	10.75	10.15				
Reading	Mean	2.77	47.52	4.57	60.33	18.11	38.09				

Aloud (n = 33)	SD	2.88	27.34	4.62	27.63	8.17	8.71	F = 2.52 (df = 2) p = .09	F = .04 (df = 2) p = .96	F = .27 (df = 2) p =.76	F = .13 (df = 2) p =. 87
No Treatment (n = 16)	Mean	4.12	47.34	5.06	60.53	20.00	39.50				
	SD	4.77	31.02	4.23	37,84	11.17	11.83				

Possibilities for the Discussion Section

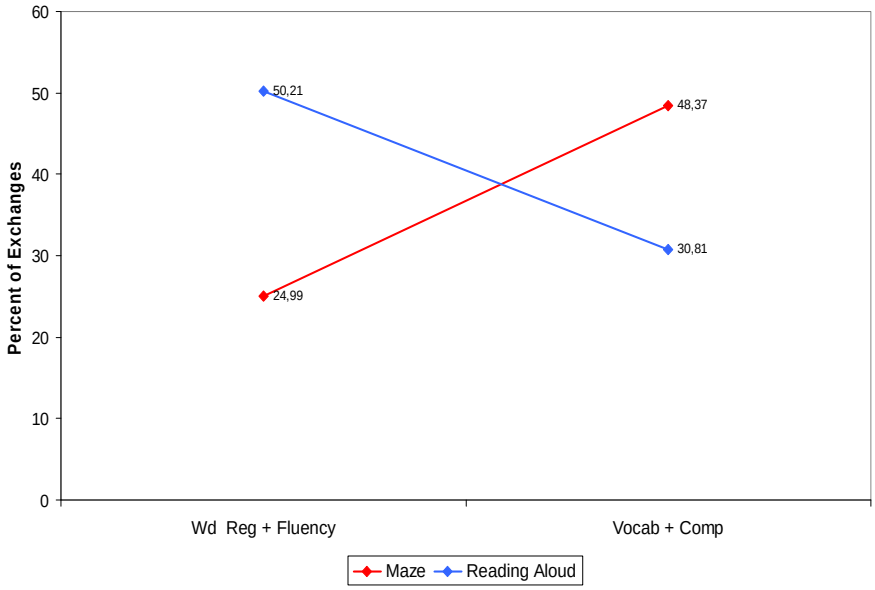
Thoughts from the PCRC presentation

Larger amounts of comments about measurement change within the maze condition.

- Meaning of maze data
  - The maze chart does not fairly represent her fluency improvement. (G4)
  - I cannot conclude anything from the maze because she's bouncing all around (G5, ELL)
- Problems with maze process / ease of fooling system
  - She is currently having a difficult time on the maze test when the choice words are on two different lines (G5, ELL)
  - She tends to get 70% of the passage correct, but makes 2 errors in a row in the beginning. So I do not feel her true ability is showing. (G8)
  - When he does maze, he skips the beginning of the story and only reads the multiple choice sentences (G5)
- Appropriateness for young children
  - I am really frustrated with maze. I know she is guessing. She rushes through and just circles her answers unless I sit right next to her and point to each one. My guess is that our students are too young for this kind of data collection. It doesn't give me the information I like anyway. (G2)
  - She is a 1st grader and it was a difficult task to get her through the maze procedures (G1)

Main question Conclusions

- Does type of measure influence student performance and teacher instruction (attention given to instructional variables)?
  - Student performance: No
    - Virtually no differences RA vs. Maze
    - Little growth at all?
  - Teacher attention to instruction: Yes
    - RA: Teachers focused more attention on word recognition and fluency aspects of reading
    - Maze: Focused more on vocabulary and comprehension aspects of reading
    - Phonics: No differences related to monitoring approach



Other conclusions

- Teachers mentioned making many changes in the measurement procedures, especially for maze selection
  - Function of maze procedure itself?
  - Function of teachers experience with RA vs. their inexperience with maze?
  - Teachers noted issues with maze selection that may be worth examining further
  - Maze reflected very, very low rates of growth

Speculate about the varied levels of CBM knowledge that teachers might have had, especially the teacher who had marked only self-study and conferences (p. 3).

References

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman test of educational achievement* (2nd ed.). Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman test of educational achievement: Comprehensive form manual* (2nd ed.). Circle Pines, MN:AGS Publishing.

National Reading Panel (NRP). (2000). *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.* (NIH Publication No. 00-4754). Washington, DC: U.S. Government Printing Office.

QSR International. (2008). NVivo 8 [Computer software]. Doncaster, Australia: QSR International.

