An Observation Instrument for Measuring Classroom Implementation of K-3 Reading Reform

Janice A. Dole
University of Utah

Laurie Lacy
Utah State Office of Education

DRAFT ONLY

An Observation Instrument for Measuring Classroom Implementation

Of K-3 Reading Reform

There is a distinct advantage to learning how to read early in a child’s schooling. Research has shown that those first few years are critical to the learning-to-read process. For example, Juel (1988) reported in her study that most children who were reading substantially below grade level at the end of the first-grade seldom caught up. Further, she cited data from Sweden (Lundberg, 1984) and New Zealand (Clay, 1979) suggesting that regardless of children’s age or language or the instructional method used, “a child who does poorly in reading in the first year is likely to continue to do poorly” (p. 444). Thus, early reading success appears to be crucial for later success in reading and in school.

There is an ever-increasing body of research to substantiate the importance of early reading success. This body of research was summarized in the National Research Council’s book, Preventing Reading Difficulties in Young Children (Snow, Burns, & Griffin, 1998). The research suggests that strong, research-based reading programs can now be developed to increase significantly the number of children who experience early success in reading. In other words, we have the knowledge base to make significant changes in reading instruction in American schools so that almost all children can become successful readers by the end of the primary grades.

Because of the knowledge base now available about beginning reading and the belief that schools can utilize this knowledge base to improve instruction, in 1998 the U. S. Congress recently allocated over a half-billion dollars to school reform in reading. Congress authorized the U. S. Department of Education to conduct a competitive grant, the Reading Excellent Act (REA) Grant. The purpose of the REA grant was to improve reading instruction at the K-3 levels in high-poverty schools across the U.S. In three series of competitive grants, almost all 50
states were awarded grants to implement K-3 reading reform in designated high-poverty schools not making satisfactory progress in reading achievement. The focus of the programs was on helping teachers, schools and districts implement high-quality reading instruction in the primary grades. The expected outcome was significantly improved reading achievement by the end of the third-grade.

The theoretical foundation for reading instruction in the REA Grant was the research base from the National Academy of Science’s Committee published in the *Prevention of Reading Difficulties in Young Children* (National Reading Council, 1998). Based on the theoretical model presented by the National Research Council, the REA Grant employed a model of beginning reading consisting of six components: phonemic awareness, alphabet knowledge, alphabet principle, fluency, vocabulary and comprehension. These six components served as the foundation for the curricular model to be used by REA designated schools in states awarded funds.

Central to the scope of the REA program is the implementation of high-quality reading instruction using the curricular model outlined by Snow and her colleagues. But a major question arising is: *How do we evaluate the extent to which teachers are using that curricular model?* In other words, *how do we determine if teachers are implementing the REA instructional program in their classrooms?*

The purpose of this paper is to present preliminary, quantitative data related to an observation instrument designed to measure the implementation of reading reform in K-3 classrooms involved in the Reading Excellence Act (REA) Grant. We do this through an analysis of the reliability and construct validity of an instrument we developed for this purpose, the Profile of Scientifically Based Reading Instruction (PSBRI). Before we discuss the Profile,
we place the Profile and our efforts on measuring classroom instruction within the larger framework of the monitoring of the implementation of reform.

Reform Implementation Monitoring

Over the last several decades, state and national policymakers have promoted systemic reform in a number of subject areas, including reading, science and math. Systemic reform is believed to be at the heart of increased achievement for students, especially students of poverty. As systemic reform has taken shape across the country, it is related to what reformers call “opportunity to learn” (Stevens, 1993). This term refers to the fact that poor and minority students must have equitable and excellent opportunities to learn if higher achievement is expected of them.

Over the last 30 years, there have been many reform efforts to try to increase students’ opportunity to learn and their achievement. These efforts have involved systemic changes in the way schools and districts operate and make decisions and also changes in specific teaching practices and classroom instruction. As research mounted in the 1970s, and indeed continues today though, researchers often find that despite reform, achievement does not improve. The question then becomes: Did achievement not improve because the educational reform was ineffective or because it had not actually taken place (Hall & Hord, 1987)? Thus, it becomes critical to look carefully and monitor districts, schools and classrooms to see whether reform actually occurs as intended.

