Evaluation of Reading First in Michigan
Technical Report, #2.1
September 2005

Reading First Within the Context of NCLB:

Are Title 1 Parts A and B Compatible Federal Policies?

Joanne F. Carlisle

Ruth Athan

University of Michigan

Suggested citation: Carlisle, J.F., & Athan. R. (2005). *Reading First Within the Context of NCLB: Are Title 1 Parts A and B Compatible Federal Policies?* Evaluation of Reading First, Technical Report #2.1. Ann Arbor: University of Michigan

Reading First Within the Context of NCLB

Are Title 1 Parts A and B Compatible Federal Policies?

Title 1 of the No Child Left Behind legislation of 2001 (P.L. 107-110) has been both heralded and disparaged as legislation designed to improve the academic achievement of U.S. students. Signed into law by President George W. Bush on January 8, 2002, the Title 1 Part A section of the law holds schools responsible for improving achievement in reading and math for students in grades 3 through 8. Less well known and understood is Title 1 Part B, which establishes a federal program called Reading First. This is a 6-year, 6-billion dollar program to assist high poverty schools in improving reading instruction in the early elementary years. Part A is more fully understood than Part B (Reading First) because of the attention paid by researchers, educators, politicians, and the popular press to the Part A requirement that schools administer annual assessments to track their progress in improving students' academic achievement every year.

Title 1 Parts A and B both include the requirement that schools show progress in the percentage of students meeting grade-level expectations in reading. While both focus particular attention on the schools making the least progress in reading, Parts A and B differ in many important respects—including not only how progress is measured, but also more broadly in the nature of the programs. Part A principally holds schools accountable for the progress of their students, with a structured system of consequences (termed "school improvements") for schools not meeting achievement gains, whereas Part B (Reading First) provides substantial direction and support for the improvement of the teaching of reading. The U.S.

Department of Education characterizes Reading First as "the academic cornerstone of No Child Left Behind, which recognizes the importance of both improving student reading achievement and implementing programs and strategies scientifically proven to be effective" (U.S. Department of Education, 2002, section A-3).

As this quotation suggests, Reading First was designed as a program with specific provisions to realize the goal of improving students' reading achievement. Miskel and Song (2004) contend that Parts A and B emerged from different political settings and contexts. They maintain that the Department of Education played "a very limited role in either creating or lobbying for the Reading First legislation" (p. 96); rather, an "elite among the elite… an insider group of five …played critical roles in passing Reading First" (p. 101). The different genesis of Parts A and B may partially explain the challenges that Reading First schools may face as they work to satisfy requirements of both parts of the legislation.

In our role as evaluators of the Reading First initiative in Michigan, we set out to examine the dual impact of Parts A and B on high poverty schools with high levels of underachievement in reading. In doing so, we drew on our reading of the law, supporting state and federal documents, research reports, commentaries, and our experiences with Reading First schools in Michigan. These provide a context for examining questions concerning the compatibility of Parts A and B. At the outset, we wondered what it was like for schools to implement a Reading First plan and at the same time address NCLB progress requirements and sanctions.

To lay the groundwork for this investigation, we first review the basic provisions of Parts A and B. Next, we delineate the accountability measures of both parts of the legislation, and we present profiles of Reading First schools in Michigan to show the relation of the accountability measures in Parts A and B. We examine the impact of Part A on school culture and capacity for Reading First schools. Finally, we reflect on alternate models of accountability that focus on the quality of instruction in reading as well as students' achievement.

Comparison of Parts A and B of Title 1

Basic Provisions

The goals and operational criteria of the two parts of NCLB Title 1 provide the basic framework needed for our analysis of their compatibility. The stated goals of the two parts are similar. Both seek to increase the responsibility of schools for the academic achievement of the students, especially in reading and mathematics. Schools must also demonstrate achievement gains for disadvantaged and at-risk students comparable to those of their peers. Both parts contain measures designed to improve the quality of teaching in today's schools (U.S. Department of Education, 2002). A hallmark of the NCLB Title 1 legislation is the emphasis in both Parts A and B on the accountability of states and local education agencies (LEAs) for students' academic achievement. Annual assessment of student achievement is the primary mechanism through which the federal government determines that the state and LEAs are doing an adequate job of raising student achievement. This legislation has also placed unprecedented requirements on improvement of the academic performance of students in risk categories (i.e.,

students in special education, students with economic disadvantage, and English language learners) and students from specific racial/ethnic groups. The law requires that the performance of children in risk categories and in specified ethnic/racial groups be disaggregated so as to determine whether students in these groups are making progress at the same rate as their grade-level peers. [1]

On the other hand, the means by which these goals are to be achieved are quite different. Part A requires that each state select (or design) measures of reading and math aligned with the state's curricular goals for each grade level 3-8 and administer these measures annually to grade 3-8 students, showing improvement over time in the percentage of students deemed "proficient" in these subject areas. Part A is required in order for states to receive Title 1 federal funding, whereas Part B is voluntary. While Part B has funds committed for each state based on the percentage of children from low-income families, states are required to submit a Reading First plan, and this plan had to be approved by a federal review board in order for the state to be awarded its funding. (We note that, although Part B is voluntary, as of 2005 all states had an approved Reading First plan in place.) Within a state, eligible districts and Local Education Agencies (LEAs) can apply for funding on a voluntary basis.

---

[1] The US DOE has made some changes in the initial requirements for English language learners and students with disabilities (see www.ed.gov). Other proposed changes are being reviewed by US DOE as of spring 2005.