Appendix A

Moodle Illustrations

Figure A1. Illustration of Moodle forum

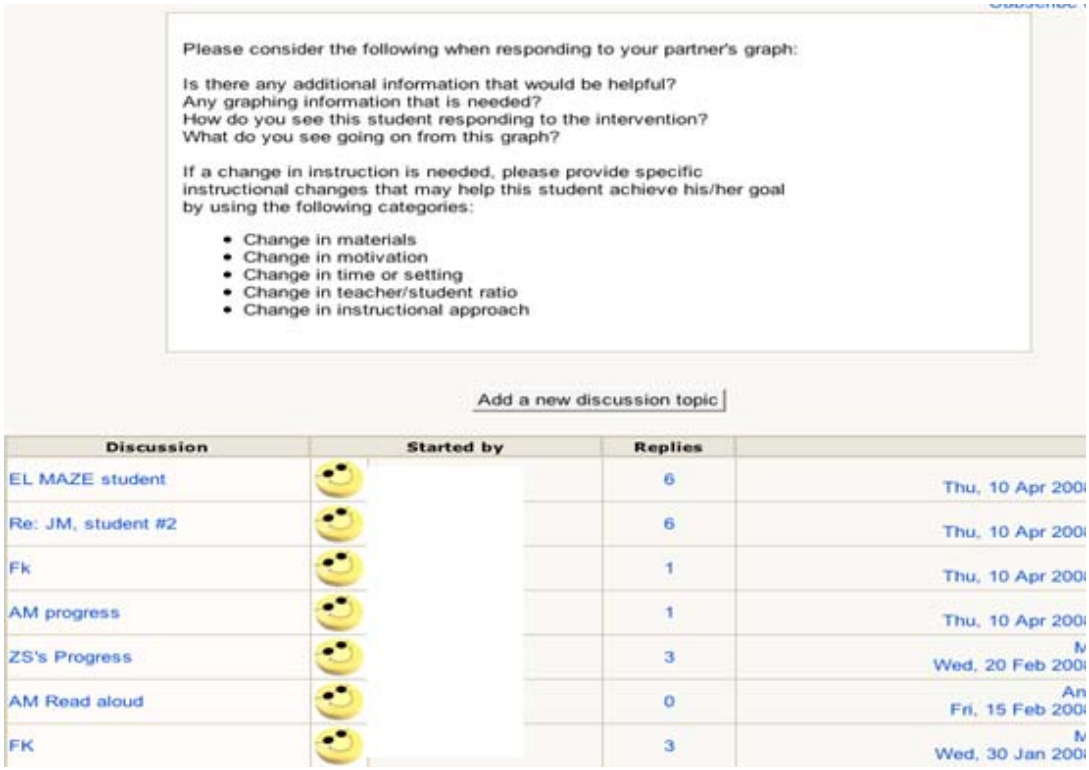
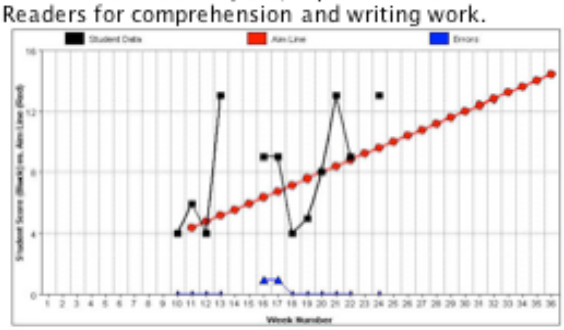


Figure A2. Illustration of a posting and replies within a Moodle forum

Initial posting – 1502 – March 2, 2008

AE is usually at or above the aimline. He is a 4th grade student monitored with a 3rd grade MAZE. He's conscientious about his work. He's made great progress in his ability to decode multi-syllable words and his comprehension has never been an issue. He self-corrects or rereads when things don't make sense. He reads 50 wpm on a 4th grade passage, which is at the 11th percentile.

We use Read Naturally 3.0, Explode the Code workbooks for independent work, and New Practice Readers for comprehension and writing work.



Reply – 1507 – March 3, 2008
What a great thing to have a student at or above the aim line! What are the New Practice Readers you are using for comprehension? They may be helpful for a few of my students.

Reply – 1502 – March 3, 2008
New Practice Readers are simply books that introduce vocabulary before a non-fiction story approximately 4 paragraphs long, followed by comprehension questions. They are as old as the hills and come in different levels, from PIC. I use level A, which I believe to be at the 2nd grade level. AE's next "strategy change" will probably require him to do it independently. For now, we read it out loud together, discuss questions, and then they write the answers independently. We practice skimming the text for information, learning how to ask questions about what we read, how to answer a question in complete sentences, etc.

Reply – 1507 – March 4, 2008
I will have to investigate those. I think they would be something I could use with two of my students who are not participating in the project. Thanks for the information about this – it is helpful.

Appendix B

Teachers' Replies to Questions about the Maze Selection Probes

1. Have there been any formatting problems with the CBM probes?
    - There were no problems. (Eleven teachers)
    - Some word selections are obvious due to previous context clues so the students can get the answer without reading; they are using compensatory skills instead. (Two teachers)
    - The three selection choices should be positioned on a single line; one student made 2 selections if the choices were split onto two lines.
    - Make the font bigger for young students.