One of the first models of reform monitoring was the Concerns-Based Adoption Model (CBAM). This model was developed from over a decade of research on educational innovations and their impact—or lack of impact—on achievement (Hall & Hord, 1987). CBAM was a measure that evaluated the extent to which teachers had adopted educational innovations.
CBAM measured two critical aspects of reform adoption—teachers’ stages of concern about an education innovation and their levels of use of an innovation.

Other models of monitoring teacher change and development have been developed from research on how teachers come to adopt or fail to adopt educational innovations (Fullan, 1992; Fullan & Hargreaves, 1992; Joyce & Showers, 1995). These models can be used to monitor the extent to which an innovation is being considered, supported and used by teachers.

Within this body of literature on educational reform monitoring, surprisingly little research has been conducted recently on measuring teacher behavior and classroom instruction itself. In 1977, Borich and Madden classified and annotated a multitude of instruments designed in the 1960s and 70s to examine classrooms and classroom instruction. But more recently developed instruments are difficult to find. We found several recent instruments that look at teacher professional development (Commeyras & DeGroff, 2000; Danielson, 1996; St. John, Manset, Hu, Simmons, & Michael, 2000) and teachers’ self-reports of their instruction (Baumann, Hoffman, Duffy-Hester, & Ro, 2000; St. John, E. P., Bardzell, J. S., & Associates (1999). But we found very few reports of instruments to look at the actual instruction itself. One recent important exception to this is the CIERA project identifying the characteristics of schools that beat the odds (Fisher & Adler, 1999; Taylor, Pearson, Clark, & Walpole, 1999). The current study is in line with the CIERA project in that we, like they, developed an observational instrument that looked at the reading instruction in the classroom.

**Method**

**Participants**

Subjects for the study were 108 kindergarten, first, second and third-grade teachers from three rural districts in one state in the Southwestern United States. The teachers included 22
kindergarten teachers and 88 grades 1-3 teachers between the ages of 23 and 58. The majority of the teachers were white and female, with 1 of the grades 1-3 teachers being male and 10 of those same teachers being female Native-American. All primary grade teachers in the 12 schools in the three districts participated in the REA Grant. Further, as part of their normal instructional practice within the REA Grant, they were all observed for this implementation study.

**Data Collection Tools**

The **observation instruments**. Two observation instruments were developed to measure teacher implementation of the REA curriculum. The Profile of Scientifically Based Reading Instruction K (PSBRI-K) was developed for kindergarten, and The Profile of Scientifically Based Reading Instruction Grades 1-3 (PSBRI-Grs.1-3) was developed for grades 1-3. The purpose of these two instruments was to determine the extent to which teachers were implementing high-quality reading instruction codified in the Snow et al. text (1998) and used as a foundation for the REA grant.

The PSBRI instruments were developed, piloted and revised by a team of eight researchers and educators, including individuals with expertise in research design and statistics, reading, early childhood education, and educational policy and administration. Instrument development began with the multiple curricular recommendations for a research-based reading program outlined by the National Research Council (1998). At the kindergarten level, the National Reading Council identified seven areas of the curriculum that were critical for reading and language arts development (see Table I for these areas). For example, kindergarten areas included “oral language activities that foster growth in receptive and expressive language and verbal reasoning” and “reading and book exploration with children for developing print concepts and basic reading knowledge and process.” Grades 1-3 included nine areas such as “explicit
instruction and practice that lead to the understanding that spoken words are made up of smaller units of sounds” (see Table 2 for these areas).

The research team operationalized these areas into low-inference teacher practices. In order to generate the teacher practices, we asked ourselves the questions, “What would this look like in a kindergarten classroom? What would an observer expect to see if this recommendation were being followed?” The answers to these questions were then used to develop individual items for each of the areas. For example, oral language activities were operationalized through such items as the following:

- Teacher provides an environment wherein students can talk about what they are doing.
- Teacher encourages students to talk about their experiences and discuss their home culture.
- Teacher models and/or encourages students to ask questions during class discussions.

The teacher practices were recorded under two different categories. The first category documented whether the instructional practice was actually observed during the observation period. There were three sub areas within this category: 1) Observed, 2) Clear Evidence, and 3) Not Observed & No Evidence. Observed Evidence was indicated when the behavior was observed directly. Clear Evidence was indicated when the instructional practice was not directly observed, but there was something in the classroom or something said by the teacher to indicate that the instructional practice did occur in the classroom. For example, Clear Evidence would be checked if an observer did not see writing directly taught, but saw student-writing folders with student work in it. Not observed and No Evidence was indicated when a particular practice was not observed and, after talking with the teacher, there was no evidence that the practice took place in the classroom on a regular basis.
The second category of observation measured the quality of the teacher instructional practices that were observed. There were three boxes for observers to check, 1) Excellent, 2) Good, and 3) Needs Improvement.