An important difference between Parts A and B is the time span of the two initiatives. Part A requires states to formulate an Adequate Yearly Progress (AYP) plan for assuring that all students are at the proficient level in reading and math by 2014. Between 2002 and 2014, they must develop or select an assessment system and begin testing 3rd through 8th graders annually in reading and mathematics (along with high school students at one grade level). In contrast, Part B Reading First is authorized for only six years and involves younger students (kindergarten through grade three). Thus, the timeline for making significant progress in Reading First schools is relatively short. Within this six-year program, the US Department of Education oversees the implementation of the state Reading First programs through submission of annual performance reports. As stated in the law and the guidance, at the end of three years, an appointed federal review panel determines whether the state is satisfactorily implementing its Reading First program and is improving the reading achievement of the students in grades 1-3.

The provision in Part B but not Part A of funding for schools to improve reading instruction appears to be a critical difference between the two parts of the law. For Part B, states with approved grants are required to dispense 80% of their federal funding to districts and LEAs with approved Reading First plans (U.S. Department of Education, April, 2002). Among other regulations, the law requires that LEAs use their funding to provide teachers with instructional materials and programs supported by scientifically-based reading research (SBRR) and provide professional development for teachers.

Why might Part A affect the ability of schools that are awarded Reading First grants to initiate and maintain an effective plan for improving reading achievement in grades 1-3? We noted earlier that Reading First schools are eligible for funding from Title 1 Part B because of high levels of poverty and underachievement in reading. This very fact led us to expect that schools with new Reading First funding might also be coping with school improvement provisions from Part A. The consequences of not showing adequate progress for Part A might increase the challenge schools face in trying to improve students' reading skills, making it more difficult to develop school culture and capacity that are conducive to improving the teaching and learning of early literacy instruction. For those involved in the evaluation of Reading First implementation, consideration of the complexity of challenges facing Reading First schools that have not made adequate yearly progress, and therefore are facing school improvement measures, is potentially critical in order to understand the results of the accountability requirements of NCLB legislation.

Accountability: Focus on Test Results or Program Implementation?

Part A accountability requirements. To assess improvement in students' reading for Part A, states have to establish a system for determining progress, most notably through the use of a cut point to distinguish those students who are proficient and those who are not. Some states use indices that give full credit for students at or above the cut point and partial credit for those below the cut point (Linn, 2003). A cut point is an arbitrary point on a performance scale such as the

50th %ile. For Part A, it is set according to the state plan for reaching the goal of having all children meeting proficiency status by 2014.

Proficiency performance levels differ widely by state, as Linn, Baker, and Betebenner (2002) found by comparing state proficiency levels in reading and math to proficiency levels on the National Association for Educational Performance (NAEP) assessment results. For example, Linn et al. found that Michigan's standards at that time were closer to NAEP proficiency standards than were the standards of Colorado. After the passage of the NCLB law, states were required to submit AYP plans to comply with the requirements of NCLB Title 1 Part A, and in the process of developing such plans, a number of states redefined the cut point for determining proficiency so that it would be more likely that students would show gains in the percent of students proficient in reading and math as 2014 approaches. States have been allowed some flexibility in devising their AYP plan; currently, some states continue to appeal to the federal government for changes to their AYP plans (Erpenbach, Forte-Fast, & Potts, July 2003; Olson, 2005). For example, in 2004, South Carolina and Washington received permission to average two to three years of data when calculating the percentage of students scoring at proficient levels on their state tests (Center on Education Policy, October 2004). Michigan has an approved dual statewide accountability system. The state uses a school report card, giving a grade for each grade level that represents a composite for the areas assessed; in addition to computing this composite grade, a final "filter" is applied to determine whether AYP standards have been met (Erpenbach et al., 2003).

In Part A, a critical aspect of accountability is the consequences for schools that do not meet the progress requirements of the state. Schools that fail to meet the state criteria for AYP for two consecutive years are identified as needing school improvement for the following year. The first "phase" of school improvement involves choice and transportation. Students are given an opportunity to transfer to a different public school within their district, and the school identified for improvement must pay for the transportation costs to the new school (see U.S. Department of Education guide, Creating Strong District School Choice Programs)[2]. In addition, Michigan schools at this level need to develop a two-year plan to improve their students' academic performance. Schools that fail to meet AYP criteria for three consecutive years are required to offer supplemental educational services to disadvantaged students, in addition to the corrective measures listed above (see U.S. Department of Education brochure, *Extra Help for Student Success*)[3]. Schools that fail to meet the criteria for AYP for four consecutive years, in addition to the supplemental services, are required to make more extensive changes, such as extending the school year or developing a restructuring plan. Schools that fail to meet AYP criteria for five consecutive years are likely to be required to restructure the school (e.g., replace the staff, reopen as a charter school, although in many states there are not adequate funds or human resources to fully implement these changes) (Olson, 2002).

Part B accountability requirements. Part B accountability, like Part A, requires states to report students' proficiency as measured by performance on a

---

[2] Available at www.ed.gov/news/newsletters/extracredit/2004/06/0602.html
[3] Available at www.ed.gov/news/newsletters/extracredit/2004/10/1012.html

standardized test. In Part B, states are required to report annually the percentage of students in grades 1-3 reading at grade level. However, Part B does not have the strictly prescribed accountability requirements of Part A. States are not required to have a specified percentage of students that must reach proficiency levels in their Reading First schools. Rather, states are required only to show "significant" achievement gains of students reading at grade level at the end of three years of funding, as demonstrated by performance on scientifically based assessment instruments. Gains must also be evident for students in risk categories (U.S. Department of Education, 2002, section G-4).

