

Measuring Program Implementation: A Tale of Two States

James A. Salzman
Cleveland State University
216-875-9977
j.salzman@csuohio.edu

Tania Jarosewich
Censeo Group LLC
440-788-2321
Tania@CenseoGroup.com

Sharon A. Brown
Cleveland State University
216-875-9977
s.a.brown54@csuohio.edu

Alisa Dorman
Oklahoma State Department of Education
405-522-3395
alisa_dorman@sde.state.ok.us

A Paper Presented at the American Educational Research Association,

New York, NY, April, 2008

Introduction

Reading First, a federally funded effort to improve early reading skills among students in participating schools nationwide, provides resources and guidance to help ensure that children read at or above grade level by the end of third grade. The three priorities of the Reading First program are to improve the quality of classroom instruction, to base instruction on scientifically proven methods, and to provide professional development for educators in reading instruction. In addition to emphasis on these three priorities, the screening and diagnosis of reading difficulties along with targeted instruction are expected to contribute to higher reading achievement among Reading First grant participants.

The districts that are participating in the Reading First grant program are among the lowest-performing and highest-poverty districts in each state. In the years prior to Reading First, many of the Reading First districts in Ohio and Oklahoma, the two states on which this paper focuses, had previously received other federal and state grants that had the same goal of improving reading instruction and student achievement. An important difference between earlier reform efforts and those represented by *NCLB*'s Reading First is that "schools are accountable not for *delivering* education to students, the historical norm, but for actually educating them – and to high academic standards" (Chubb, Linn, Haycock, & Wiener, 2005, p. 10). Although other grant programs required accountability for spending, none has required the level of accountability and monitoring that is integral to the Reading First program.

What the Research Says about Monitoring Fidelity of Program Implementation

Typically, state and federal agencies audit schools for meeting the fiscal aspects of a grant's expectations and have developed processes for schools to report financial information for review but examination of program implementation is less common (Melde, Esbense, &

Tusinski, 2006; Zvoch, Letourneau, & Parker, 2007). Evaluations of grant implementation that do occur often rest with internal evaluation units or external evaluators hired to examine the extent to which the grant met expected outcomes. Often funding agencies are satisfied with allowing an evaluation of grant-funded programs to describe schools' success in meeting goals and, implicitly then, assuming fidelity to the expectations of the grant. However, without measuring fidelity of implementation, the connection between grant activities and outcomes is weak. Drawing the conclusion that a program has not been effective without adequately examining whether the program was fully implemented has been called Type III error (Kalafat, Illback, & Sanders, 2007). That is, observed findings are erroneously attributed to the studied intervention (Dobson & Cook, 1980).

Many strategies to measure the performance of schools are based on research that has been done outside of the field of education (see for example, Horst et al., 1974; Schmidt, Scanlan, & Bell, 1979; Wholey, 1998, 2001). And only in recent years has there been a concerted effort to measure the fidelity of the implementation of programs for the purpose of evaluating the program's impact. Unfortunately, no consensus currently exists on what fidelity of implementation is or what it exactly involves (Dusenbury, Brannigan, Falco & Hansen, 2003; Fullan & Pomfret, 1977; Scheirer & Rezmovic, 1983), though recent efforts are working to define the field (see Dane & Schneider, 1998), especially as it relates to program implementation across multiple sites (see Ruiz-Primo, 2006; Zvoch, Letourneau, & Parker, 2007).

Before policy makers can judge, even inferentially, whether or not a particular program instituted at the school level across multiple sites affected student achievement, they must determine the program's essential components. Five components of fidelity that evaluators must determine based on Dane and Schneider's (1998) model include : 1) whether the program was

followed as it was designed; 2) whether all of the elements of, and resources necessary to carry out, the program were in place; 3) whether the program was delivered consistently across sites; and 4) whether teachers and students received the expected-quality “dosage” of the program. Each of these issues is critical in identifying elements of a program that may be making a difference in student achievement (Dusenbury, Brannigan, Falco, & Hansen, 2003; Melde, Esbense, & Tusinski, 2006). Measuring the impact of each of these elements can also help identify which program elements to sustain after a program ends or how to replicate a program’s success (Kalafat, Illback, & Sanders, 2007).

The Importance of and Potential for Monitoring Fidelity of Program Implementation in Reading First

Over the past 5 years, the federal government has provided states with nearly \$5 billion to implement Reading First. At least 80% of that money has gone directly to school districts to support their implementation of the program at the school level. Each state has been able to use up to 20% of their state’s allocation for administering the program, providing professional development in effective reading instruction and technical assistance to support implementation, and evaluating the impact of the program. At each state, the Reading First programming is occurring in multiple school districts across the state. In many states, professional development, technical assistance, and evaluations have spanned these multiple sites.

Recognizing that a multiple site implementation provides unique opportunities for measuring fidelity of implementation, Zvoch et al. (2007), identify several concerns that must be assuaged. In this paper, we focus on two of the concerns. First, there is a plethora of research (see Cook, Carey, Razzano, Burke, & Blyler, 2002; Leff & Mulkern, 2002; Rog & Randolph, 2002) that notes that multiple stakeholders must work across multiple sites to be certain that all

participants receive appropriate training, support and resources to implement the program as it was intended. This concern is minimized in Reading First in most states because of the states' having sufficient monetary support and a commitment to provide consistent training and support across the state. The second concern that must be addressed for successful multi-site program evaluation is the "systematic site-based collection of data that index location, provider, and target characteristics and the fidelity with which the planned treatment intervention was delivered to recipients" (Orwin, 2000 as cited in Zvoch et al., 2007). This concern, of collecting systematic data regarding implementation, is the focus of this project.

Objectives of the Paper

In the Guidance Document (USDE, 2003), states engaged in Reading First were told that "Each State educational agency that receives Reading First funds must assess and evaluate, on a regular basis, the progress of local educational agencies that receive subgrants in meeting the goals of the Reading First program" (p. 18). The guidance further identified four areas that must be addressed: 1) implementation at the state level; 2) achievement gains by students; 3) program effectiveness on both the state's and schools' parts; and 4) reducing the number of students reading below grade level. Clearly, this invocation provided a great deal of autonomy on the parts of states to determine the best way to meet these expectations.

This paper provides two different ways of meeting these federal expectation within Ohio' and Oklahoma's Reading First programs. The authors compare and contrast the two states in four areas. First, they describe the process of developing program monitoring instruments. Second, they describe the implementation of the measurement process Third, they describe the use of the measurement feedback, especially as they related to funding decisions. And, finally, they reflect on the convergences and divergences of the different processes.