In addition to the instructional practices, we also measured student behaviors. We reasoned that the evidence for instruction taking place could be measured not only by what teachers did but also by what students were doing. For example, if the teacher was conducting a read-aloud with children, we would expect to see children highly engaged in listening. Similarly, if a teacher was encouraging students to sound out words by stretching them out, we would also expect to see students sounding out words by stretching them out as they composed texts. For these items, we asked observers to rate how many students were engaged in the targeted student behaviors—none, some, most, or almost all.

The researchers developed, pilot-tested and revised numerous drafts of the observation instruments. This process was iterative and occurred over a period of eight months. Researchers developed a draft of items and then tested their usefulness through careful observations in classrooms that were similar to REA classrooms in terms of teachers and students. After the observations, the researchers would return to the drawing board, revise the items and test them again.

The final PSBRI-K instrument consisted of 33 items grouped across the seven curriculum areas identified by the National Research Council. The final PSBRI-Grs.1-3 instrument consisted of 45 items grouped across the nine curriculum areas identified by the NRC.

The instruments were designed for use by trained observers with expertise in reading. Observers attended to the general recommendations of the National Research Council as well as their own expertise in reading/language arts to evaluate the quality of the instruction itself. The
researchers trained all observers for two days on how to use the instruments, including practice observations and debriefing to calibrate ratings across observers.

The instruments were designed to study teaching behaviors that occurred during the reading/language arts block in K-3 teachers’ classrooms. For all but the kindergarten teachers, this block was a three hour block of time, since the state REA Grant stipulated that teachers spend that amount of time each day for reading/language arts. Thus, each observer was required to observe during the whole reading/language arts block of time for each teacher.

Additional data sources. In order to examine the construct validity of the PSBR, the researchers correlated scores on the PSBRI with scores on several other instruments used in the larger study of the REA Grant. One test was part of the state’s mandatory system of educational assessment; other tests were developed specifically for the REA Grant.

State’s educational assessment instrument. At the end of each school year, the state tests each child from grades 1-6 using a criterion-referenced test, the Language-Arts End-of-Level Test. The numbers of items per test ranges from 62 at grade one to 105 at grade six. The concepts measured by tests at grades one and two include: graphophonic awareness, building reading vocabulary, comprehension, writing conventions and listening. At grade three, the subtests include vocabulary, comprehension, writing process, writing conventions and listening. All of these tests have been subjected to in-depth psychometric analyses and show strong patterns of technical properties (Fox & Nelson, 1999; Nelson & Fox, 1999). These tests were used as evidence of student learning outcomes in reading.

Student Accomplishment Forms. This tool is an informal measure of student performance in reading. The tool was developed for teachers to evaluate their students’ ongoing performance in reading. Four checklists were developed, one for each grade level, K-3. The
forms were intended to help teachers track individual students’ progress and to identify ongoing instructional needs. Teachers were provided with a form with benchmarks identified by the National Research Council. For example, at the kindergarten level, book exploration benchmarks included, “knows where to begin reading a book,” “knows the parts of a book and their functions, including title, illustrations and author.” Teachers were encouraged to use the chart over the year to identify and record when each child had mastered each benchmark for that grade level. Extensive pilot testing was conducted with each of the four grade level forms built for this evaluation. A concerted effort was undertaken to make the form highly usable by teachers.

*The Elementary Reading Attitude Survey.* This survey is a public domain elementary reading attitude inventory designed to measure elementary children’s attitudes toward reading. It is informally known as the Garfield Survey because the character of Garfield is drawn on each item and children circle the Garfield that most represents their attitudes towards reading. This survey (McKenna & Kear, 1990) was given to all students in the REA grant, grades 1-3 as a measure of their attitudes toward reading.

The Elementary Reading Attitude Survey produces two sub scores, recreational reading and academic reading, in addition to a total score. McKenna and Kear (1990) conducted extensive research on this survey. Their paper provides evidence of strong psychometric properties for the survey as well as an adequate norm base for each grade level.