As for consequences, as mentioned earlier, each state submits a performance report annually, and the quality of implementation is evaluated at the end of three years. The Reading First guidance indicates that states not making significant progress in "meeting the purposes of the Reading First program" might have further funding withheld or else might be provided with additional technical assistance (U.S. Department of Education, 2002, section G-7). Accountability is presented in terms of compliance with implementation requirements as well as progress in improving reading achievement. States must evaluate the progress of districts and LEAs in meeting such Reading First goals as adoption of instructional materials and assessments and implementation of state and local professional development programs.

Implications of Similarities and Differences in Parts A and B

In sum, making schools responsible for the academic achievement of their students is a core principle for Title 1 Parts A and B. Both adhere to the philosophy

that achievement testing, carried out annually, can be used to evaluate the effectiveness of instruction and foster schools' efforts to improve reading achievement. On the other hand, the relation of assessment and instruction is markedly different for Parts A and B. For Part A, the test used to determine AYP is generally linked to the state's curricular benchmarks. In other respects, Part A leaves the process of coordinating assessment and instruction to schools. Because no financial or instructional resources are provided to help schools improve the alignment of instruction and assessment, the law appears to assume that schools have the resources they need to improve instruction—they just need to make better use of them. This has possibly drastic effects on high poverty schools, which are most likely to have fewer certified or high-quality teachers (Darling-Hammond, 2004). For Part B, while funding is available for only those schools serving high poverty and underachieving students, considerable resources are made available to eligible schools so that they can improve reading instruction. Furthermore, annual reporting focuses not only on improvement in reading achievement but also evidence of effective implementation of curricular and professional development programs.

Compatibility of Parts A and B: The Case of Michigan

Assessing Academic Progress in Michigan's Reading First Schools

To qualify for Reading First funding in 2002, Michigan districts or schools needed 40% or more students (or 50 or more students) scoring below proficient on the Michigan Educational Assessment Program (MEAP) for 2 of the last 3 years (Michigan Department of Education, 2002). Because significant underachievement

in reading is a criterion for eligibility for funding, Michigan's Reading First schools are very likely to be identified as needing improvement by the state regulations governing AYP at the time they begin their Reading First school plans. Even in the first year, before the infusion of funds and support from Reading First can make a difference, such schools are likely to be simultaneously making school-wide changes needed to implement Reading First and school-wide changes to fulfill their school improvement obligations.

This situation leads to potential disruption of school-wide instructional programs and initiatives. For example, Michigan schools that have not made AYP for three years must provide supplemental educational services for disadvantaged students (as is true nation-wide). Because of limited staff resources, the pressure is on schools to allow private agencies or businesses to provide tutoring services (Saulny & Beller, April 4, 2005). In Detroit, MI, for example, the district has compiled a list of approved providers, and schools can make arrangements with any of these to provide supplemental instruction. However, Saulny and Beller (2005) point out that the federal government offers no regulation or oversight of such tutoring arrangements, causing concern among school districts, elected officials, and industry executives. Further, they report that at present there are no scientific studies to indicate whether students in failing schools receive any academic benefits from such tutoring.

There are no criteria for tutors who work for companies that provide supplemental instruction to have credentials as academic tutors. This situation is at odds with the NCLB drive to improve teacher quality, which is largely focused on

requiring teachers to be certified to teach in the area in which they are teaching. While Darling-Hammond (2004) has found that the schools with the high levels of poverty and underachievement also have the least qualified teachers, we found that a large percent of the teachers in Michigan's Reading First schools had a master's degree (68%), and 62% had permanent or standard certification. It is possible that teachers are more able than tutors to provide appropriate instruction for struggling readers, if only they had the time to do so.

As part of the Reading First program in Michigan, teachers are taught to provide appropriate instruction for students who are not making timely progress in reading. For example, they are taught to use classroom assessments to monitor the progress of the students toward grade-level goals and to use this information to decide on the instructional needs of the students. In many schools, the students making the slowest progress are selected to receive an additional block of reading instruction from the teacher, outside the 90-miute literacy block required by the state's approved Reading First plan. The same students who are receiving extra instruction from their classroom teacher might also be receiving supplemental education from a tutor who is not familiar with their progress in the classroom or the current goals and methods being used by the teacher.

There is no requirement that the efforts required by Parts A and B to provide supplemental instruction for struggling students be coordinated. As a result, lessons from the outside providers are likely not to be connected to the classroom curriculum or the instructional regime followed by the classroom teacher. The

supplemental instruction provision of Part A could well result in a lack of coherent, high quality instruction for exactly those students it was designed to help.

How likely is it that Michigan schools entering Reading First are placed in improvement status based on Part A state assessment results? Of the 44 Michigan Reading First "round one" public schools (those with plans approved in spring/summer 2002), approximately 86% had been identified in a school improvement phase for 2001-2002 (that is, before they began their Reading First programs).[4] As Table 1 shows, of these public schools, approximately 80% were still in a school improvement phase the 2002-2003 school year (that is, after one year as a Reading First school). In the summer of 2003, additional schools became part of Michigan's Reading first initiative (termed "round 2" schools). As Table 2 shows, in the 2003-2004 school year, there were 110 public schools (including the continuing round one schools). Of these, approximately 59% had been identified in a school improvement phase for 2002-2003. After one or two years of a Reading First program (depending on the start-up year), approximately 49% of these schools remained identified in a school improvement phase for the 2003-2004 school year.

[4] There were actually 49 Reading First schools in the 2003-2003 school year, but some weren't public schools (i.e., charter schools), and some did not have students in the grades that take the MEAP (e.g., 4th grade)—thus, there were 44 Reading First schools with MEAP results in 2003-2003. Similarly, there were 115 Reading First schools in 2003-2004, but only 100 had MEAP results.