Program Monitoring Expectations

Context in Ohio

Reading First – Ohio is built upon three program standards based on the current knowledge bases of scientifically-based reading research (SBRR) (National Reading Panel, 2000; National Research Council, 1998) and professional development (National Staff Development Council, 2001: (a) Standard One: Professional Learning, (b) Standard Two: Comprehensive and Coherent Program Design, and (c) Standard Three: Systematic and Explicit Reading Instruction. These standards are further described in a series of 18 indicators (See Appendix B). that describe the programmatic aspects that districts are implementing in their Reading First professional development, instructional strategies, and materials. The *Reading First – Ohio* application states that, “[e]ach program standard must be fully met to create a high quality reading program that will result in higher reading achievement for all children” (Ohio Department of Education, 2003, p. 4)

The *Reading First - Ohio* grant indicated that school level grant implementation would be measured through a “rubric reflective of the state’s accountability system that rates implementation as showing *little or no progress, minimum progress, and adequate progress*” (Ohio Department of Education, 2003, p. 157). Regional consultants, independent contractors with the project, provided technical assistance to schools as they implemented their state-approved grant within the federally approved guidelines. Each consultant worked with specified schools and districts to review fidelity of implementation three times in each year of the first three years of the grant; this was modified to twice a year in the fourth year, a reduced funding year for the districts as they worked to sustainability. School district personnel, generally resource coordinators assigned to individual schools, gathered artifacts and wrote summary

statements for each of the 18 indicators to provide evidence of implementation for the document review. Each school was required to show progress in meeting grant requirements, with the stipulation that they would lose Reading First grant funding if they did not show progress toward full implementation after their second year in the Reading First program. The intent of using this process was to provide districts numerous formative assessments to help them improve program fidelity prior to the high-stakes assessment in the spring of their second year in the program. In other words, the grant identified high expectations that districts would implement their grants with fidelity and equally high expectations that the *Reading First – Ohio Center* (Center) would design and utilize a rubric that would hold districts accountable for attaining and sustaining fidelity.

Context in Oklahoma

In contrast to Ohio's grant proposal, the Oklahoma Reading First grant did not include specific information about how the Oklahoma State Department of Education (SDE) would monitor program implementation or evaluate district and school adherence to grant requirements. Oklahoma's Reading First grant specified that the external evaluator was to develop a system by which to measure the level to which school districts and schools implemented their Reading First grant. Oklahoma's grant began in the 2003-2004 school year but the external evaluator was not hired until the spring of 2006. In the first three years of the grant, Reading First Reading Specialists, employees of the State Department of Education, assisted schools and districts in their implementation efforts. The Reading Specialists visited schools five times each year to support local grants through technical assistance and on-site consultation regarding professional development. Although these visits helped the SDE to monitor implementation for and understand statewide technical assistance needs, they did not provide a systematic way to

measure strengths and weaknesses in particular grants to made decisions about whether the level of grant implementation was high enough to warrant continuation of grant funding.

Upon the start of the grant evaluation in Oklahoma, evaluation staff and the State Department of Education team discussed procedures by which to implement a statewide monitoring system that would allow for the state staff to make the high-stakes decision of which districts would be continued in the Reading First grant.

Instrument Development

The Reading First – Ohio Program Monitoring Tool (PMT).

The Center was charged with developing the Reading First Program Monitoring Tool (PMT) to evaluate school and district implementation of *Reading First - Ohio*. While the PMT was a tool that was developed to also assist school personnel with self-reflection regarding how well they were meeting the expectations of the grant, the major function was to determine district and school compliance with the three standards identified in Ohio's application to the United States Department of Education. These standards were: professional learning; comprehensive and coherent program design; and systematic and explicit reading instruction. The PMT, and the protocols associated with its use, provides a system for schools to provide documentation of the fidelity of implementation of the program standards. This document review process was judged to be the most efficient, given the number of Reading First schools while providing the schools with the ability to display a chain of evidence that supports their claim of implementation.

Initial Development. Development of the Program Monitoring Tool (PMT) spanned over one and one-half years and began with the work of a contractor (Evaluation & Research Associates) who, prior to the creation of the Center, developed a prototype rubric for assessing

districts' success at meeting the three standards that were part of Reading First grant applications. The contracted development team, comprised of content experts in reading and literacy, reviewed the three program standards, each of the 18 indicators (see Appendix B) subsumed under the program standards, and the descriptions of the standards as explicated in the United States Department of Education (USDE) grant. As much as possible, the items of the PMT used verbatim or with minor changes the language of the grant. The original version of the PMT was developed in spring 2003 as a five-point scale with the five performance levels taken directly from the Ohio grant to the USDE.

Content Validation Process. Because of its high-stakes nature as a method of collecting data that would contribute to decisions about whether or not to continue a district's grant funding, Ohio's PMT underwent a lengthy development and rigorous review process. The instrument was reviewed by the state director of *Reading First-Ohio*, modified based on feedback, and sent for review to the Central Regional Reading First Technical Assistance Center (CRRFTAC) at the University of Texas in November of 2003. While being careful to state that they were not validating the content of the tool, reviewers at CRRFTAC confirmed that the PMT appeared to be aligned with the program standards in Ohio's grant to the USDE. Minor revisions were made based upon feedback from CRRFTAC reviewers.

Anticipating the piloting of the instrument in the winter of 2004, the Center delivered information and training sessions for district-level Reading First staff (i.e., District Coordinators and principals), in December 2003 and January 2004. Participants expressed overwhelming agreement that the PMT, with its 5-point scale, would be useful for staff to self-asses their programs and for others to evaluate district-level program implementation. However, prior to implementation of the PMT, the State Director halted the process of piloting, and during spring

2004, changed the instrument from a five-point to a three-point scale, in accordance with her understanding of the reporting requirements of Ohio's USDE-approved *Reading First - Ohio* grant application (Ohio Department of Education, 2003, p. 157). Center staff agreed that the instrument could maintain its basic integrity and format with a three-point scale. These changes delayed the piloting of the instrument, a delay that the Center staff felt was detrimental to the first cohort of schools in the program.

In the early spring of 2004, the Center collaborated with ODE to list additional sources of evidence that districts could use to meet the standards of the three-point scale. ODE sent out the instrument for a second review by CRRFTAC, this time for format and clarity. After that review, both the State and the Center were comfortable with the instrument's shape and form

In late spring 2004, several schools volunteered to pilot the tool to review both the efficacy of the tool and the site visit protocol. The schools provided the Center with documents that described the schools' level of implementation of each of the indicators. ODE Reading First staff, Center staff, and Reading First field faculty (university professors who deliver professional development to the Reading First school personnel), made minor changes to the PMT based on the documents that schools provided. This lengthy development process provided all stakeholders with reasonable estimates of content validity, as well as confidence in the viability of the process.

