



TECHNICAL REPORT #32:

**Teacher Use Study: Progress Monitoring With and Without
Diagnostic Feedback**

*Christine Espin, Stan Deno, Kristen McMaster, Rebecca Pierce,
Seungsoo Yeo, Amy Mahlke, and Beth Zukowski*

RIPM Year 5: 2007 – 2008

Dates of Study: October 2007 – May 2008

September 2009

The College of Education
& Human Development

UNIVERSITY OF MINNESOTA



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

The purpose of this portion of the study was to examine factors that affect teachers' use of progress monitoring data for designing instructional programs. Specifically we compared the use of progress monitoring data alone with progress monitoring data combined with diagnostic feedback. The hypothesis is that progress monitoring data coupled with diagnostic feedback will increase teachers' use of data in designing student instruction, will broaden the type of interventions teachers use with students, and will effect greater student achievement gains. The study was conducted with the following question in mind: Does the use of a diagnostic feedback system coupled with progress measurement affect teachers' use of Curriculum Based Measurement (CBM) data?

Method

Setting and Participants

Setting. The data collected for this portion of the study were gathered from eight different districts and 21 different teachers' classrooms in southeastern Minnesota. The various school demographics are listed in Table 1.

Table 1
Demographics of School Districts

	Total Pop	% male	% white	% Hispanic	% black	% Asian	% FRL	% Title 1	% ELL	% Sped
School A	869	52.2	97.0	2.1	0.5	0.3	19.2	8.1	0.3	10.5
School B	849	53.9	95.9	0.5	2.9	0.6	29.0	4.4	0.1	15.5
School C	1779	50.9	93.5	1.1	3.5	1.5	19.7	5.0	0.0	11.4
School D	1358	50.4	94.5	2.9	1.4	1.1	22.2	10.0	1.1	13.8
School E	856	51.1	95.6	2.6	0.8	0.6	34.9	14.5	0.9	13.9
School F	1021	52.2	87.6	8.1	1.0	2.5	10.0	15.4	3.4	8.8
School G	3746	52.0	87.0	2.6	5.2	4.5	35.8	3.7	3.0	16.3
School H	1688	52.0	92.7	6.4	0.5	0.2	18.2	5.4	3.1	10.6

Teacher participants. All special education teachers in grades 1-12, along with reading specialists and other remedial reading service providers were invited to participate in the study. The invitation was extended via email to both teachers and principals. Follow-up letters through school mail and follow-up phone calls and/or emails were also conducted to teachers who had previously participated in studies or had expressed an interest in participating in future studies. Twenty-two teachers responded affirmatively, but one teacher later removed herself from the study.

The 21 participating teachers were certified in the following licensure areas: Elementary Education ($n = 11$); SLD ($n = 14$); EBD ($n = 4$); other special education categories ($n = 8$); and Reading licensure ($n = 1$). Experience levels for the teachers ranged from 3 years to 37 years (mean = 15.38). The teachers reported using CBM procedures for between 1 year and 16 years with most teachers having 3-6 years experience. Most teachers had received training in collection of CBM data through either school- or district-level training ($n = 19$). Some teachers received training via university level coursework ($n = 2$). All teachers were Caucasian; age ranges were reported as being 20-29 ($n = 3$); 30-39 ($n = 3$); 40-49 ($n = 5$); 50-59 ($n = 8$) and 60-69 ($n = 2$).

Student participants. Once teachers were secured and consent for their participation signed, teachers identified students that met eligibility criteria of: having a reading goal, being able to read a minimum of 10 words from connected text in 1 min, and demonstrating a consistent attendance record. Parental consent forms were sent home with identified students. A gel pen was used as an incentive for returning the form, regardless of parent giving consent or not. A total of 125 students returned signed permission forms. Students had the study explained to them by the research assistant conducting the pretesting. Students then signed assent forms prior to pretesting.

Participating students from each teacher's caseload were ranked according to their mean score on the pretest oral reading measures. Within each teacher's group, students with similar mean oral reading scores were paired. If more than two students' scores were close, the students with the least mean pretest difference between correct maze choice and incorrect maze choice were paired. A total of 56 pairs were created ($n = 112$). All students who were not paired were considered as a no treatment group. After similar pairs of students were created, the partners were randomly assigned using the following process:

- Using Excel, a list of random numbers between 1 and 100 was created.
- Even numbers were equated with the SARF condition; odd numbers were equated with the No SARF condition
- The first student of a pair was assigned to the condition equated with the next number on the list. The second student was assigned to the remaining condition.

Each teacher participant had one, two, or three pairs of students in treatment conditions. These were the students whom the teachers progress monitored and discussed in online forums. Once the paired students had been assigned to the conditions (CBM data with or without diagnostic feedback), a t-test analyzed the average mean oral reading fluency score for both conditions. There was no statistically significant difference between the two groups ($t = .39, df = 110, p = .70$).

One teacher pulled out of the study which eliminated six students, another seven students either moved or no longer received the reading support and were also removed from the final analysis. The matched partners of these students were also removed. Additionally, 13 students were not placed into a treatment group. Thus, the results in this study were determined from data

collected from 92 students (46 in the CBM with diagnostic feedback group and 46 in the group without feedback). This sample included 76 students who received reading assistance from special education services and 16 who received reading assistance from Title 1 services. The primary disability for 72 of the students in special education was a high-incidence disability (e.g., learning disabilities [$n = 47$], emotional/behavioral disabilities [$n = 7$], other health disabilities [$n = 5$], autistic spectrum disorders [$n = 6$], developmental cognitive disabilities [$n = 5$], and speech/language [$n = 2$]); the primary disability for 2 of the students was physically impaired and for the last 2, deaf and hard of hearing. See Table 2 for demographic information of students in treatment groups.