Table 1:MEAP Results for Round 1 Michigan Reading First Schools (N = 44)

	2001-2002 Phase		2002-2003 Phase	
	Count	*Percent*	*Count*	*Percent*
Made AYP	7	14%	9	20%
Did not make AYP (in a school improvement phase)	37	86%	35	80%

Table 2:MEAP Results Round 1 and 2 Michigan Reading First Schools (N = 110)

	2002-2003 Phase		2003-2004 Phase	
	Count	*Percent*	*Count*	*Percent*
Made AYP	46	41%	56	51%
Did not make AYP (in a school improvement phase)	64	59%	54	49%

While it appears that there is a noteworthy decrease in the number of Reading First schools in school improvement, the context from which we derived these figures is rather complex. First, it is impossible to determine the impact of AYP status on the reading achievement of students in the 86% of Reading First schools already in an AYP phase of school improvement in 2001-2002 (before they began the Reading First program). It is important to understand that Michigan began its AYP requirement in 1997-1998, before the passage and requirements of NCLB. By 2001-2002, Michigan schools not making AYP for four years in a row (*n* = 24 or over 50% of these round one schools) were already in level 3 of AYP, causing them to enact such sanctions as plans for restructuring and provision of tutors for students. Given the challenges faced by these schools, we are struck by

the accomplishment of two of those 44 schools that were able to make AYP that first year of their Reading First participation.

The decrease in Reading First schools in school improvement might have been affected by the differences in round 1 and 2 applicants. That is, many of the most seriously underachieving schools, as judged by poverty and performance indicators, received funding through their district in the round one competition; round 2 schools had somewhat higher percentages of students reading at grade level. Perhaps even more important, however, is the fact that Michigan English Language Arts (ELA) MEAP assessments were changed in 2002-2003 (both the content and the scoring rubrics), and no direct comparison can be made with previous test scores (Michigan Department of Education, 2004). As scores dramatically and immediately improved across the state with the new ELA MEAP assessment, it appears that the new AYP criteria were easier to meet. Therefore, while there seems to be a dramatic decline in the percent of Reading First schools in school improvement, it is too early to know if this trend will continue. It is also difficult to determine whether the gains are a valid reflection of achievement gains related to the Reading First program, or whether they are an artifact of the changed assessment. In either case, almost half of Reading First schools in 2004 faced the dual responsibility of implementing the Part B Reading First program while attending to Part A school improvement sanctions.

Michigan's NCLB Part A requirements are slated to become more stringent in the upcoming years. At present, all schools must have a 95% test participation rate for all grades tested by state requirements (4th, 7th and 11th for ELA and

mathematics, and 5th, 8th and 11th for science and social studies), and all designated subgroups must meet AYP benchmarks. Beginning the 2005-2006 school year, all grades 3 through 8 and 11th grade and all subgroups will need to meet the required AYP percentages for proficient reading and mathematics, increasing the possibility that schools will not make AYP in any given year. Additionally, the required AYP percentages for proficiency in reading and mathematics will increase each year as schools near the 2014 100% proficiency mark, further increasing the possibility that struggling schools in large, urban, and poor districts will not meet AYP (Center on Education Policy, 2004).

Reading First schools must worry about meeting the accountability requirements for improvements in reading established by both Part A and Part B. In Michigan, Reading First schools administer the Iowa Test of Basic Skills (ITBS) in the spring to determine students' progress in reading. The state Reading First administration uses two cut points to examine progress in reading on the ITBS, Reading Total measure. One is the 50th %ile; this is used to determine the percentage of students' meeting grade-level expectations. The other cut point is the 25th %ile; students below the 25th %ile are considered to be significantly underachieving in reading. Both are indices of progress in reading that are found in Part B (The No Child Left Behind Act 2001, P.L. 107-110). The goal is to show progress by increasing the percentage of students at or above the 50th %ile and decreasing the percentage of students below the 25th %ile. At the end of the first year of Michigan Reading First (spring 2003), only 19 percent of the third graders in the participating Reading First schools were reading at or above grade level on

the Reading Total subtest of ITBS, as Table 1 shows. The percentage reading at or above grade level increased to 29% in spring 2004 and 32% in spring 2005. The improvements in the first and second grades were greater. Each increase in the percentage of students reading at or above grade level for all three grades from 2003-2004 and 2004-2005 was statistically significant.

Table 3: Percent of students in Reading First Schools Reading At or Above Grade Level or Underachieving in Reading on ITBS Reading Total

	Percent at or above grade level (50th %ile)			Percent underachieving (below 25th %ile)		
	2003	2004	2005	2003	2004	2005
Grade 1	31	35	37	46	42	37
Grade 2	22	31	35	52	42	36
Grade 3	19	29	32	50	39	36

Note. This includes 49 schools for 2003, 115 schools for 2004, and 108 schools for 2005.

With regard to the goal of decreasing students reading significantly below grade level, Table 1 shows that there were decreases at all three grade levels from spring of 2003 to spring of 2004 and also from spring 2004 to spring 2005. It is also evident from this figure that in 2003 about half of the students in Reading First schools had reading achievement below the 25th %ile. While we expect Reading First schools to be characterized by high levels of underachievement in reading, the large percentage of students whose reading was below the 25th %ile is quite sobering. These figures help us appreciate the challenge schools face in trying to

raise the reading achievement of students so that they perform at or above grade level.