The Oklahoma Reading First Monitoring Tools

The process used to develop monitoring tools for Reading First Oklahoma was different from that employed in Ohio's Reading First project. Because Oklahoma's Reading First grant did not identify the method by which the state was to monitor implementation, Censeo Group evaluation staff and the Oklahoma State Department of Education (SDE) team collaborated to

develop the procedures by which to implement a statewide monitoring system. The instruments developed to measure implementation were a school self assessment form and protocols and procedures for school site visits.

The first tool that the evaluation team developed in the fall of 2006 was the Oklahoma Reading First School Self-Assessment (OK RF Self-Assessment). The OK RF Self-Assessment was designed to help school Leadership Teams determine the degree to which their school was implementing the elements that they had identified in their application for a Reading First grant: instructional strategies, programs and materials, intervention strategies, instructional assessments, instructional leadership, district and site-based professional development, and evaluation strategies and sustainability. The goal was for leadership teams to gain an understanding of grant implementation in fall 2006 to address areas of deficiency and improve program outcomes by spring 2007.

The categories and the specific items of the instrument were based on the rubric that the SDE used to review the initial school district Reading First grant applications. The 26-item tool included some items that asked schools to indicate the presence of implementation with a *yes* or *no* response and others that asked respondents to choose one of three statements (*not at all, to some extent, consistently*) to best describe the level of implementation of that element.

Respondents could provide comments or examples to support their ratings.

Implementation of the Measurement Process

Ohio - Implementation at the District and School Levels

Staff responsible for overseeing the monitoring created a protocol for administering the instrument, engaging in data verification, providing several levels of oversight and checks, and giving districts feedback that they could use to adjust their programs. The site visit protocol

provided clear guidelines at all points in the monitoring system to facilitate the timely review of monitoring data and sharing of that information among stakeholders

On-Site Monitoring of Implementation in Ohio.

The system and methods of program monitoring sent the message that Reading First was being implemented at the building, not the district, level. Although monitoring was conducted through a document review, and it was not necessary for it to occur at school sites, staff noticed immediately a qualitative effect of conducting the monitoring in buildings. Regional consultants visited Reading First sites to steep themselves in how the program played out at that level, though in the first year of program monitoring, many of the site visits took place at a central location in the district for time efficiency.

The first set of site visits lasted between three and four hours, but as regional consultants became more proficient at using the tool, and school personnel became more knowledgeable about both the implementation at the school level and adept at accurately portraying this in artifact and summary statements, the reviews were streamlined to about two hours. Throughout the process, school-level staff members commented that the length of time that the regional consultants spent in conducting the review communicated to principals and district coordinators an unspoken but implicit message of the high expectations for meeting program standards.

Oklahoma - - Implementation at the District and School Levels

School Self-Assessment.

The Oklahoma Reading First Director introduced this school self-assessment instrument at the September 2006 Leadership Team meeting. At the meeting, the Director instructed the teams on how to use the form and the form itself included directions for rating the items. Leadership Team members were to complete the assessments independently, meet to discuss

differences between ratings, and come to consensus about the level of implementation. The discussions were intended to help teams to identify the strengths and weaknesses of their own implementation in order to improve the level at which the school was addressing each element. The teams were asked to submit the completed forms to the external evaluator. Of the 73 participating Reading First schools, 67 did so. The external evaluation team compiled the results and presented them to the SDE. In addition to its anticipated usefulness to the schools, the OK RF Self-Assessment also provided the SDE with its first overall view of implementation, albeit self-reported, across the state.

Oklahoma Reading First Implementation Site Visits.

The way in which fidelity of implementation was measured for the purposes of continuing districts' grant funding was through implementation site visits conducted by the external evaluation team and the SDE Reading Specialists to all 73 Reading First schools in Spring 2007. Censeo Group staff visited 15 schools with SDE Reading Specialists accompanying Censeo Group staff on these visits to become familiar with the procedures and instruments used to collect information. The SDE Reading Specialists visited the remaining 58 schools. All visits were scheduled in advance with school-based Reading Coaches helping to develop the visit schedule. Leadership Teams in each school received information about the site visits prior to the visits.

Site visits included classroom observations in randomly selected classrooms, teacher focus groups, administrative staff interviews, and reviews of extant data. Censeo Group developed an observation form and teacher, principal, and Reading Coach interview protocols to allow for systematic collection of implementation information during the site visits. Censeo

Group staff provided training for SDE Reading Specialists in using the site visit observation forms and in submitting summary data to Censeo Group.

Each face-to-face interview lasted between 20 and 40 minutes and interviews were conducted either individually or in groups. Interviews included questions about the key evaluation areas related to the implementation of Reading First, including Reading First staff perceptions about the ways in which they have implemented the grant and plans for changes in implementation for the upcoming school year. In addition to allowing site visitors to clarify information obtained through observations and document review, the interviews gave participants the opportunity to identify additional relevant issues.

Interview data were analyzed to answer the evaluation questions and to uncover intended and unintended outcomes associated with the implementation and perceived outcomes of Reading First. Themes identified by each interviewee, connections between themes, and consistency in themes were examined.

Classroom observations conducted during site visits focused on the extent to which the core curriculum was used in instruction, evidence of instruction in the five components of reading, and application of differentiated instruction. Classroom observations were conducted during the 90-minute reading block in classrooms selected at random at each grade level. If more than one teacher taught at a particular grade level in a school, site visitors randomly chose which classroom to observe. If only one teacher taught at a grade level in a school, site visitors observed that classroom. Site visitors collected information through the Reading Instruction Observation (RIO) Form about ten instructional practices: explicit instruction, differentiated instruction, classroom management, efficient use of time, student engagement, adherence to scope and sequence of the core program, alignment of materials with the core program,

adherence to the 90-minute reading block, instructional grouping, and five components of reading. The elements of the observation form were based on current research in the area of reading (National Institute for Literacy, 2001; National Reading Panel, 2000; Simmons & Kame'enui, 2002), professional development (Guskey, 2000; Sparks, 2002), and school reform (National Association of State Boards of Education, 2002; Waters, Marzano, & McNulty, 2003) and collected information of a variety of areas of implementation of Reading First.

After each school site visit, site visitors rated the observed areas of practice on a five-point scale (1= very weak or very low, 2 = weak or low, 3 = good, 4 = strong, not perfect, 5 = excellent, nearly perfect). Mean ratings for the areas were calculated for each school and for all Reading First schools. Site visitors submitted all site visit data to Censeo Group who then summarized and compiled those data. In June 2006, each school received a summary of the ratings for the school as well as the statewide average rating for each area.