*Staff Developer Reading Strategy Description Forms.* Built for kindergarten and grades 1-3, the two Staff Developer Reading Strategy Description Forms (SDRSDF) were constructed by the research team to obtain global, high inference ratings on the key areas of curricular recommendations used on the PSBRI observation instruments. The content of the SDRSDF for
kindergarten consisted of nine curricular areas similar to those used on the longer PSBRI-K form (see Table 1 for these areas). The content for the SDRSDR for grades 1-3 consisted of the nineteen curricular areas similar to those used on the longer PSBRI-Grs.1-3 form (See Table 2 for these areas). For each area, the observer (staff developer) was asked to rate globally the level of implementation observed during the reading/language arts block. In other words, observers rated on a 1-5 scale, 1 being little or no implementation and 5 being very high implementation, the extent to which they directly observed or saw clear evidence of each of the curricular areas noted by the National Research Council. For example, areas for grades 1-3 included “promotion of spelling-sound correspondence understanding,” “teaching of common spelling conventions and their use in printed words,” and so forth.

It should be noted that all of the curricular areas for the SDRSDF kindergarten and grades 1-3 were incorporated into the PSBRI forms. Thus, the curricular areas identified as being important were the same, all being drawn from recommendations by the National Research Council. The difference was that on the PSBR forms, some areas were combined so that there were seven kindergarten areas and nine grades 1-3 areas, instead of nine and nineteen on the SDRSDF. Further, on the SDRSDF each area was globally rated. But on the PSBRI each area was further subdivided into specific teacher practices. Initial analyses of the SDRSDR and the PSBRI showed total score correlations between .70 and .80.

Teacher Program Description Form. An additional instrument used in the study mirrored the Staff Developer Reading Strategy Description Form but was rated by teachers instead of observers. The Teacher Program Description Form (TPDF) was identical to the Staff Developer Reading Strategy Description Form. The only difference was that teachers recorded the extent to which there was high or low levels of implementation in their own classrooms.
Thus, while staff developers rated teachers’ level of implementation of key areas of the REA curriculum on the SDRSDR forms, teachers self-rated their levels of implementation using the TPDF. For example, for the area, “promotion of spelling-sound correspondence understanding,” the observer would rate each teacher’s level of implementation of this area between 1-5 on the Staff Developer Reading Strategy Description Form. On the other hand, for the Teacher Program Description Form, the teacher would rate herself on her level of implementation of this area, again between 1-5.

Procedure

Data for this study were collected over a nine-month period of time. Expert observers completed the PSBRI-K and PSBRI-Grs. 1-3 forms during the third month of implementation of the REA Grant. Each observer spent one reading/language arts block of time observing each teacher involved in the grant. Observers were asked to administer the forms during a four-week block of time. They needed that block of time since each observation took a complete day, and each observer had a minimum of eight observations to complete. At the same time, the observers also completed the Staff Developer Reading Strategy Description Forms. Thus, observers completed two different measures of the extent to which teachers were implementing the recommended reading curriculum. The Student Accomplishment Forms were given to all teachers early in the fall. Teachers were encouraged to use the forms throughout the year to chart their students’ progress in reading and language arts.

The remaining instruments were administered at the end of the first academic year of the project. The End-of-Level Language Arts tests were administered as part of the state’s educational assessment system in grades 1-3 in May. Teachers administered the Elementary Reading Attitude Survey to their students in grades 1-3 in May as well. Finally, teachers
independently completed the Teacher Program Description Form on their own time during May.

Confidentiality of the forms was ensured during the data collection process. The two forms most in need of confidentiality were the PSBRI forms and the Teacher Program Description Forms. Observers placed the PSBRI forms directly into stamped, self-addressed envelopes to mail to the evaluation team. This ensured that other teachers or administrators did not have access to the results of the forms. In addition, teachers placed their completed Teacher Program Description Forms in stamped, self-addressed envelopes and mailed them directly to the evaluation team.

Results

Reliability of the Profile of Scientifically-Based Reading Instruction

Initial psychometric analyses of the two levels of the Profile demonstrated relatively similar patterns of reliability for both the kindergarten form and the grades 1-3 form. Tables 1 and 2 document the internal consistency reliability of sub-score areas and total scores for the two levels of the Profile. Internal consistency reliability for the total kindergarten observation form was .85. Sub-score internal consistency coefficients ranged from .37 through .82. As would be expected, scales with more items generally produced more reliable results. Internal consistency reliabilities for the grades 1-3 form were slightly higher. The total form coefficient alpha at this level was .92. Sub-score reliabilities ranged from .68 through .90.