Table 2
Student Demographics of 92 Students in Treatment Conditions

	Without Diagnostic Feedback	With Diagnostic Feedback
Gender		
Male	32	34
Female	14	12
Grade		
1 st	0	1
2 nd	6	5
3 rd	10	7
4 th	7	8
5 th	8	10
6 th	5	4
7 th	4	4
8 th	1	1
9 th	1	2
10 th	2	2
11 th	2	2
SES		
Free lunch	20	16
Reduced lunch	6	3
Normal rate	20	27
Ethnicity		
African American	1	2
Hispanic American	1	3
White American	44	41
Disability		
Special education	38	38
Title 1	8	8
Total Participants	46	46

Independent Variables

The independent variables for this study were progress monitoring data with and without feedback provided from diagnostic software. Teachers were asked to develop instructional programs based on data from the oral reading CBM progress monitoring alone or data from CBM *and* cumulative feedback on miscues committed by the students as recorded in the diagnostic system. The progress monitoring probes were oral reading passages from the AIMSweb system, which provides sufficient progress monitoring passages so the teacher can administer a new probe each week. For students in the CBM progress monitoring alone group, teachers were asked to simply administer oral reading probes and used the resulting data to inform instruction. For students in the CBM with diagnostic feedback group teachers were asked to perform an additional step. Following administration of the oral reading passage, teachers entered the type of miscue the students made into the Subskill Analysis of Reading Fluency (SARF; Scullin, Werde, & Christ, 2006) system which then created a matrix of the specific types of errors. See Appendix A for description of the various types of miscues.

Subskill Analysis of Reading Fluency (SARF) Data. The SARF system had all of the AIMSweb passages precoded for the type of word. (See Appendix A for a description of the various word types). Teachers had to determine which miscue a student committed. Teachers had received training at the beginning of the study from the research assistants in the various types of miscues. Teachers wrote what the student actually said or used miscue codes on a paper copy of the passage during administration and then transferred the information to the SARF program. Once the miscues were entered the teacher could then access a cumulative matrix of the student's

miscue patterns and word type error patterns. This information could then be used to formulate instructional plans. (See Appendix B for a sample miscue matrix)

Dependent Variables

Dependent variables were the focus of teacher instruction and student performance. To measure the effect on teacher instruction, we analyzed the discussions between pairs of teachers and the results of surveys completed by teacher participants. The effect on student reading performance was assessed using CBM pre- and posttest oral reading and maze selection measures and a standardized achievement test administered at posttest only.

Focus of Teacher Instruction

Online discussions. Once a week, teachers were asked to communicate with a partner on a password-protected Moodle website furnished by the university from which this research was conducted. After logging into this site, the teachers would click on their dyad. This action would take them to a forum in which they could post a student's progress-monitoring graph and an accompanying description/explanation of the student's progress including information regarding the student's SARF data. The following suggestions were available to guide the teachers in selecting information to post: grade level of the student; classwide screening CBM data (results of tri-annual school-wide screening at grade level); progress monitoring level (including errors) and the reasons for choosing that level; the gain per week selected for the student; an explanation of the interventions including number of students in the group and the amount of instructional time; an explanation of the response each student to instructional interventions; and reasons for making a change in interventions, monitoring level, or aimline. The teacher's partner would also post a graph and information about one student.

After posting, the teachers would reply to their partner's post. The following guidelines were available 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. Changes in instruction could include a change in materials, motivation, time or setting, teacher/student ratio, or instructional approach.

Surveys. Teacher participants completed two end-of-study surveys. Both surveys were developed by the researchers specifically for this study to gather information about how teachers used progress monitoring data (with and without diagnostic feedback) to inform instruction and to obtain feedback on the technological requirements. The analysis of the surveys resulted in no significant difference between treatment groups. For detailed information and findings about the surveys, see Tech Report #35.

Student Reading Performance

Common CBM probes. Two oral reading and two maze selection probes at the second grade reading level were administered as pre- and posttests to all student participants. The oral reading probes were selected from Peabody CBM Reading Passages & Word Counts (Vanderbilt University). The maze selection probes were selected from Project PROACT MAZE Reading Passages (Vanderbilt University). A research assistant went through each set of probes and selected five passages that were not culturally-dependent (e.g. based on a holiday, tradition, or

folktale), were not too familiar (e.g., based on a fairy tale), and were longer than 350 words. The primary investigators made the final choices. Alternate-form reliability for oral reading and maze selection were .95 and .70, respectively.

Standardized reading subtests. Two subtests from the Kaufman Tests of Educational Achievement – Revision II (KTEA-II; Kaufman & Kaufman, 2004) were administered as posttests: 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 correlation between CBM measures and Kaufman subtests. Note that these correlations were calculated on a larger sample of 186 students that participated in this study and a parallel study that took place in another district (see Technical Report #31).

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	.92	.79	.80	.74	.62
Reading aloud (post mean)		1.00	.73	.80	.77	.63
Maze (pretest mean)			1.00	.82	.65	.68
Maze (post mean)				1.00	.72	.73
KTEA-II Letter-Word Identification					1.00	.75
KTEA-II reading comprehension						1.00

Note. $N = 104$.

*Raw scores used for correlations with KTEA subtests.

Procedures

Focus of Teacher Instruction

Progress Monitoring Data. Teachers were asked to conduct weekly progress monitoring data collection using the districts' adopted system of AIMSweb. Teachers entered each student's oral reading data into the password protected AIMSweb system. This system created a graphic display of the student's progress toward the teacher-established year-end goal. Teachers set student goals using projected targets developed by AIMSweb. Most targets were developed by striving to have the student move up one level on the norms chart provided through AIMS. A

vertical line on the students' graphs signified when teachers made a change to instructional interventions.

Students in both conditions were administered weekly oral reading probes from the AIMSweb system. For students in the condition with diagnostic feedback, teachers also entered miscues into the SARF system to have student errors analyzed and categorized.