Use of the Measurement Feedback

Ohio - Providing Feedback in "Real Time"

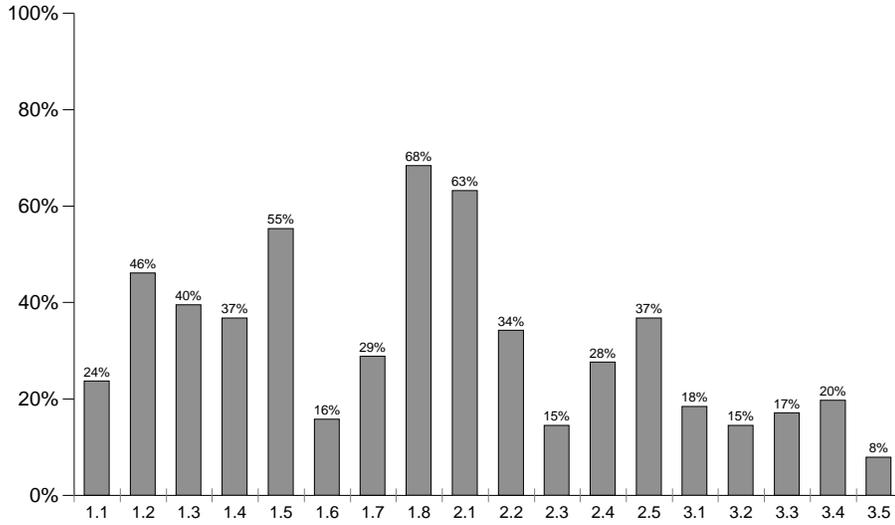
Though there was some time lag between the time that regional consultants collected program monitoring information and the time that they provided the report, in general, schools received feedback in a timely manner through a web-based reporting system. One of the problems with evaluation and research is that educators are good at collecting data but not always as good at communicating what those data mean. Most egregious on this account are the turn-around times on state and national tests. By the time schools receive information, it is usually too late to do anything meaningful for the students who were tested. The *Reading First - Ohio* program monitoring feedback has been provided to schools, on average, within 2 to 3 weeks of the visit. The regional consultants provided the schools with the same written summary

that they submitted to the Center. The formal report is intentionally succinct, containing details necessary to justify the regional consultant's judgment according to the criteria. Once the Center electronically submitted the report to the district, the regional consultant met with the district coordinator and building personnel and provided in-depth feedback and technical assistance to schools to improve their implementation based on the report.

Results from two reporting periods are indicative of the changes that districts made in response to technical assistance provided based upon the PMT process (See Figures 1 & 2). The percentages that are shown are the percent of schools that made *adequate progress*, the highest rating possible, on each of the indicators. Figure 1 shows the percentages of schools that made adequate progress on each indicator in the Fall of 2005, the first reporting period when all three cohorts were in place. Early results showed that schools were more able to demonstrate progress on their professional development being integrated within a comprehensive change process (Indicator 1.8) and linking their reading goals to state standards (Indicator 2.1) but were not as able to use effective grouping practices (Indicator 3.5). This is contrasted with the results from the Spring of 2006 after regional consultants had a chance to work with districts and the schools became more aware of what the expectations were (see Figure 2). All but two indicators were being adequately addressed by over 80% of schools in Reading First Ohio.

Figure 1

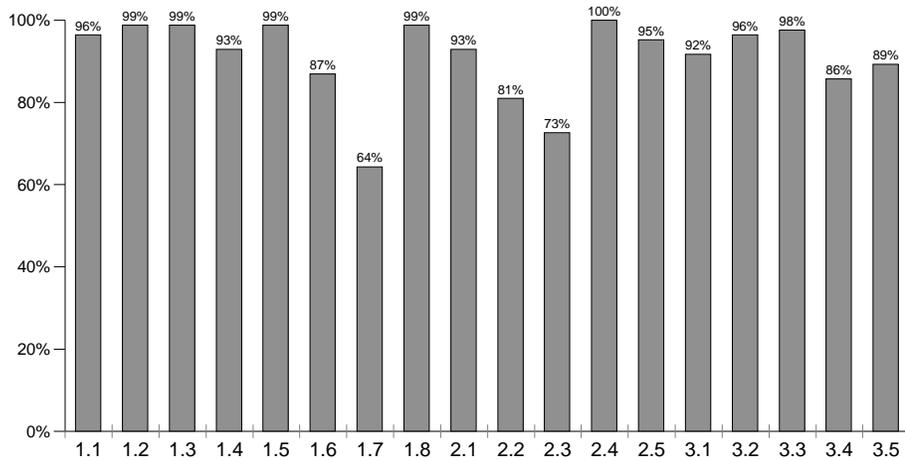
Ohio's Percentage of Schools Meeting Adequate Progress Expectations in the Fall of 2005



Note: N = 120 schools

Figure 2

Ohio's Percentage of Schools Meeting Adequate Progress Expectations in the Spring of 2006



Note: N = 120 schools

The Role of Program Monitoring in Funding Decisions.

Given the enormity of the fiscal implications for districts of being discontinued from the Reading First grant based on poor implementation, discussions about the role that program monitoring would have in terms of continued funding for Reading First were thoughtful and far-reaching. In April of 2005, the Director of the Office of Literacy at the Ohio Department of Education gained approval from the USDE for a standard for making Reading First funding decisions (Davids, personal communication):

The building must have earned a minimum score of 90% (at least 49 of 54 total points) on the Program Monitoring Tool in May of the second year of implementation. (“Little or No Progress” on an indicator = 1 point, “Minimal Progress” on an indicator = 2 points, and “Adequate Progress” = 3 points.) If the building fails this requirement, then funding will be terminated for year three.

Up to this point, Center staff had been preparing district coordinators to expect that they would have to meet *adequate progress* on all 18 indicators. Although the instrument designers never intended for the PMT to become an interval scale, one that could be converted into a percentage, they were not opposed to this use. The standard was clear and measurable and districts seemed to appreciate the extra cushion.

Modifying the Process for High-Stakes Monitoring.

In the first five program monitoring assessments in which each school participated, regional consultants conducted the site visit at the school building or within the district. However, for the sixth program monitoring for a school, the program monitoring that was tied to the decision to continue or discontinue a school, districts sent their PMT materials to the Center

in early June. In 2005, the end-of-the year PMT review was conducted for 65 schools in 12 school districts. By this time, many schools had begun to streamline their documentation while others wanted to be certain that they included all possible evidence of implementation. Some schools sent for review as few as two three-inch binders while one school sent 18 two-inch binders, one binder for each indicator. During the week-long review process, two independent reviewers rated each school's materials and sent final reports to the State Department of Education to await student achievement data, which would influence the final funding decision. The independent reviews by two different consultants in the spring of 2005 showed that consultants were consistent ($r = .902, p < .001$) in their judgments of the documents that they reviewed in contrast to concerns raised by several school districts that regional consultants did not always apply the instrument consistently across the board (Westat/Learning Points Associates, 2005). This level of inter-rater reliability was a necessary pre-condition to meeting the expectations of using PMT scores for high-stakes accountability measures as required by the State's definition of progress. In the spring of 2006, with a cohort of only six districts and 21 schools, regional consultants were able to attain an even higher coefficient of consistency ($r = .996, p < .001$).