There is no reasonable way to compare performances on the MEAP and ITBS measures. The ITBS is a nationally normed test that reports students' scores, as related to the scores of other students nationally at the same grade level. The MEAP is a criterion-referenced assessment that was designed by Michigan to be aligned with state curricular benchmarks. It reports scores by four levels; those students scoring at one of the top two levels are considered to be proficient for AYP percentages. Nonetheless, it might be instructive to examine the number of Reading First schools that do not meet the standard for adequate achievement on each of these measures. We noted earlier that 49% of the 44 schools with MEAP results were in school improvement levels in 2003-2004. At the end of the same school year, 53% of these Reading First schools had fewer than 29% of their 3rd graders reading at grade level (29% representing the Reading First state average for this grade level). Thus, for both accountability systems, about half the Reading First schools in Michigan were performing below acceptable reading proficiency standards, given the goals and criteria set by Parts A and B (one or both) of NCLB Title 1.

Profiles of Michigan Reading First Schools

What happens when Reading First schools already in school improvement levels embark on their Reading First program? Do the resources of the Reading First program provide sufficient impetus for improvements in third-grade reading so that the school, overall, is able to meet Part A AYP criteria the following year?

In our search for preliminary answers, we selected seven schools that started their Reading First plans in the fall of 2002 and that were at MEAP phase 4 for that school year (2002-2003). Information about MEAP levels and ITBS results for these schools is shown in Table 4.

Table 4: ITBS and MEAP Results for Schools that Began Reading First at Level 4 of School Improvement

School	MEAP Phase		ITBS at or above 50th %ile (percent)		ITBS below 25th %ile (percent)	
	2003	2004	2003	2004	2003	2004
A	4	5	6	19	60	50
B	4	5	6	11	76	52
C	4	5	31	30	49	28
D	4	4	12	26	53	52
E	4	4	7	16	58	58
F	4	5	13	24	26	27
G	4	5	25	34	45	32

The table shows that 5 of the 7 schools moved to MEAP level 5 in the 2003-2004 school year. Two of the schools made sufficient progress so that they were not placed in phase 5, but they would need another year of meeting progress expectations in order to move out of school improvement status. Table 3 also shows the ITBS results for the two measures of progress (increasing the percentage of 3rd graders at grade level and decreasing the percentage of third graders significantly

below grade level). Most of the schools made noteworthy progress on the Reading First ITBS Reading Total between 2003 and 2004. School E made the least progress on both ITBS measures; school C made little to no progress in increasing the percentage of students at or above the 50th %ile, but it showed a large decrease in the percentage of students below the 25th %ile. Researchers warn us that changes over a two-year period are likely to be unreliable (e.g., Linn, 2002), but given the short time frame of Part B, progress from year to year is important to follow.

This preliminary examination of a sample of schools that started Reading First at Level 4 provides some support for our concern that the two accountability systems are likely to send different messages to Reading First schools. While we note again that the ITBS and the MEAP assessments are very different, we are also aware that in the coming years MEAP will be administered to 3rd graders as well as 4th graders, making it very likely that implicit or explicit comparisons of ITBS and MEAP results will be made. Many scenarios can be imagined. For example, what if there is a significant discrepancy for 3rd graders in a particular school on the ITBS and MEAP tests, such that only 30% of the 3rd graders perform at or above the 50th percentile on ITBS, while 60% are considered proficient on the MEAP? Or, conversely, what are the implications for a Reading First school that does not make AYP because its 3rd graders do not meet MEAP proficiency percentages, yet the school shows a large increase in the percent of students reading at or above the 50th percentile on the ITBS? An administrator would need to dole out praise for teachers' efforts to improve reading, based on ITBS scores, and disappointment in the same efforts because of a failure to meet required AYP requirements. The

likelihood that there will be mixed messages is great, and these are bound to affect the school climate and teachers' evaluation of their success in teaching their students to read. Teachers' effectiveness in the classroom is to a large extent reflected through their sense of self-efficacy; teachers, not just programs, contribute to students' achievement gains (AERA, Summer 2005).

Impact of Part A on Reading First School Culture and Capacity

The discussion thus far has documented our expectation that Reading First schools, which have large percentages of students underachieving in reading, are likely to be in a "school improvement" phase and may eventually require restructuring. The consequences of not making AYP are such that some aspect of the school organization is likely to be disrupted, and this disruption might affect the implementation of the school's Reading First program. Stability at the level of school organization and continuity of the staff and student body are desirable so that Reading First programs have optimal opportunity to become established and valued within the school community – characteristics associated with successful implementation (Elmore, 2002a).

The importance of school culture and teachers' satisfaction with their work conditions has been the topic of several recent research reports. The lead article in the American Educational Research Association publication *Research Points* (AERA, Summer 2005) suggests that more consideration be given to evidence indicating the effects of teachers on students' gains in academic achievement. In a similar vein, the Southeast Center for Teaching Quality (November 2004) points out that teachers' working condition are, in effect, students' learning conditions.

Such factors as time to prepare lessons and administrative support for their work affect the attitudes and performance of teachers. Conditions that place them under extraordinary stress or that detract from their ability to prepare lessons and run orderly classrooms are likely to deepen the problems faced by schools at various AYP school improvement levels.