Scoring, based on rules for scoring in the AIMSweb program, for the oral reading passages was as follows:

- Words would be supplied if the student paused for 3 seconds
- Omissions would be counted as errors
- Mispronunciations would be errors
- Substitutions would be errors
- Dialect would not be counted as an error
- Insertions would not be errors

Continuation of study in HVED. Because of technical difficulties with the SARF program, it was not loaded onto teachers' computers until end of January to mid February (between the weeks of January 21st to the week of February 11, 2008). Due to the span of time from initial training and the actual use of the SARF system research, assistants were available to give teachers a review session on miscue coding. In addition, teachers were asked to extend the study by 6 weeks (extending from the original ending date of April 4th to May 16th). During this time the teachers were asked to mail weekly copies of the students' SARF Matrix reports. The purpose for doing this was to encourage teachers to look at the data weekly. Teachers were asked to continue with Moodle discussions during this time; however, minimal discussion occurred as teachers were more focused on end-of-year routine duties.

Fidelity Observations. The duration of this study was divided into three rotations. The first two rotations lasted about 7 weeks. As mentioned previously, the third rotation was extended to about 12 weeks. In the middle of the first rotation and the beginning of the third rotation we checked the fidelity of CBM administration and error entry into SARF. During the first fidelity check, a research assistant used a fidelity checklist to observe the number of administration step teachers followed as they administered a weekly CBM probe and categorized errors. The second fidelity check was conducted differently; in addition to being observed, the teachers taped their weekly administration, and then sent the tape and a copy of the resulting SARF matrix report to a research assistant. For both observations, the teacher and a research assistant determined total words read correctly. Interscorer agreement was 99%. Teachers demonstrated 99% accuracy in following standardized CBM administration procedures.

Online Discussions. Teachers were asked to share and discuss CBM graphs in a password-protected online forum using the procedure previously described. This procedure had been field-tested in a pilot study that preceded this study. In the pilot study, a password-protected forum was developed using Moodle software and directions for posting graphs and writing information was given to three special education teachers. Each teacher posted information and a CBM graph about one of their students, and responded to the postings of the other two teachers. Afterwards, teachers completed a brief survey. The results suggested the Moodle site was easy to access, user-friendly, and more feasible for discussing and sharing data than face-to-face meetings.

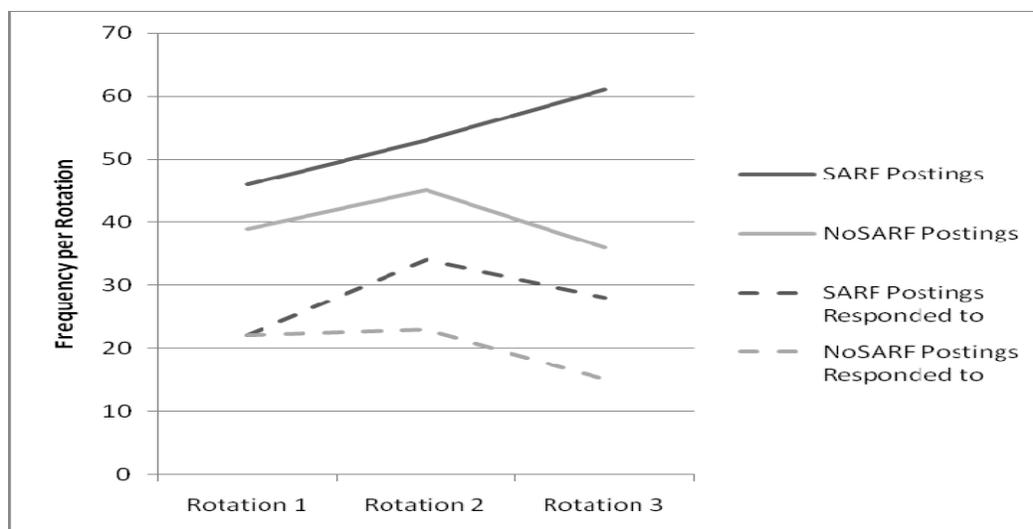
Teachers were paired for Moodle discussions by grade levels whenever possible. This was done by first dividing teacher participants into grade levels (primary, intermediate, higher

grades), and then randomly assigning partners within each stratum. Teachers were paired with new partners for each rotation.

Frequency of online interaction. Each teacher was asked to post a student graph and accompanying information six times (approximately once a week) during each rotation, and were encouraged to continue posting once a week during the six-week extension of the study. In addition, they were to respond to their partner's postings. If a teacher had three pairs of matched students, information about each student was posted once during a rotation. For those with two pairs, information about each student was posted once and the teacher selected two students about whom to post a second time. If a teacher had one pair, information about each student was posted three times during each rotation.

In actuality, most teachers did not post or respond as frequently as requested by study requirements. Actually levels of postings and responses can be seen in Figure 1. Although teachers were expected to post and respond a total of 18 times (6 for each of 3 rotations), the actual number of postings per teacher ranged from 1 to 25 and the number of responses from 1 to 16. In total, teachers made 120 postings and 60 responses when discussing students in the condition without diagnostic feedback. When discussing students in the condition with diagnostic feedback, they made 160 postings and 84 responses. The total number of words written about students in the condition without diagnostic feedback was 11,495 (range per teacher = 0 to 2,233); 15,717 words (range = 38 to 2,072) were written about students in the condition with diagnostic feedback.

Figure 1. Postings and postings responded to by rotations.