Results in Oklahoma Reading First

School Self-Assessments.

Descriptive statistics were calculated for each item of the OK RF Self-Assessment. Initial results were reported to Reading First Leadership Teams at the November 2006 Reading First Coaches' meeting to help the teams interpret the findings of their own review and use results to guide program implementation.

Results of the self-assessments suggested that schools were confident that they were meeting the requirements of the grant in the majority of implementation areas. Almost all of the school Leadership Teams indicated that teachers adhered to the core program, provided a 90-minute uninterrupted, daily, instructional block, used Reading First approved assessments, and had access to assessment data. The majority of teams also agreed that their reading materials were integrated into a cohesive program, they provided effective intervention, schools successfully supported teachers who were struggling with Reading First implementation, and teachers collaborated to improve instruction. The items that were ranked the lowest included teachers' use of effective classroom strategies and provision of differentiated instruction, suggesting that schools consistently agreed that these elements were not as well-implemented as they could be.

Implementation Site Visits

The average site visit rating was 3.88 ($SD=0.51$) on a scale of one of five. This suggested that schools were, in general implementing their Reading First grants with fidelity. Four schools' implementation was significantly lower than the mean, falling two standard deviations below the average rating.

The Oklahoma SDE determined that Cohort 1 schools that received an overall implementation rating of two standard deviations below the mean would be included in the list of schools considered for discontinuation(see Table 1). These four schools also had achievement scores that did not meet the continuation criteria. An additional 13 schools met implementation criteria but did not meet their achievement goals. These 17 schools were discontinued from the Reading First grant. An additional 15 schools received an average implementation rating between one and two standard deviations below the mean. They were not considered for

discontinuation, but were selected to receive state-level technical assistance and support in the 2007-2008 school year.

Table 1. Discontinuation Decisions

N Schools	Implementation goal	DIBELS goal	ITBS goal
4	Not met	Not met	Not met
2	Met	Not met	Not met
11	Met	Met	Not met

Each Reading First school received a school report at the regularly-scheduled June 2007 leadership meeting. Each school report included the mean score for that site as well as the average score across the state for each category. Appendix A includes an example of a summary report that each school received.

Standards for Continued Funding in Oklahoma

In Oklahoma, similar to Ohio, the high-stakes decision regarding which districts would continue in the Reading First grant also includes the requirement of full grant implementation and student achievement outcomes. According to Oklahoma’s state Reading First grant, schools had to show effective grant implementation, improvement in student performance on the ITBS assessment, and improvement in student performance on DIBELS assessment to continue to receive Reading First grant funding after their third year of implementation. Prior to the third year of grant implementation and the hiring of an external evaluator, the Oklahoma State Department of Education did not have a way to measure the level of implementation of Reading First among the participating grantees and had not identified the expected growth on ITBS or DIBELS assessments to be used in continuation decisions. Over the course of the 2006-2007 school year, Oklahoma Reading First staff collaborated with Censeo Group evaluators to develop the criteria for Cohort 1 schools (those in their fourth year of Reading First implementation) to

continue to receive funding. The USDE approved the following criteria in the spring of 2007: evidence of full implementation, and 60% of students at or above 50th percentile or 30% decrease in students below 50th percentile on the ITBS, and 60% of students at or above 50th percentile or 30% decrease in students below benchmark on the DIBELS.

Educational Importance of the Study

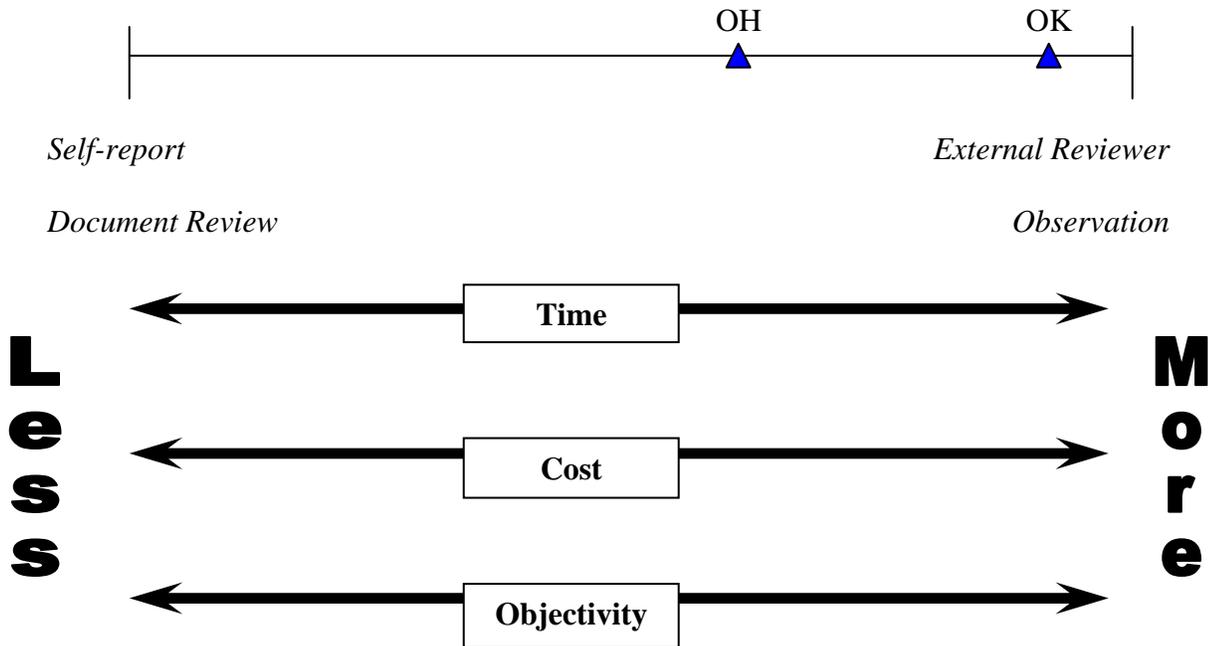
Ohio's PMT and Oklahoma's self-assessment and implementation site visits provided each state department of education and other Reading First stakeholders with information for making decisions about which schools had implemented their grant sufficiently enough to continue to receive Reading First funding, information to plan for technical assistance and professional development, and an overall review of fidelity of implementation across the state and at the school level. The program monitoring information also provided important information for Reading First school teams to identify areas of strength in their implementation and areas in which they could improve. This was particularly important in Oklahoma since self-assessments conducted by school teams suggested that the majority of schools thought that they were implementing areas of their Reading First grants with fidelity, results that were not necessarily consistent with site visit information, and not surprising according to other researchers (Dusenbury, Brannigan, Falco, & Hansen, 2003; Lillehoj, Griffin, & Spoth, 2004). Most importantly, these are two similar models with substantive differences that may allow other evaluators an opportunity to consider the relative strengths of each as they choose or develop fidelity of implementation measures, particularly when evaluating grants that are implemented in multiple sites and require data collection methods that will gather consistent data across sites.