To give an example, we describe the problems faced by a Michigan Reading First school that has been struggling to meet school improvement requirements for some time. This school is in Detroit; it reached level 5 of school improvement after two years in Reading First and was required to go through some form of restructuring. In the fall of 2004, the school changed its name and mission statement; the student body remained stable for the most part, but all teachers and administrators were new to the school. State evaluation of the implementation of Reading First showed that the school was facing formidable problems. Because of the turn-over of staff, some of the K-3 teachers had not received training in the comprehensive reading program adopted by the district or in the procedures for administering and interpreting the results of the classroom progress-monitoring system called Dynamic Indicators of Basic Early Literacy Skills (DIBELS, 2002-2005). Only some of the teachers had participated in the professional development required by the state (Language Essentials for Teachers of Reading and Spelling, published by Sopris West), as grade-level meetings were not held on a regular basis. Staff buy-in to Reading First was not school-wide. Missing were the motivation and initiative needed to unite the staff members in a coherent effort to implement the district's Reading First program. As of spring 2005, this new version

of the school had again not met AYP criteria and showed no improvement in reading achievement on the ITBS. In fact, a considerably larger percentage of students in grades 1-3 were underachieving in reading in the spring of 2005 than had been the case in 2003 or 2004.

With regard to the accountability measures in Part A, Elmore (2002b) stated unequivocally that, “in the history of federal education policy, the disconnect between policy and practice has never been so evident, nor so dangerous.” He contends that one of the biggest problems of Part A is that it is based on a “naïve” view of what it takes to improve student learning. Elmore distinguished between *internal accountability (or capacity)* and *external accountability*. Internal accountability takes into account that school personnel share an explicit set of norms and expectations about what an effective school looks like before they wrestle with external accountability outcomes (test scores):

> The work of turning a school around entails improving "capacity" (the knowledge and skills of teachers)—changing their command of content and how to teach it—and helping them to understand where their students are in their academic development … You can't improve a school's performance, or that of any teacher or student in it, without increasing the investment in teachers' knowledge, pedagogical skills, and understanding of students … the increased pressure of test-based accountability alone is likely to aggravate the existing inequalities between low-performing and high-performing schools and students (p. 3).

Elmore argued further that most high-performing schools reflect the “social capital” of their students (they are primarily schools with students of high socioeconomic status), rather than the internal capacity of the schools themselves. Many low-performing schools cannot rely on the social capital of students and families; instead, they must rely on their organizational capacity. With little or no

investment in capacity, low-performing schools get worse, relative to high-performing schools: “the result is an enormous distortion in the relationship between accountability and capacity—a distortion that is being amplified rather than dampened by [current] federal policy” (p. 3).

We have found it helpful to characterize Title 1 Parts A and B using Elmore’s (2002b) concepts of capacity (internal accountability) and external accountability. Although certainly not as clear-cut a phenomenon as that presented by Elmore, we see Part A as focused primarily on external accountability, while Reading First is focused on both external accountability and building capacity. The provisions of Reading First address issues of school culture and capacity: e.g., professional development for teachers and using classroom assessments to tailor instruction for struggling readers. Because Reading First schools are characterized by high poverty, they may therefore lack the “social capital” that Elmore associates with internal capacity, thereby increasing the challenge of meeting the external accountability standards.

This framework might help us better appreciate what it is like to work with Parts A and B simultaneously from administrators’ and teachers’ perspectives. First, teachers and administrators must spend more time preparing their students for large-scale assessments and administering these tests. That is, not only must they administer both MEAP and ITBS, but also they must thoroughly prepare students to take these different tests in order to offset further negative consequences of student underachievement. Teachers are also asked to take responsibility for the results. In some cases, AYP outcomes have been used to evaluate principals and teachers,

leading to replacements of staff members and principals whose students appear not to be making adequate progress. For Part B, too, test results might be used to evaluate the quality of instruction of the teachers or leadership skills of the principal. Such external forces clearly affect the academic and social structure of schools.

Part B focuses not just on accountability but also capacity-building. It requires that steps be taken to improve teachers' knowledge about reading and reading instruction, to upgrade the instructional materials and practices that they use, and to use classroom assessment to differentiate instruction for struggling readers. Additionally, principals are required to be part of Reading First professional development, and to regularly meet with other Reading First principals, the state facilitator, their school Reading First literacy coach, and the teachers. In these ways, Part B (Reading First) is addressing issues of school capacity and culture that have an effect on instruction, achievement, and therefore external accountability measures.

As we consider the scope of the changes in school organization and school culture that are part of Reading First, we need to remind ourselves that programs of curricular reform take time to implement and settle into place. The pressures of Part A notwithstanding, a short-sighted feature of Part B is the time span—it provides funding for eligible states, districts, and schools for only six years. While Reading First might be reauthorized, no one can assume that this will happen. Because studies have shown that five years minimum are typically needed to show the enduring effects of programmatic changes in reading instruction (Borman, Hewes,

et al., 2003; Chatterji, 2004; Desimone; 2004). The Reading First grant period barely allows sufficient time for schools and states to show stable effects of the changes brought about by the Reading First program features.

Aligning Assessment and Instruction

AYP and Reading First Impact on Teaching and Learning

Ultimately, simplistic or uni-dimensional accountability systems do not provide helpful information about school quality (Elmore, 2002b; Raudenbush, 2004). Even when, as recommended by experts, multiple years of test results and demographic factors are taken into account (Linn, 2003; Raudenbush, 2004), the results cannot tell us what good teaching looks like or what aspects of a particular school are functioning to support or undermine school improvement. AYP proficiency percentages alone do not provide information to make decisions about necessary improvements in the quality of instruction in a given school.

Part A has been criticized for unintended consequences that have a negative impact on the quality of education for all schools, whether they are in school improvement or not (Elmore, 2002b; Shepard, 2002). For example, schools are likely to narrow the instructional focus to emphasize the content covered in reading and math; content domains not covered by the high stakes assessments are likely to receive less emphasis or perhaps be eliminated altogether (Linn, 2003). Thus, the sanctions of Part A might inadvertently lead to a regression in meaningful instruction and learning, particularly for those students in struggling schools and districts.