Coding scheme. Research assistants organized the archived online discussions by teachers so each protocol contained all the postings and related responses for students who were monitored with or without diagnostic information. Once the protocols were formatted, they were imported into NVivo, a software program that researchers use to store and systematically code data (QSR International, 2008). A primary investigator and a research assistant independently identified units of information or *statements* within each protocol. A statement contained only one thought specifically related to reading. Disagreements were discussed until an agreement was reached. After the statements were selected, each was assigned a code. First, the investigator and research assistant worked together as they assigned codes to comments within six protocols. Next, the two researchers independently coded two randomly selected protocols and calculated point-by-point agreement (total agreement/total units of information), demonstrating 92% and 93% agreement. The rest of the protocols were coded independently by one or the other of the two researchers.

Coding online discussions. We began the analytical process by using the five components of a balanced reading program as a priori codes: phonemic awareness, phonics, fluency, comprehension, and vocabulary. However, we made changes as the coding progressed so ended

with the following main categories: phonics (subsumed phonemic awareness), word recognition, fluency, practice reading, comprehension, vocabulary, general errors, fluency/accuracy, curriculum, material change, change measurement system, and motivation. General errors was added as a category because the focus of the diagnostic feedback system was error analysis. In addition to general errors, we also gathered information about errors specific to each of the reading codes: phonetic errors, fluency errors, practice errors, word recognition, vocabulary errors, and comprehension errors. As we coded, we observed that teachers occasionally made comments about wanting a student to read *slower* in order to improve accuracy, thus, we created a category (fluency/accuracy) to capture such comments.

Finally, we noticed that teachers tended to discuss either student characteristics or instructional programming. Thus, we split each of the main categories listed above into two subcategories: student characteristics and instructional interventions. Statements about student characteristics included comments about a learner's achievement levels or abilities (e.g., "She has great potential, but does not apply herself. She refuses help and does not complete work or turn in assignments."). Statements about instructional interventions included strategies a teacher used or considered using to improve a learner's reading skills (e.g., "My key to is to find out what helps her to stay focused on her work.").

Code description. Eight of the categories focused specifically on reading skills: phonics, word recognition, fluency, practice reading, comprehension, vocabulary, specific errors (within the previously listed categories), and fluency/accuracy. The phonics category contained statements about decoding words using phonemic awareness, applying phonetic rules, reading word parts, and chunking word components (e.g., "It sounds like he has accomplished a lot this year with the sight words and CVC."). Statements that discussed building automaticity in word

reading were coded as word recognition (e.g., “Two days a week, we work on Edmark level 2 on the computer to work on sight words.”). A fluency code was given to comments that discussed reading text in a smooth, effortless manner, including statements about repeated readings and chunking groups of words in phrases (e.g., “He is reading much more smoothly now and with expression.”). Practice reading reflects statements that simply discuss opportunities that increase the time a student spends reading at home, reading with a partner, or reading silently (e.g., “The reading assignment every night is to read 15 min.”). This category ended up being subsumed under fluency during analysis.

The categories comprehension and vocabulary are related to understanding text. The comprehension code was assigned to statements that discussed the understanding of text meaning, including the use of context clues, answering and generating questions, and summarizing and recalling ideas (e.g., “I am also doing a lot of strategies such as visual imaging or making connections.”). Statements were coded as vocabulary reflected understanding about word meanings, so incorporated comments about word study and finding definitions (e.g., “She has a very low vocabulary, so I am always reminding myself to check for understanding with her, even on the most basic things.”).

The last two reading codes were sub-divisions of the categories listed above and dealt with specific reading errors and slowing a reading rate to improve accuracy. Within each reading category, comments made specifically about miscues a student made were coded. An example of a comment coded as a word recognition error is, “Nearly all of her errors are irregular words.” In contrast, an example of a comment coded as a phonetic error is, “His mistakes are more often than not with dropping suffixes.” The fluency/accuracy code includes comments that discuss improving a student’s accuracy by having them slow down their reading rate (e.g., “I have also

been doing a lot of paired reading with him in an attempt to get him to slow down and not skip words.”).

Five of the categories were not directly related to reading instruction: general errors, curriculum, material change, change measurement system, and motivation. Comments were coded as general errors if they were general in nature (e.g., “He makes very few errors.” or “One of the questions I would have is what kind of errors is this reader making?”). Many of these comments centered on errors made during weekly progress monitoring sessions. Statements were coded as curriculum when they referred to a specific curriculum by name (e.g., “We work from the Sopris-West Language program.”). If teachers discussed or desired a change in materials, their comments were coded as material change (e.g., “The J & J Readers though are so boring and truly need a lot of guidance from me to understand what some of the bizarre words are that I don’t think they would be useful as text to analyze.”). Occasionally, the focus of dyadic conversations centered on making changes to the measurement system rather than to the instructional programming (e.g., “We have tried to do a lot of brain exercises and activities to stimulate his brain before doing the fluencies [progress monitoring timings].”). These comments were coded as change measurement system. Finally, teachers frequently made comments that considered the influence of motivation on student learning. These were coded simply as motivation (e.g., “He does have a great attitude and wants to learn to read” or “It seems like motivation is lacking in a lot of students these days.”).

Student Reading Performance

Administration of pretests and posttests. Research assistants administered pre-and posttests to all student participants in the students’ schools. First, two common maze selection probes (2 min each) were administered using a group format. Next, two common oral reading

probes (1 min each) were administered individually. During posttest sessions, the oral reading probes were followed by individual administration of the two KTEA-II subtests: Letter-Word Identification and Reading Comprehension (in that order). See Appendix C for the script and administration procedures used by examiners. There were approximately seven months of instruction between pre- and posttests.

Scoring fidelity. Two research assistants scored all maze selection probes independently. We scored a word selection 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 stopped after the first set of two adjacent errors, and only the portion of the passage preceding the cutoff point was used to determine CMC and IMC. Any differences were discussed until both parties were in agreement. For data analysis, we used CMC – IMC scores.