The Implementation Evaluation Continuum

If one thinks of evaluation methods on a continuum of choices, each of which affects their capacity for change, impacts on budget, and values different types of information, then one can choose purposefully among these choices (see Figure 3).. Each state chose a different point on the continuum of methods for evaluating implementation. Each method, obviously, comes with advantages and costs associated with it. For Ohio, the program directors followed the requirements of the original grant to the USDE. The PMT aligned with the expectations of the grant and, with training, regional consultants were able to apply that tool consistently. When district coordinators were surveyed at the end of the third year of the grant, they cited accountability of meeting program standards through the document review as the third most significant grant element, just behind reading coaches and professional development, that led them to successful implementation of a stronger core reading program. As described earlier, the program monitoring was applied three times per year with regional consultants spending between six and eight hours reviewing the document folders, analyzing evidence of implementation, and writing a report (See Appendix B for an example). The document review-based monitoring enabled regional consultants to obtain a whole-school perspective in a relatively short period of time, while minimizing the costs for such a comprehensive view of the school-level program. While the regional consultants maintained an appropriate professional objectivity in reviewing the data, they also provided these districts technical assistance and so they may not have been as unbiased as an outside observer who did not have a technical assistance role. That said, the end of year analysis of inter-rater reliability, as described previously, supported their ability to minimize their personal biases and concentrate on the criteria and evidence before them.

Figure 3

Implementation-Evaluation Continuum



One last point that was unique to Ohio and should be mentioned here is that the program monitoring protocol complemented other aspects of the professional development, technical assistance, evaluation system. While regional consultants did not do site-based classroom observations as part of this protocol, there were two other groups who were in classrooms. The Center provided mentor coaches, who supported teachers and school-level coaches as they engaged in classroom practices, to all schools. And Westat/Learning Points Associates, the external evaluators, did classroom observations for their external evaluation report to the state. The Center staff responsible for program monitoring made a deliberate choice not to “step on” these other participants and to use the document review for assessing fidelity of implementation.

Similar to Ohio’s practices for measuring program implementation, Oklahoma’s practices were intended to provide schools and the state department of education with information about fidelity of implementation at school and district levels. The methods used in Oklahoma, similar

to those in Ohio, included a visit to each Reading First school, although in Ohio the visits were conducted to review documents, while in Oklahoma, site visitors observed instruction and conducted interviews. In both states the school visits provided the state departments of education with information about the context in which the grant was being implemented and an overall assessment of the fidelity of implementation. The goal of the school visits in both states was to assess *how well* each school was implementing the elements of the grant, not simply to provide an assessment of whether the elements were present or not. The focus on the quality of implementation was important for technical assistance and professional development planning. School visits, although falling on the higher end of the continuum in terms of time and cost, provide critical information about the quality of implementation. The site visits in Oklahoma were conducted in one day, which provided a snapshot view of classroom instruction. In contrast, Ohio's method allowed school districts to compile evidence that spanned the three month period between program monitoring visits. Although the observations conducted on one particular day in the Oklahoma visits might not be representative of overall instruction, the interviews and opportunity for schools to offer documentation to support their claims provided a way to verify the validity of the observation data. Also, in Oklahoma, additional information about implementation was collected through participant surveys. The results of the surveys were not used in the decision to discontinue schools but were included in descriptions of successes and concerns of implementation statewide (CBAM REFERENCE). The balance between cost, time, and the collection of data that can give a picture of the quality of implementation and the regularity with which a program is implemented with fidelity results in the use of various measures and methods.

Also, the site visits in Oklahoma were conducted not only by external evaluation staff but also by state department of education staff, who already had a relationship with the schools through their frequent school visits and technical assistance support. The training, consistent procedures, and common protocol resulted in high inter-rater reliability among the observers. The Reading Specialists noted that the protocols and procedures that they used in the implementation site visits resulted in more comprehensive and useful information than that which they had collected in other visits to the schools. For example, previously they had observed teachers but had not interviewed them about instructional practices. The interviews provided the Reading Specialists with insights that helped them to structure technical assistance and further professional development. In this instance, the support of an external evaluator to develop the site visit protocols also supported internal evaluation efforts and built capacity for evaluation.

Transparency and the Importance of Protocols

On a larger scale, the methods developed for monitoring fidelity of program implementation in these two statewide federally-funded grants provide an example of how other evaluators can work with their stakeholders to develop effective and efficient monitoring systems. Maybe more importantly, though, the examples from these two states can serve as models for schools to monitor their own efforts at improving instruction in their schools and determining prescriptions for that improvement. Elmore (2007) noted that “accountability pressure has drawn administrators more deeply into issues of instruction and many practitioners say they are ill-equipped to organize and manage around the improvement of instruction” (p. 22). This is a critical need since instructional leadership is one of the factors that predicts the fidelity of implementation of new programs (Kurki, Boyle, & Aladjem, 2006). The procedures

(structured observations, review of documents) and materials (self-assessment form and PMT) could be used by district and school-level administrators to conduct their own monitoring of program implementation and to improve the delivery of services.

Though different, both states' implementation evaluations were highly-structured and provided protocols that reviewers followed, mirroring Elmore's prescription that protocols can provide a powerful way to structure the work of instructional leadership by giving a predictable structure and separating "the person from the practice" (p. 22). As instructional leaders ask teachers to change practice, it is important for teachers to know the standards to which they must adhere and for all stakeholders, administrators and teachers alike, to separate *who* the teacher is from *what* he or she is doing. It is also important for the examination of practice to be consistent and multi-faceted, particularly when high-stakes decisions rest on the outcomes.

One Last Point: From Ritual Compliance to Authentic Implementation

As stated previously, the authors modified Dane and Schneider's (1998) five components of fidelity into four indicators that evaluators must determine are present before they can make any connection to student outcomes. It was clear that the first three indicators were attained by nearly all schools. In other words, the fidelity measures in both states were able to determine the extent to which: 1) the program was followed as it was designed; 2) all of the elements of, and resources necessary to carry out, the program were in place; 3) the program was delivered consistently across sites. What is less clear is whether teachers and students received the expected-quality "dosage" of the program.