2. Have you experienced any difficulty as a result of the conversion of reading aloud passages to maze selection passages?
    - No problems were experienced (six teachers).
    - Stopping after two incorrect may present a false picture of a student's progress, especially at the beginning of a story. Students seem to get more correct as they get into the story.
    - An improvement was seen; students were confused at the beginning, but seemed comfortable with the process by the end.
    - Students seem to do better with reading aloud because the teacher can correct errors.
    - Teacher likes the maze selection measure.
    - One student zipped through a passage because it was familiar. Weekly maze selection has helped in another student's reading strategy because she applies the "use sense to fill in the blank" strategy of maze selection to other reading situations.
    - Although maze selection seems to have face validity, progress is difficult to measure because it is slow.
    - Students didn't seem to compare the two (reading aloud passages and maze selection passages). A challenge with maze was the inability to tell wpm for IEP progress reporting.
    - Teacher prefers reading aloud for young students; however, a fellow teacher who uses maze for older children has commented that she finds it very effective.

Appendix C

Posttest Sequence

1. Group administration
    a. Practice probe - Model first blank, provide guidance for second blank, and let students work independently on third blank. Review answers.
    b. MAZE probes - Follow the script below.
2. Individual administration, use the following order:
    a. Read Aloud probes - Follow the script below.
    b. KTEA-II Letter-Word Identification – Follow published administration procedures.
    c. KTEA-II Reading Comprehension – Follow published administration procedures.

Administration Script for Progress Monitoring Maze Selection Passages

"**Put your first and last name on the cover of the booklet. Put your pencil down. Do not start until I tell you to. You will be reading two stories. First, I want you to read the first story to yourself. When you come to a part where there are three underlined words in very dark print, choose the one word that makes sense in the sentence. Circle that word. You will have 2 minutes to work. Don't worry if you do not finish. Turn the page. Ready…Begin**." After **2 minutes**, say to the students:

"Stop. Put your pencils down. Turn past the blank page to the second story. Ready… Begin." After **2 minutes**, say to the students:

"Stop. Put your pencils down." Collect the packet and pencils.

Administration Script for Progress Monitoring Reading Aloud Passages

"**When I say, 'Begin', start reading aloud at the top of this page. Read across the page. Try every word. If it takes you too long, I will tell you the word. Keep on reading until I tell you to stop. Remember to do your best reading**. **Ready…Begin."**

- Start the timer when the student begins reading (accurately monitor **60 seconds**).
- If the student misses the first 10 words, discontinue the passage and record "zero" words as the score.
- Mark an error (see box below) with an X and mark the time limit with a vertical line.

Scoring the Reading Aloud Passages

- Supply the word for the student after a 3 second "stall".
- A self-correct is NOT an error.
- Do not penalize a child for dialect.

Count as **incorrect**:

- Any word mispronounced.
- Any omitted word.
- Any word on which the student stalled for 3 seconds.

Appendix D

Table D1

All Units of Information by Category and Subcategory

	Maze Selection Condition			Reading Aloud Condition			Across Conditions	
	Statements	Percent of Condition	Percent of Total	Statements	Percent of Condition	Percent of Total	Statements	Percent of Total
Phonics Total	**49**	**15.86**	**7.13**	**45**	**11.90**	**6.55**	**94**	**13.68**
Student Characteristics	26	8.41	3.78	24	6.35	3.49	50	7.28
Instructional Intervention	23	7.44	3.35	21	5.56	3.06	44	6.40
Word Recognition Total	**16**	**5.18**	**2.33**	**42**	**11.11**	**6.11**	**58**	**8.44**
Student Characteristics	7	2.27	1.02	13	3.44	1.89	20	2.91
Instructional Intervention	9	2.91	1.31	29	7.67	4.22	38	5.53
Fluency Total	**30**	**9.71**	**4.37**	**77**	**20.37**	**11.21**	**107**	**15.57**
Student Characteristics	14	4.53	2.04	33	8.73	4.80	47	6.84
Instructional Intervention	9	2.91	1.31	21	5.56	3.06	30	4.37
Practice Reading Text	7	2.27	1.02	23	6.08	3.35	30	4.37
Vocabulary Total	**12**	**3.88**	**1.75**	**22**	**5.82**	**3.20**	**34**	**4.95**
Student Characteristics	2	0.65	0.29	8	2.12	1.16	10	1.46
Instructional Intervention	10	3.24	1.46	14	3.70	2.04	24	3.49
Comprehension Total	**77**	**24.92**	**11.21**	**51**	**13.49**	**7.42**	**128**	**18.63**
Student Characteristics	32	10.36	4.66	26	6.88	3.78	58	8.44
Instructional Intervention	45	14.56	6.55	25	6.61	3.64	70	10.19
Motivation Total	**36**	**11.65**	**5.24**	**75**	**19.84**	**10.92**	**111**	**16.16**
Student Characteristics	19	6.15	2.77	33	8.73	4.80	52	7.57
Instructional Intervention	17	5.50	2.47	42	11.11	6.11	59	8.59
Materials	**7**	**2.27**	**1.02**	**20**	**5.29**	**2.91**	**27**	**3.93**
Measurement Change	**82**	**26.54**	**11.94**	**46**	**12.17**	**6.70**	**128**	**18.63**
TOTAL	309	100.00	44.98	378	100.00	55.02	687	100.00