We scored the oral reading probes during administration. Total words (TW) and errors (IW) were counted. A word was scored as incorrect if it was mispronounced, if it was omitted, or if the student stalled on it for more than three seconds. Differences in dialect or speech patterns were not counted as errors. Each administrator tape-recorded 12% of the probes. Later, another research assistant independently scored the taped passages. Interscorer agreement (total agreements / 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, standards scores, and percentiles for the KTEA Letter-Word Identification subtest, Reading Comprehension subtest, and reading composite. Differences were resolved by going back to the data. 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 #31). 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.

Our cross-grade sample was made up of students who exhibited poor reading skills so standard scores would likely reduce the variability in scores because they are based on age-level norms. Thus, 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.

Results

Teacher Instruction

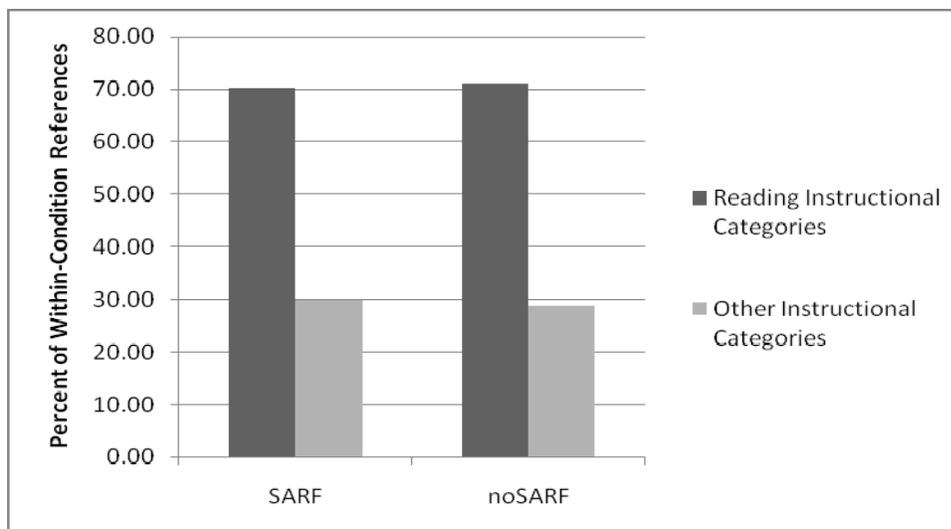
The results discussed in this section were determined by analyzing data collected from online dialogues that occurred during the second and third rotations. We limited our content analysis to the last two rotations because SARF was not working properly during the first rotation; thus, teachers did not benefit from diagnostic feedback until the second rotation.

Using NVivo, postings and accompanying responses were sorted by conditions, progress monitoring with and without diagnostic feedback. Using the Matrix Query function of NVivo we determined the number of statements made in each category (described above) by condition. Next, we determined the percentage of statements by condition and by total statements. We ran the analysis twice, first using all statements within all categories as the total, and next using only statements in the categories that reflected reading components as a total: phonics, word recognition, fluency (including practice reading), vocabulary, and comprehension.

Overall content of teacher exchanges. In the first analysis, it was determined that teachers made slightly more comments (57% of total statements) about students for whom they received

additional diagnostic information. Within both conditions, more comments were made specifically about reading instruction (phonics, word recognition, fluency, practice reading, vocabulary, and comprehension) than about instruction and measurement procedures in general (general errors, motivation, change measurement system, material change). See Figure 2.

Figure 2. Percent of References (Statements) about Reading Instruction vs. General Instruction

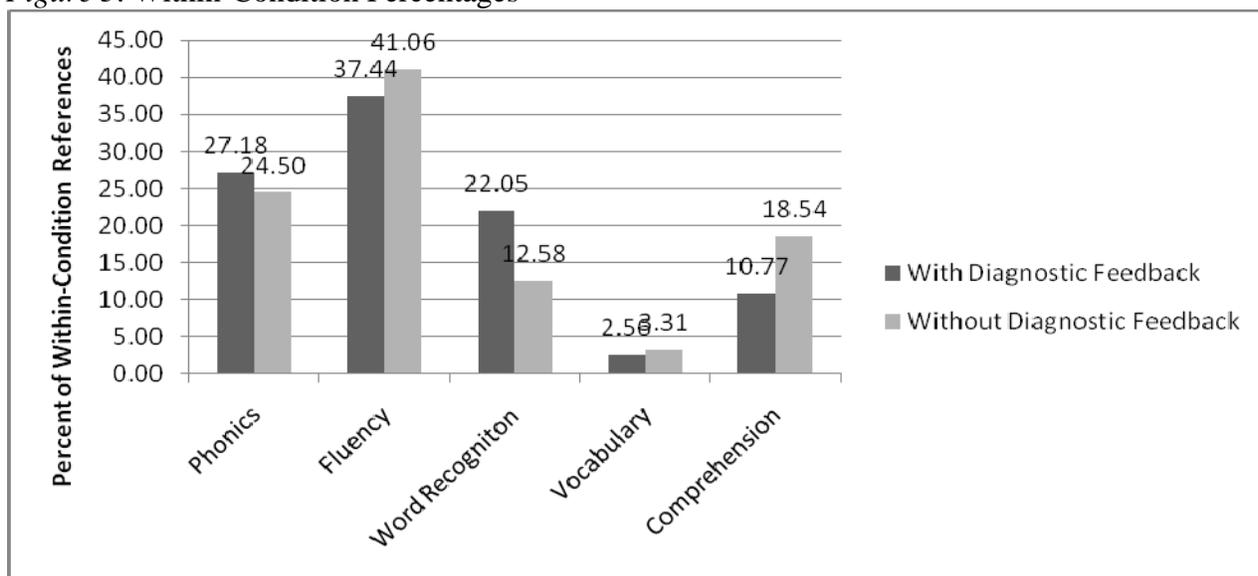


Within conditions, teachers made relatively more comments about general errors (7.91% of conditional references) and changes to the measurement system (7.55%) when discussing students with CBM data and diagnostic feedback, then when discussing those with only CBM data (5.19% and 6.60% respectively). In contrast, more comments were made about motivation (16.04%) when discussing students for whom feedback consisted of only CBM data then when discussing students for whom feedback consisted of CBM data plus diagnostic feedback (13.67%). Comments about general changes to materials were fairly equal between conditions: .94% for the condition without diagnostic feedback and .72% for the condition with additional feedback.