By and large, the experiences in Ohio and Oklahoma saw the vast majority of schools at least able to demonstrate that they were complying with the relatively-prescriptive mandates of their state's Reading First grant. One of the reasons we were able to see such high levels of

fidelity of implementation was that funding was tied explicitly and clearly to *following the program*.

That is, there was clear accountability for implementing the required elements to which schools agreed when they signed up for Reading First. Without this high level of accountability, the likelihood that teachers and principals would make significant changes in practice to lead to improved student outcomes could have been diminished. The purpose of building the measures of program implementation used in Ohio and Oklahoma, then, was to provide a clear set of expectations for participants and to serve as one metric in determining whether schools and districts were successful in implementing their program. This type of monitoring helped to ensure that the soft accountability of previous grant programs and less-than-expected outcomes of those programs would not be duplicated.

Accountability to clear standards, however, is necessary but not necessarily sufficient to ensure the outcomes we all desire. In both Ohio and Oklahoma there were a number of schools that showed high levels of fidelity of implementation, but not the expected student gains. There is still a difference between ritually complying with grant expectations and authentically implementing a school reform or making adaptations in the reform that can affect outcomes (Hill, Maucione, & Hood, 2007). As indicated by Oklahoma's experience, the tendency for staff in schools to overestimate the extent to which they are meeting standards, no matter how clearly articulated, is part of the learning curve for implementing a program with fidelity. Salzman et al. (2005) described a similar tendency to overestimate performance when teachers used assessments with clear benchmarks to compare their students' performance to a set standard. In both cases, teachers and staff in low-achieving schools may engage in a sort of peer-norming; that is, comparing their current performance or students' performance to previous performance.

This may lead to an exaggerated sense of progress or achievement because they can see how far they've come but not how far there is to go.

Works Cited

- Chelimsky, E. (1985). Comparing and contrasting auditing and evaluation: Some notes on their relationship. *Evaluation Review*, 9, 485-503.
- Chubb, J., Linn, R., Haycock, K., & Wiener, R. (2005). Do we need to repair the monument? Debating the future of No Child Left Behind. *Education Next*, 5(2), 9-19.
- Dollitt, C., & Summa, H. (1997). Performance auditing: Travelers' tales. In E. Chelimsky & W. R. Shadish (Eds.), *Evaluation for the 21st century: A handbook* (pp. 86-108). Thousand Oaks, CA: Sage.
- Dusenbury, L., Brannigan, R., Falcom, M. & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research*, 18, 237-256.
- Elmore, R. F. (2007 April). Professional networks and school improvement. *School Administrator*, 20-24.
- Guskey, T.R. (2000). *Evaluating professional development*. Thousand Oaks, CA: Corwin Press, Inc.
- Hill, L. G., Maucione, K., & Hood, B. K. (2007). A focused approach to assessing program fidelity. *Prevention Science*, 8, 25-34.
- Horst
- Kalafat, J., Illback, R. J., & Sanders, D. (2007). The relationship between implementation fidelity and educational outcomes in a school-based family support program: Development of a model for evaluating multidimensional full-service programs. *Evaluation and Program Planning*, 30, 136-148.

- Kurki, A., Boyle, A., & Aladjem, D. K. (2006). Implementation: Measuring and explaining the fidelity of CSR implementation. *Journal of Education for Students Placed at Risk, 11*, 255-277.
- Melde, C., Esbense, F., & Tusinski, K. (2006). Addressing program fidelity using onsite observations and program provider descriptions of program delivery. *Evaluation Review, 30*, 714-740.
- National Association of State Boards of Education. (2002). *The report of the NASBE study group on low-performing schools. From sanctions to solutions: Meeting the needs of low performing schools*. Alexandria, VA: NASBE.
- National Institute for Literacy (2001). *Put reading first. The research building blocks for teaching children to read. Kindergarten through grade 3*. Washington, DC: Author.
- National Reading Panel. (2000). *Report of the National Reading Panel, teaching children to read: An evidenced-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: National Institute of Child, Health, and Human Development.
- National Research Council. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- National Staff Development Council. (2001). *National Staff Development Council's standards for staff development*. Oxford, OH: Author.
- Ohio Department of Education. (2003). *Reading First Ohio*
- Salzman, J. A., Clay, D. L., Brown, S. A., Rosemary, C. A., Lenhart, L. A. (2005). To DIBELS or not to DIBELS? That is the question: Measuring DIBELS ability to predict outcome measures. *Ohio Journal of English Language Arts, 46*(1), 29-34.

- Schmidt, R. E., Scanlon, J. W., & Bell, J. B. (1979). *Evaluability assessment: Making public programs work better*. Rockville, MD: U. S. Department of Health and Human Services, Project Share.
- Simmons, D.C. & Kame'enui, E.J. (2002). *A consumer's guide to evaluating a core reading program grades K-3: A critical elements analysis*. In The Secretary's Reading Leadership Academy. Washington, DC: U.S. Department of Education.
- Sparks, D. (2002). *Designing powerful professional development*. Oxford, OH: National Staff Development Council.
- United States Government Accountability Office (2006). *Government auditing standards (GAO-06-729G)*. Washington, DC: Author.
- Waters, T., Marzano, R.J., & McNulty, B. (2003). *Balanced leadership: What 30 years of research tells us about the effect of leadership on student achievement. A working paper*. McREL. Obtained at http://www.mcrel.org/PDF/LeadershipOrganizationDevelopment/5031RR_BalancedLeadership.pdf.
- Westat/Learning Points Associates. (2005). *Reading First – Ohio: Year 2 Annual Report*. Washington, DC: Author.
- Wholey, J. S. (1998). Assessing performance measurement systems. *The Public Manager*, 27(3), 23.
- Wholey, J. S. (2001). Managing for results: Roles for evaluators in a new management area. *American Journal of Evaluation*, 22(3), 343-347.

Wholey, J. S. (2004). Using evaluation to improve performance and support decision making. In

M. C. Alkin (Ed.), *Evaluation roots: Tracing theorists' views and influences* (pp. 267-

275). Thousand Oaks, CA: Sage.

Zvoch, K., Letourneua, L. E., & Parker, R. P. (2007). A multilevel multisite outcomes-by-

implementation evaluation of an early childhood literacy model. *American Journal of*

Evaluation, 28, 132-150.