Part B might also have negative unintended consequences associated with external accountability measures (e.g., narrowing the curriculum because of the amount of time spent on literacy). However, Part B provides substantial funds and an infrastructure for improving instruction in early literacy in Reading First schools. In fact, subgrants to the LEAs must be of sufficient size and scope to enable them to improve students' reading achievement (U.S. Department of Education, April 2002). Sufficient funding affords the LEA grantee substantial improvements in resources (materials and human services) needed to improve reading instruction. While funding alone certainly does not assure successful implementation of any school reform program, Reading First includes features that specifically address instructional and organizational school capacity. A commentary on NCLB accountability published in the Rand Review echoes this theme; the authors (McCombs & Carroll, 2005) make this point: "Demanding accountability without providing adequate resources can be an evasion of accountability by setting up public schools for failure."

As noted earlier, Reading First provides resources in a multi-faceted effort to improve instruction. Three particular resources stand out because they provide support for schools in the critical areas of instructional materials, leadership, and the alignment of instruction and assessment. While the legitimacy of the construct of SBRR is in question (Pearson, 2004), the requirement that materials and programs have research support has focused everyone's attention on the quality of materials used for reading instruction. In writing the law, legislators and their advisors decided to require grantees to use the results of federally supported

research initiatives to design instruction for underachieving readers. Thus, for example, Part B dictates that there must be instruction in the five "essential components" of reading instruction (phonemic awareness, phonics, fluency, vocabulary, and comprehension) in Reading First classrooms, on the basis of previous research that has shown that these areas play a role in the successful acquisition of reading skills (National Reading Panel, 2000).

Educational leadership is also given a central role in Part B, starting with the formation of a leadership team at the state level. In Michigan, administration of the grant has both state and local components. The state hires facilitators who provide support for implementation of Reading First programs in five or six schools. Each school hires a literacy coach who provides support for the teachers. This takes the form of grade-level meetings to discuss methods and problems of reading instruction, meetings with teachers to provide advice or to assist in procuring materials, and assistance in carrying out classroom-based assessments of reading. Done well, this system should provide an integrated commitment on the part of state, district, and school educators.

Particularly relevant for us, given the purposes of this paper, Part B provides a framework to align assessment and instruction within the classroom context. In Michigan, Reading First schools are required to use classroom-based measures of reading to assess students' progress in reading three times during the year (Dynamic Indicators of Basic Early Literacy Skills). Classroom and student profiles provide information about the progress of individual students, and teachers are taught to use this information to adjust methods and materials of reading instruction. Classroom-

based measures offer information that is useful for teachers during the school year, unlike year-end standardized test results, which are deemed by teachers not useful for purposes of instruction (Ingram, Louis, & Schroeder, 2004; Wixson & Carlisle, 2005).

Because of the support for research-supported reading instruction that is built into Part B, Reading First is promising as a federal policy that might help failing schools to improve their students' reading achievement. Nonetheless, it is important to point out that the provisions of Part B might not be sufficient to offset the deep-seated difficulties faced by high-poverty schools with limited resources attempting to improve their students' reading achievement. These difficulties might be compounded by Part A sanctions. While states other than Michigan will soon begin reporting on the progress in reading made by their Reading First schools, it will be a number of years before this federal program can be adequately evaluated overall and before the impact of Part A on Part B can be assessed.

Looking Forward: Broadening School Accountability Perspectives in the Evaluation of Reading First Schools.

Some experts have offered suggestions of ways to address proposals the shortcomings of accountability policies in NCLB Title 1, Parts A and Parts B. Leading educational researchers (Chatterji, 2004; Darling-Hammond, 2004; Elmore 2002b; Raudenbush, 2004) sound a clarion call for more complex, nuanced models of accountability that evaluate school practice and organization—issues of capacity—versus assessment outcomes alone. Raudenbush (2004) has this to say:

> Identifying evidence-based best practices is far more difficult than holding schools accountable for outcomes alone. But this hard work appears essential

> if schools are to be held accountable in ways that are scientifically defensible, fair, and effective. A mix of evidence based on outcomes and assessments of practice appears essential if accountability is to achieve its potential to improve schools (p. 37).

Raudenbush (2004) acknowledges that employing an accountability system that evaluates "evidence-based best practices" is a difficult endeavor. There are the obvious issues of resources, in terms of both the financial support and the infrastructure necessary to implement (and evaluate) instructional and organizational school reforms. Perhaps less obvious and more complex are the philosophical and ideological problems that would have to be addressed to instantiate such an accountability system (cf. Chatterji, 2004; Elmore, 2002ab; Pearson, 2004).

Still, Raudenbush gives us suggestions about what a useful and defensible accountability system might look like. He mentions the combination of evidence-based outcomes and information about practices in the classroom. In addition, information on the management or organizational structure of the school should be taken into account. Granted, it is difficult to gather trustworthy information about the use of evidence-based "best practices" at the school and classroom level, but it is information of this kind that will provide a solid basis for examining schools' accountability. How else will we be able to determine the extent to which schools and teachers are doing all they can to improve the reading skills of their students?

Somewhat similarly, in discussing best practice for carrying out evaluations, Chatterji (2004) recommends what she calls the *extended-term mixed-method* (ETMM) design. ETMM designs follow the life-span of programs and policy initiatives within particular environments, employing descriptive research methods

in the early stages of program adoption and implementation, followed by experimental designs at a subsequent stage. Chatterji cautions against using only short-term quantitative or experimental evaluation models that focus on assessment outcomes alone:

> What often results [from these models] is an atheoretical, poorly conceptualized "black box" evaluation where little is unveiled as to the reasons and conditions under which a program worked … it is no longer sufficient for the academic and professional community to continue discussions amongst themselves on methods for obtaining best evidence. It is time to step up, inform, and as necessary change federal policy, before it is too late (pp. 4-12).