Inter-rater reliability for each of the two forms was assessed using intra-class correlation. Total score intra-class reliabilities for average ratings ranged from .89 to .95, depending on the
specific weighting system being used. Intra-class reliabilities for individual raters ranged from .73 through .86, again depending on specific weighting techniques.

The reliability analyses to date on the two observation forms are extremely promising. Both the internal consistency indicators and the inter-rater measures suggest strong psychometric properties for these forms.

Relationships Between the Profile of Scientifically Based Reading Instruction and Student Outcome Measures

Because of the intensive psychometric development work accomplished during the study in the area of assessing program implementation, several measures of levels of reading program implementation were available that could be correlated with various outcome measures. Causal explanations should not be attributed to these relationships, but when these correlations are interpreted in the context of other evaluation results, they take on added importance. Because of the small group of kindergarten teachers available with complete data for all outcome and implementation measures (n=22), only the grade one through three findings are presented here.

Table 3 lists three major implementation measures including the Profile-Grs. 1-3, Staff Developer Reading Strategy Form, and the Teacher Program Description Form. Three major outcome measures are also presented including Utah Language Arts End-of-Level Tests, Student Accomplishment Forms, and the Elementary Reading Attitude Survey. For each of these six variables only the total score correlations between implementation measures and outcome measures are reported.
The Profile showed the strongest pattern of relationships with the Language Arts End-of-Level Tests, .54. This correlation provides evidence of a moderate relationship between the Profile and student achievement, as would be expected. We would also expect, and did find, that this relationship is greater than the relationship between the Profile and the reading motivation survey, which is .34. Further, we would expect the Profile to have stronger relationships to the other outcome measures than to any of the other instruments. Indeed, this does occur. The Profile and the student accomplishment forms correlated .37, and the Profile and the reading motivation survey correlated .34.

Interestingly, the shorter, high-inference Staff Developer Reading Strategy Form demonstrated the same pattern of relationships but at slightly lower levels. The Staff Developer Reading Strategy Form correlated .50 with the Language Arts End-of-Level Tests, .36 with the student accomplishment forms and .22 with the reading motivation survey.

Teacher ratings of their own level of implementation of the reform efforts also showed significant but lower magnitude correlations with the reading achievement test, .30 and the student accomplishment forms, .29. The relationship between the high inference staff developer reading strategy form and the reading motivation survey was not significant, .22. Further, the relationship between the teacher ratings of their level of implementation and the reading motivation survey was also insignificant, .10.

Taken as a whole, the information in Table 3 makes a strong argument for the empirical validity of the instruction dimensions of the Profile. The Profile, both in its long and short forms, relates most strongly to the reading achievement test that, in turn, has the strongest psychometric properties.
Most importantly, the correlations between the Profile and the reading achievement tests reinforce what we know to be the substantial relationship between teaching and learning.

**Conclusions**

The purpose of this study was to evaluate the psychometric properties of an instrument designed to determine the extent to which teachers were utilizing reading reform strategies in their classrooms. This instrument was designed by one state’s REA research team to evaluate the extent to which REA teachers implemented the reading curriculum recommended by the National Research Council. We found that the Profile has good reliability. The internal consistency of the instrument was determined to be .85 for the Profile-K instrument and .92 for the Profile-Gr. 1-3 instrument. Further, we found some evidence for the construct validity of the instrument. The Profile-Gr. 1-3 correlated more closely with reading achievement, as determined by the Language Arts End-of-Level Test, than it did with a teacher-determined reading accomplishment form or a reading motivation survey. Certainly, other evidence of construct validity is warranted, but there are promising data here to suggest that the Profile has sound psychometric properties.

There are other issues that need to be addressed related to the reliability and validity of the instrument. First, we measured reliability of the Profile through internal consistency. The reliabilities were high, both for the Profile-K and the Profile-Gr. 1-3, .85 and .92 respectively. However, test/retest reliability would be most helpful as well. This is because our evaluation was based on one, albeit three hours long, observation. These observations were longer than the usual observation of a given reading lesson—typically one-half hour. In addition, the observation instrument was designed to capture teacher instructional practices that should occur almost, if not every, day. Therefore, if an individual instructional practice was not observed or
there was no evidence of the practice, this would be considered a problem in terms in program implementation. Nevertheless, it would be beneficial to measure the same teacher for at least another day to establish test/retest reliability.