Content of comments focused on reading instruction. In the second analysis, the total reflected only comments made specifically about the five components of reading instruction.

This analysis also included the subcategories that included either comments about errors or about slowing down a student's reading rate to improve reading accuracy. For overall within-condition percentages, see Figure 3.

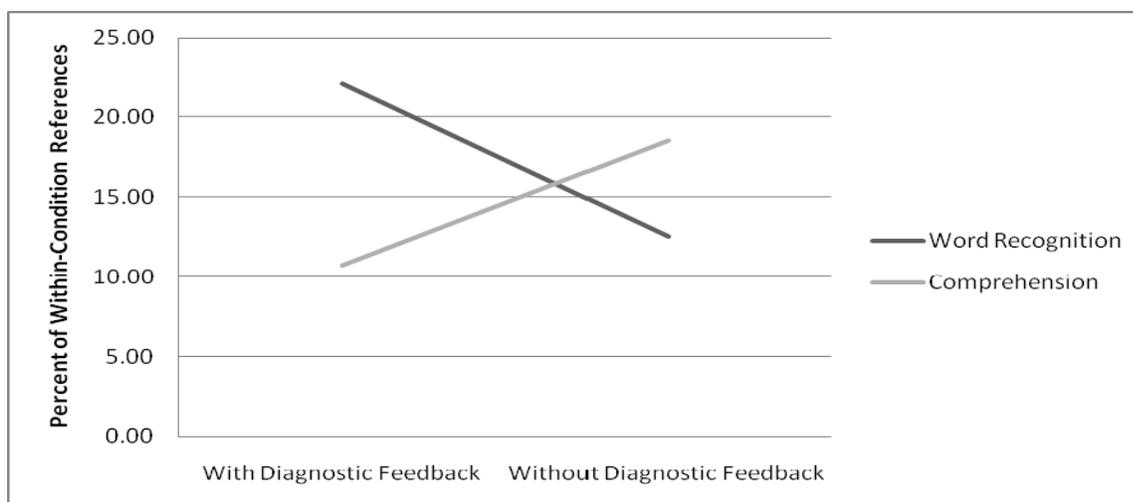
Figure 3. Within-Condition Percentages



Conditional differences were most evident in word recognition and comprehension.

Teachers made more comments about word recognition when discussing students in the with-diagnostic-feedback condition and more comments about comprehension when discussing students in the without-diagnostic-feedback condition (See Figure 4).

Figure 4. Conditional Differences in Word Recognition and Comprehension



Conditional differences were also noted in the comments teachers made about errors specific to the categories of reading instruction, most notably in the phonics and word recognition categories (see Table 3).

Table 3

Percentage of Within-Condition References about Errors Specific to Reading Categories

Type of Error	Percentage of With Diagnostic Feedback Condition	Percentage of Without Diagnostic Feedback Condition
Phonic Errors	2.65	4.10
Fluency Errors	1.32	1.03
Practice Errors	0	0.51
Word Recognition Errors	3.31	11.28
Vocabulary Errors	0	0.51
Comprehension Errors	0	1.03
Total Errors	7.28	18.46

No conditional differences were evident in the comments teachers made about slowing down fluency for the purpose of improving accuracy (15% for the condition without diagnostic feedback and 14.89% for the condition with diagnostic feedback).

Instructional interventions vs. characteristics. As well as being coded as belonging to one of the categories, statements were also coded as being either comments about instructional intervention or student characteristics. Within and across conditions (with or without diagnostic

feedback), the percentage of comments made about instructional interventions and student characteristics were similar (see Table 4). For a more detailed summary of student characteristics vs. instructional interventions, see Appendix D.

Table 4

Units of Information about Student Characteristics vs. Instructional Interventions

	Without Diagnostic Feedback	Without Diagnostic Feedback	Across Conditions
Student Characteristics	44.96	48.52	46.98
Instructional Interventions	55.04	51.48	53.02

Curriculum. Teachers in different grade levels discussed different curricula (see Table 5)., Teachers mentioned a curriculum 122 times when discussing students monitored without diagnostic feedback; curricula was mentioned 164 times for students monitored with diagnostic feedback.

Table 5

Curricula by Grade Level

1 st	2 nd	3 rd
ROAR	Language!	Language!
PALS	PALS	Read Naturally
Language!	6-minute fluency	Accelerated Reading
	Read Naturally	Rewards
	Kansas Learning Strategy- Word Identification	Edmark
	Rewards	High Frequency Flashcards
	Edmark	J & J Readers
	Earobics	6-Minute Fluency
	Accelerated Reading	Soar to Success
	Houghton Mifflin	Rigby Reading series
		SRA Spelling
		Early Reading Empowerment
		Phono-Graphix
4 th	5 th	6 th
Language!	Language!	Rewards
Read Naturally	Rewards	Language!
Handprints by EPS	Foss Science	Reading Strategies That Work
SRA Spelling	Read Naturally	6 Traits of Writing
Dr. Fry's Instant Word Practice Book	Math U See	6-Minute Solution
	Success Maker	Rewards Plus Science