Appendix A



**OKLAHOMA READING FIRST
EXTERNAL EVALUATION**

2006-2007 Implementation Site Visit Report

12547 Ridge Road • North Royalton • OH • 44133 • 440-788-2321

Sample Elementary School

This report summarizes the results of the 2006-2007 Oklahoma Reading First external evaluation site visits that Censeo Group and/or the Oklahoma State Department of Education Reading First Specialists conducted in early 2007. The following tables include the ratings provided for your school and the statewide average ratings for elements that the site visitors observed during classroom instruction and gathered through interviews and from documentation that your school provided. Each element was rated along a 1 to 5 scale, with 1 indicating a low level of evidence and 5 indicating a high level of evidence that the component was fully implemented. The overall rating is an average of the other ratings and provides an indication of the degree to which your school has successfully implemented Reading First as compared to other schools in the state.

CLASSROOM OBSERVATIONS		
Observation Area	Your School Rating	Statewide Average Rating
Adherence to 90-minute uninterrupted core instruction	5.00	4.83
Average number of the five components observed during core instruction	2.25	2.74
Evidence that teachers followed scope and sequence of the core program	4.25	4.12
Use of explicit instructional strategies	4.00	3.60
Evidence of differentiated instructional practices	1.00	2.48
Varied grouping practices	1.50	2.16
Effective classroom management	4.25	4.08
Efficient use of time during instruction	3.75	3.76
High level of student engagement	4.00	3.94
Use of Instructional materials that are aligned with the core program and PASS	4.75	4.42
Average score for observational data	3.48	3.61

OBSERVATION, INTERVIEW, AND DOCUMENTATION

Rated Area	Overall Rating	Statewide Average Rating
Evidence that teachers followed the scope and sequence of the core program and used explicit instructional strategies	3.5	3.59
Staff access to and use of data to make instructional decisions	4.0	3.65
Clear criteria to identify and monitor effectiveness of intervention	4.0	3.99
Appropriate frequency and content of intervention	4.0	3.89
Extent to which teacher professional development needs have been met	3.0	3.71
Effectiveness of coach's support, feedback, and professional development	4.0	4.07
Evidence of plan for sustaining Reading First-aligned instruction after completion of grant	4.0	3.41
Effective monitoring of grant implementation and student outcomes	3.0	3.72
Extent to which Reading First is coordinated with other programs	4.0	3.58
Level of Leadership Team involvement in grant implementation	3.0	3.65
Level of principal involvement in Reading First grant implementation	4.0	3.78
District support of Reading First	4.0	3.90
Level of teacher engagement in the Reading First grant	4.0	3.63
Timeliness and success of addressing SDE targets	4.0	4.20
Average Rating	3.75	3.77

REQUIRED GRANT ELEMENTS		
Element	Overall Rating	Statewide Average Rating
Use of approved assessments	5.0	4.44
Attendance at Reading First professional development	1.0	3.56
Reading First Coach has obtained or is in the process of obtaining Reading Specialist certification	1.0	4.74
Average Rating	2.33	4.25

OVERALL RATING FOR 2006-2007 SCHOOL VISIT	
Sample Elementary School Rating	3.19
Statewide Average Rating	3.88

Appendix B

Reading First-Ohio Center

Program Monitoring Report #1

District:	Building:
Building Principal:	Resource Coordinator:
Data Manager:	Literacy Specialist:
 Program Reviewer:	 Regional Consultatnt
Review Date:	November 4, 2004

Standard One: Professional Learning	Little or No Progress	Minimal Progress	Adequate Progress
1.1 What evidence is presented that the district has aligned their educators' learning goals with student performance goals?	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
1.2 What evidence is presented that the district has involved educators in planning for their own professional development?	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
1.3 What evidence is presented that the professional development in reading is school-based and integral to the school operations?	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
1.4 What evidence is presented that the professional development in reading relates to individual needs and is organized, in part, by collaborative problem solving?	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
1.5 What evidence is presented that the professional development in reading is continuous with follow up and support (internal and external) for further learning?	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
1.6 What evidence is presented that the professional development in reading is evaluated based upon multiple sources of information?	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
1.7 What evidence is presented that the professional development in reading provides opportunities for developing a theoretical understanding of the reading process, reading development and the assessment of reading growth?	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
1.8 What evidence is presented that the professional development in reading is integrated with a comprehensive change process?	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Standard Two: Comprehensive and Coherent Program Design			
2.1 What evidence is presented that the district's reading goals for each grade are linked to Ohio's Academic Content Standards in Reading?	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
2.2 What evidence is presented that the district focuses instructional content at each grade level on five essential components of reading?	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.3 What evidence is presented that the learning activities are implemented using SBRR instructional materials and strategies shown to improve essential reading?	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
2.4 What evidence is presented that the district's evaluation of reading skills is based on reliable, valid assessments and is grounded in the four types of assessment information?	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
2.5 What evidence is presented that program coherence has been monitored through systematic observation?	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Standard Three: Systematic and Explicit Reading Instruction			
3.1 What evidence is presented that the district's systematic and explicit reading instruction is guided by an Assess-Revise/Plan-Teach model?	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
3.2 What evidence is presented that the district's systematic and explicit instruction focuses on the five essential components of reading (information and documentation for this standard is additionally provided in 2.3)?	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
3.3 What evidence is presented that the district's systematic and explicit instruction uses tools that reflect scientifically based reading research? Does it employ an effective core program, proven interventions and research-based supplemental services? Are teaching and learning strategies supported by scientific research?	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
3.4 What evidence is presented that the district's systematic and explicit reading instruction applies diagnostics in instruction?	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
3.5 What evidence is presented that the district's systematic and explicit instruction uses effective grouping and intervention practices?	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>

Selected Examples of Descriptive Feedback for PMT Report #1

1.3 ●○○ Little or No Progress What evidence is presented that the professional development in reading is school-based and integral to the school operations?

Evidence shows little to no progress that PD in reading is school based and integral to the school operations. Coaching logs and reading workshop agendas show that PD in reading is ongoing and reflective of teacher need, however, the sessions take place outside of the contractual day.

1.4 ○●○ Minimal Progress What evidence is presented that the professional development in reading relates to individual needs and is organized, in part, by collaborative problem solving?

*Evidence shows minimal progress that PD in reading relates to individual needs and is organized, in part, by collaborative problem solving. Michael R. White's grade level meetings show that there is a collaborative effort in place. Further documentation will validate that this process is taking place (e.g., **completed** intervention forms, identifying how students are selected - PD to address areas of need).*

1.5 ○○● Adequate Progress What evidence is presented that the professional development in reading is continuous with follow up and support (internal and external) for further learning?

Clear evidence shows adequate progress that PD is continuous with follow up and support (internal and external) for further learning. There is also strong evidence of PD being continuous with ongoing consulting (Harcourt/district PD). Documentation shows that school administrators provide support to teachers in advancing skill development.