Second, part of the issue of construct validity relates to the extent to which the Profile is consistent with and indeed matches the recommendations of the National Research Council for high-quality reading instruction. As a group of researchers and educators, we took our best guess at translating the recommendations from the Snow et al. text into actual teaching and student behaviors that we could identify. Our best guess was based on the group’s cumulative knowledge, expertise and experiences working in classrooms of primary grade children as well as dozens of concurrent observations of primary grade teachers in their classrooms.

But, ultimately we do not know if our best guess reflected that of the National Reading Council. Translating from general principles to specific and detailed teacher practices is risky business. However, we believe that the total instrument itself, rather than any number of items or subtests, is probably an accurate measure of the extent to which a given teacher is implementing high-quality reading instruction.

It is important to understand what the observation instrument can and cannot do. The observation instrument was designed to look primarily at the reading curriculum and instruction—that is, what it is that the teacher says and does to promote early reading achievement. The instrument looks at the curriculum and the instruction itself—is the teacher reading aloud to children every day, is the teacher explicitly teaching the letter-sound relationship, is the teacher teaching spelling? The instrument does not look at classroom management—is the teacher effectively moving and managing children from place to place, is the teacher making sure students are engaged in their work? The instrument also does not look
at the teacher’s affect—is the teacher a kind and caring person, do children want to engage with the teacher, do children feel safe and comfortable in their classroom?

A specific anecdote can make this caveat clearer to the reader. As we were piloting the instrument, we asked one principal to allow us to observe one of her kindergarten teachers. The principal showed us to her best kindergarten teacher. As we observed this teacher all morning, we were amazed at her wonderful rapport with her children. She was warm and engaging, and her students appeared to love her. We watched the children move comfortably throughout the morning activities with ease and success. However, by the end of the morning, we had checked off fewer than a handful of teaching and student behaviors on our observation instrument. Clearly the instrument did not measure what we saw. Similarly, the teacher did not do what the observation instrument recommended.

Based on our observations, we concluded that here was a gifted teacher in many ways, a warm, caring teacher with superb management skills who was teaching a 30-year old curriculum. She did not allow her students to compose their own texts, she did not show students the print and pictures as she read aloud to them, she did not teach her children letters and sounds in combination, she did not ask students questions and listen to their answers. We strongly believed that all of these activities were within her reach, but she did not know how to do them—either because she had not been exposed to them or because she believed that her old methods were better.

The point of the anecdote is this. The Profile allows observers to look only at reading curriculum and instruction, not at all aspects of teacher and student behaviors in a classroom. The Profile misses some critically important teacher instructional practices, but it was not
designed to measure such practices. It was designed to measure the extent to which the latest recommended *curricular* and *instructional* practices are in place in classrooms.

Interestingly, though, we also believe that it is very difficult for teachers to do well on the observation instrument without management skills and a sense of caring. We believe that effective management and a sense of caring are necessary prerequisites to high-quality reading instruction. The few master teachers we did observe scored almost perfect scores on the observation instrument. They masterfully carried out almost all activities on the observation instruments with ease and grace. Their children were writing pages of stories, they were reading from a variety of printed texts including informational texts, picture books and leveled books. They were writing reports about topics of study in science and social studies. They were learning new vocabulary related to social studies and science. But, they were superb managers of children, and they were caring and responsive individuals as well.

**Value of the Study**

The observation instrument we developed is one of only a handful of reported instruments designed to measure high-quality reading instruction. Researchers have asked teachers about their classroom instruction, including their knowledge and beliefs, their organization and grouping practices, their materials, and their instructional practices. But teacher self-reports cannot be the only way to determine whether change has taken place. Few researchers have developed observation instruments that look at specific instructional practices of teachers to determine the extent to which high quality reading instruction was happening in a given classroom.