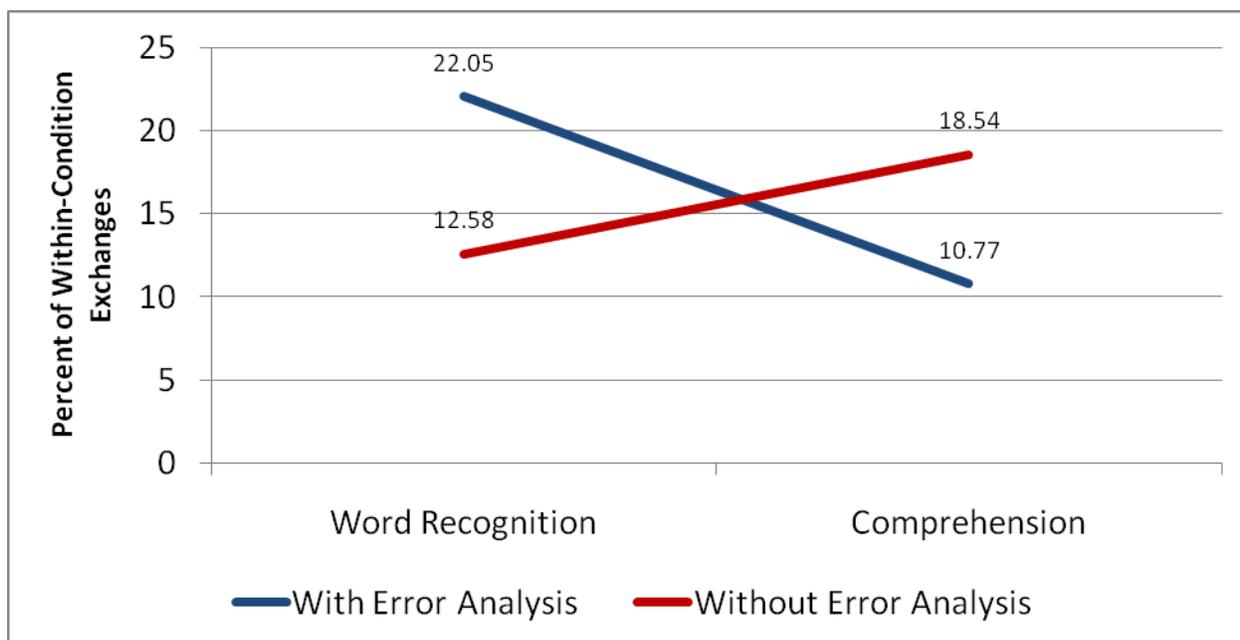
Phono-Graphix Success Maker		
7 th	8 th	9 th
Language! Rewards	Rewards Rewards Plus Science	Rewards
Reading Strategies That Work 6-Minute Solution Rewards Plus Science Read Naturally	Rewards Plus Social Studies	

Student Reading Performance (ADD MORE DATA)

HVED ($n = 92$)											
		Pretest		Posttest		Kauffman test		Maze vs. Reading Aloud (condition)			
Condition		Maze	Reading Aloud	Maze	Reading Aloud	Reading Comprehension	Letter Word ID	Difference in Maze posttest	Difference in RA posttest	Kauffman test	
										Reading Comprehension	Letter Word ID
Sarf	Mean	7.22	76.13	11.34	90.07	30.74	46.98	$F = 1.63$	$F = 1.04$	$F = .18$	$F = .06$
	SD	5.53	34.10	6.95	33.75	10.66	9.17				
	Skewness	.38	.24	.29	.38	-.07	-.19				
	Kurtosis	-1.1	-.28	.52	1.10	-.30	-.42				
No-Sarf	Mean	7.14	75.93	10.20	92.85	31.30	47.74				
	SD	5.43	30.86	6.29	30.40	10.83	8.08				
	Skewness	.79	.39	.56	.22	-.07	-.29				
	Kurtosis	-.08	.52	-.18	.55	-.38	1.19				

Possible Topics for Discussion
(Taken from CEC presentation, 2009)

- Main question: Does the addition of error analysis influence student performance and teacher instruction (attention given to instructional variables)?
 - Student performance: No
 - Virtually no differences with error analysis vs. without error analysis
 - Not unexpected – time needed for instruction to change and effect achievement?
 - Little growth at all?
 - Teacher attention to instruction: Yes
 - With Error Analysis: Teachers focused more attention on word recognition and errors of reading.
 - Without Error Analysis: Although still a small percentage, more focus was on comprehension.
 - Phonics and Fluency: No differences related to the addition of error analysis.



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.

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

International.

Scullin, S., Werde, S., & Christ, T. J., (2006). *Subskill analysis of reading fluency (SARF) 3.4: A*

review of miscue analysis and informal reading inventories (Tech. Rep. No. 1).

Minneapolis: University of Minnesota.

Appendix A

Table A1
SARF Word Type Categories

Word Type Category	Description
Consonant-Vowel-Consonant (CVC)	Three letter words with a Consonant Vowel Consonant configuration.
Words with blends in the beginning (C_V_C)	Words with multiple consonants in the beginning or ending position of the word. These sounds have their regular sounds. The vowel sound is its short sound.
Vowel-Consonant-silent E (VCE)	Words with a long vowel sound with the silent 'e' spelling pattern.
Letter combinations (combo)	Words that contain consecutive letters that produce a specific sound that is the typical sounds of the letters when presented independently.
Prefix/suffix (Pre_Suf)	Words that include a base word that is modified by a identifiable prefix or suffix.
Compound (Comp)	Compound words: composed of two or more words that could stand alone.
Multisyllabic (Multi)	Words that have three or more syllables.
Dolch	Words which are included on the Dolch list of the most frequently encountered in reading.
Instant 1	Words on Fry's list of instant words. This group includes words 1-300 and should be a sight word by the end of 2 nd grade.
Instant 2	Words on Fry's list of instant words. This group include words 301-1000 and should be sight words by the end of 3 rd grade.
Irregular	Words that are not included on the instant word lists. The letter correspondences are not typical and/or low frequency.