We see this observation instrument as helpful to school districts, administrators and reading specialists as they evaluate whether teachers are implementing high quality reading
instruction in their primary grade classrooms. The instrument can be used as a teaching tool for these groups of educators to learn what high quality reading instruction looks like in the classroom. It will not do educators much good to just read the instrument. But observations of teachers in their classrooms using the instrument will make clearer to educators just what is happening instructionally in those classrooms and how those events relate to high-quality reading instruction.
References


Table 1

Coefficient Alpha Reliability Analyses For the Observation of Key Reading Instructional Areas
Using the Profile of Scientifically-Based Reading Instruction Kindergarten Level¹

<table>
<thead>
<tr>
<th>Area</th>
<th>Number of Items</th>
<th></th>
<th></th>
<th>Coefficient Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Daily Activities</td>
<td>Periodic Activities</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td>I. Oral Language Activities</td>
<td>9</td>
<td>0</td>
<td>9</td>
<td>.61</td>
</tr>
<tr>
<td>II. Reading Aloud</td>
<td>7</td>
<td>2</td>
<td>9</td>
<td>.73</td>
</tr>
<tr>
<td>III. Book Exploration</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>.72</td>
</tr>
<tr>
<td>IV. Writing Activities</td>
<td>4</td>
<td>3</td>
<td>7</td>
<td>.71</td>
</tr>
<tr>
<td>V. Extending Understanding</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>.37</td>
</tr>
<tr>
<td>VI. Print-Related Activities</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>.81</td>
</tr>
<tr>
<td>VII. Phonemic Analysis Activities</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>.53</td>
</tr>
<tr>
<td>VIII. Sight Vocabulary/ Alphabetic Principal</td>
<td>3</td>
<td>5</td>
<td>8</td>
<td>.48</td>
</tr>
<tr>
<td>Section I - Daily Activities</td>
<td>36</td>
<td>0</td>
<td>36</td>
<td>.82</td>
</tr>
<tr>
<td>Section II - Periodic Activities</td>
<td>0</td>
<td>20</td>
<td>20</td>
<td>.72</td>
</tr>
<tr>
<td>Total Form</td>
<td>36</td>
<td>20</td>
<td>56</td>
<td>.85</td>
</tr>
</tbody>
</table>

¹ Based on 26 classroom observations
Table 2

Coefficient Alpha Reliability Analyses For the Observation of Key Reading Instructional Areas Using the Profile of Scientifically-Based Reading Instruction Grades 1-3 Level\(^1\)

<table>
<thead>
<tr>
<th>Area</th>
<th>Number of Items</th>
<th>Coefficient Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Phonemic Analysis</td>
<td>5</td>
<td>.73</td>
</tr>
<tr>
<td>II. Word Recognition and Fluency</td>
<td>6</td>
<td>.68</td>
</tr>
<tr>
<td>III. Spelling</td>
<td>6</td>
<td>.76</td>
</tr>
<tr>
<td>IV. Independent Reading</td>
<td>5</td>
<td>.77</td>
</tr>
<tr>
<td>V. Comprehension Strategies/Teacher</td>
<td>10</td>
<td>.76</td>
</tr>
<tr>
<td>VI. Comprehension Strategies/Students</td>
<td>4</td>
<td>.77</td>
</tr>
<tr>
<td>VII. Writing</td>
<td>4</td>
<td>.75</td>
</tr>
<tr>
<td>VIII. Daily Assisted Reading</td>
<td>4</td>
<td>.90</td>
</tr>
<tr>
<td>IX. Reading Outside School</td>
<td>1</td>
<td>N/A</td>
</tr>
<tr>
<td>Total Form</td>
<td>45</td>
<td>.92</td>
</tr>
</tbody>
</table>

Based on 80 Classroom Observations
Table 3

Correlations Between Major Program Implementation Measures and Major Outcome Measures – Grades 1-3, Spring 2001

(Average n=87)

<table>
<thead>
<tr>
<th>Major Implementation Measures (Total Score)</th>
<th>Major Outcome Measures (Total Score)</th>
<th>Language Arts End-of-Level Tests</th>
<th>Student Accomplishment Forms</th>
<th>Elementary Reading Attitude Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profile of Scientifically-Based Reading Instruction</td>
<td></td>
<td>.54**</td>
<td>.37**</td>
<td>.34*</td>
</tr>
<tr>
<td>Staff Developer Reading Strategy Form (High Inference Form)</td>
<td></td>
<td>.50**</td>
<td>.36**</td>
<td>.22</td>
</tr>
<tr>
<td>Teacher Program Description</td>
<td></td>
<td>.30*</td>
<td>.29*</td>
<td>.10</td>
</tr>
</tbody>
</table>

*p<.05

**p<.01