Scullen et al. (2006)

Table A2
SARF Error Codes

Error Code	Description
Low Fluency (lf)	When the student takes less than three seconds to say the word, but longer than expected. Student may make attempts on the word when trying to read it aloud.
Miscue Consistent (mc)	When the student incorrectly says a word for the stimulus word, however the spoken word does not change the meaning of the sentence. (saying mom when the stimulus words is mother)
Miscue Disrupted (md)	When the student incorrectly says a word for the stimulus word and the spoken word changes the meaning of the sentence.
Omission Consistent (oc)	When the student leaves out a word when reading the provided printed passage. Leaving the word out does not change the meaning of the sentence.
Omission Interrupted (oi)	When the student leaves out a word when reading the provided printed passage. Leaving the word out does change the meaning of the sentence.
Partial Attempt (partial)	When the student is able to correctly say a portion of the stimulus word, but does not say the complete stimulus word.
Repetition (r)	This is when the student rereads a word or words in a passage.
Reversal (wr)	This is coded when students reverse the order of the letters in word.
Self Correct (sc)	The student corrects a previous error within three seconds of reading it incorrectly.
Word Given (wg)	When the student pauses for 3 seconds on a word, the administrator provides the student with the correct word.

Appendix B

Sample Miscue Matrix

Total Words Read= 420	Instant												Total**
	C*VC*	CVC	VCE	Dolch	Instant 1-300	Instant 301- 1000	Comp	Letter Combo	Prefix/ Suffix	Multi Syllabic	Irregular	Other	
Accuracy=	26	37	14	203	233	36	9	93	27	6	3	29	
	92%	97%	79%	95%	95%	100%	100%	92%	93%	50%	67%	90%	
Low Fluency Consistent	0	0	0	0	0	0	0	0	0	0	0	0	0
Low Fluency Disrupted	0	0	1	1	1	0	0	2	0	1	1	1	8
Miscue Consistent	2	0	0	1	1	0	0	0	0	0	0	0	4
Miscue Disrupted	0	0	1	1	3	0	0	3	0	0	0	2	10
Omission Consistent	0	1	0	0	0	0	0	0	0	0	0	0	1
Omission Interrupted	0	0	0	0	0	0	0	0	0	0	0	0	0
Partial Attempt	0	0	0	0	0	0	0	0	0	0	0	0	0
Repetition	0	0	1	7	6	0	0	0	0	0	0	0	14
Reversal	0	0	0	0	0	0	0	0	0	0	0	0	0
Self Correct	0	0	0	0	0	0	0	0	0	0	0	0	0
Word Given	0	0	0	0	0	0	0	2	2	2	0	0	6
Total**	2	1	3	10	11	0	0	7	2	3	1	3	43

Instant 1-300 - first 300 from Fry's Instant 1000 should be site words by end of 2nd
 Instant 301-1000 - latter 700 of Fry's Instant 1000 should be site words by end 3rd
 CVC - consonant-vowel-consonant with regular sounds short vowel
 C*VC* - multiple consonants with regular sounds short vowel
 VCe - consonants with long vowel silent-e rule
 Letter Combinations - consecutive letters that produce a particular sound that is not regular/typical of the letters independently
 Prefix/Suffix - word or modified word with an identifiable prefix, suffix or derivative at the beginning or end (can be part of base word or stem)
 Comp - Compound Word, Word composed of two or more words that could stand alone
 Multisyllabic - word with three or more syllables (often with pre/suffixes)
 Irregular (but not Instant) - Word whose letter-sound correspondences that are atypical; including low frequency letter combinations
 Other - pronouns and words that do not fit well into other categories

**Totals represent the sum of word classifications. Some words are classified in multiple categories so the sum of classifications is usually

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
Units of Information by Category and Subcategory

Table D1

	Without Diagnostic Feedback			With Diagnostic Feedback			Across Conditions	
	Units of Information	Percent of Condition	Percent of Total	Units of Information	Percent of Condition	Percent of Total	Units of Information	Percent of Total
Phonics Total	37	17.45	7.55	53	19.06	10.82	90	18.37
Characteristics	16	7.55	3.27	29	10.43	5.92	45	9.18
Instruction	21	9.91	4.29	24	8.63	4.90	45	9.18
Fluency Total	40	18.87	8.16	47	16.91	9.59	87	17.76
Characteristics	15	7.08	3.06	23	8.27	4.69	38	7.76
Instruction	25	11.79	5.10	24	8.63	4.90	49	10.00
Practice Reading	22	10.38	4.49	26	9.35	5.31	48	9.80
Word Recognition Total	19	8.96	3.88	43	15.47	8.78	62	12.65
Characteristics	10	4.72	2.04	17	6.12	3.47	27	5.51
Instruction	9	4.25	1.84	26	9.35	5.31	35	7.14
Vocabulary Total	5	2.36	1.02	5	1.80	1.02	10	2.04
Characteristics	1	0.47	0.20	2	0.72	0.41	3	0.61
Instruction	4	1.89	0.82	3	1.08	0.61	7	1.43
Comprehension Total	28	13.21	5.71	21	7.55	4.29	49	10.00
Characteristics	16	7.55	3.27	11	3.96	2.24	27	5.51
Instruction	12	5.66	2.45	10	3.60	2.04	22	4.49
Motivation Total	34	16.04	6.94	38	13.67	7.76	72	14.69
Characteristics	25	11.79	5.10	23	8.27	4.69	48	9.80
Instruction	9	4.25	1.84	15	5.40	3.06	24	4.90
General Errors	11	5.19	2.24	22	7.91	4.49	33	6.73
Change Measurement System	14	6.60	2.86	21	7.55	4.29	35	7.14
Material Change	2	0.94	0.41	2	0.72	0.41	4	0.82
Totals	212	100	43.27	278	100	56.73	490	100