The complexity of analyses that truly inform us about "reasons and conditions" is one reason this model is unlikely to receive a warm reception among policy makers. As with the system recommended by Raudenbush, an accountability system of this kind would be difficult to implement and monitor at the state and LEA levels. Still, it seems necessary at this point to take heed of Chatterji's warning that adequate documentation of the differences in implementation of programs at the district, school, and classroom level must be sought. Without such documentation, the effects of specific Reading First instructional or leadership factors, for example, would be difficult to interpret, let alone replicate.

It seems to us that accountability policies and programs are stronger if they provide resources for building more effective schools, as is the case with Part B Reading First, as opposed to Part A, which focuses on achievement results alone. Evaluation of the effectiveness of accountability measures in Reading First and the resources and provisions put in place to help schools achieve these measures are underway. Ideally, these evaluations will entail both qualitative and quantitative

assessment. Although certainly not perfect and not without its detractors (Manzo, 2005), Reading First may be a step in the right direction of eschewing scientifically unsound approaches to accountability and school improvement, especially for our most neglected students.

References

American Educational Research Association (AERA). (Summer 2004) Teachers matter: Evidence from value-added assessments. *Research Points, 2(2).*

Borman, G. D., G. M. Hewes, et al. (2003). Comprehensive school reform and student achievement: A meta-analysis. *Review of Educational Research,* 73(2): 125-230.

Center on Education Policy (October, 2004). *Rule changes could help more schools meet test score targets for the No Child Left Behind Act*. Downloaded from http://www.cep-dc.org/nclb/

Chatterji, M. (2004). Evidence on "What Works": An argument for extended mixed-method (ETMM) evaluation designs. *Educational Researcher, 33 (9),* 3-13.

Darling-Hammond, L. (2004). Inequality and the right to learn: Access to qualified teachers in California's public schools. *Teachers College Record, 106*, 1936-1966.

Desimone, L. (2002). How can comprehensive school reform models be successfully implemented? *Review of Educational Research*, 72(3): 433-479.

Elmore, R.F. (2002a). Local school districts and instructional improvement. In W.D.Hawley & D.L.Rollie (Eds.), *Keys to effective schools: Educational reform as continuous improvement* (pp.111-121). Thousand Oaks, CA: Corwin Press.

Elmore, R., F. (2002b). Testing trap. *Harvard Magazine*. September-October. Down-loaded from www.harvard-magazine.com/on-line/0902140.html

Erpenbach, W.J., Forte-Fast, E., & Potts, A. (July 2003). *Statewide educational accountability under NCLB*. Washington, DC: Council of Chief State School Officers.

Ingram, D., Louis, K.S., & Schroeder, R.G. (2004). Accountability policies and teacher decision-making: Barriers to the use of data to improve practice. *Teachers College Record, 106*, 1258-1287.

Linn, R.L. (2002). Stability of school-building accountability scores and gains. *Educational Evaluation and Policy Analysis, 24*, 29-36.

Linn, R.L. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher, 32*, 3-13.

Linn, R.L., Baker, E.L., & Betebenner, D.W. (2002). Accountability systems: Implications of requirements of the NCLB Act of 2001. *Educational Researcher, 31*, 3-16.

Manzo, K.K. (2005). Complaint filed against reading initiative. *Education Week.* June 22.

McCombs, J.S., & Carroll, S.J. (Spring 2005). Ultimate test: Who is accountable for education if everybody fails? *Rand Review*.

Michigan Department of Education. (2002). *Making Reading First in Michigan*. Application submitted to U.S. Department of Education. (Downloaded from www.mireadingfirst.org)

Michigan Department of Education. (2004). *MEAP*. (retrieved from www.michigan.gov/meap)

Miskel, C. & Song, M. (Summer 2004). Passing Reading First. *Educational Evaluation and Policy Analysis, 26. 2, 89-109.*

National Reading Panel Report. (2000). *Teaching children to read: An evidence-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.

Olson, L. (2002). 'Inadequate' yearly gains are predicted. *Education Week.* April 3.

Olson, L. (2005) Requests win more leeway under NCLB. *Education Week.* July 13.

Pearson, P. D. (January and March 2004). The reading wars. *Educational Policy*. 18, 1, 216-252.

Raudenbush, S.W. (2004). *Schooling, statistics, and poverty: Can we measure school improvement?* Princeton, NJ: Educational Testing Service.

Saulny, S., & Beller, P. (April 4, 2005). A lucrative brand of tutoring grows unchecked. *New York Times*, East Coast edition.

Shepard, L. (2002). The hazards of high-stakes testing. *Issues in Science and Technology 19(2).*

Southeast Center for Teaching Quality. (November 2004). Teachers' working

conditions are students' learning conditions. *Best Practices and Policies, 3(9*).

U.S. Department of Education, Office of Elementary and Secondary Education. (April 2002). *Guidance for the Reading First Program*. Downloaded from www.ed.gov

U.S. Department of Education. (2002). *No Child Left Behind: Executive summary*. Retrieved from www.ed.gov/nclb/overview/intro/execsumm.html

Wixson, K.K., & Carlisle, J.F. (2005). The influence of large-scale assessment of reading comprehension on classroom practice: A commentary. In S.G. Paris & S.A. Stahl (Eds.), *Children's reading comprehension and assessment*. Mahwah, NJ: Lawrence Erlbaum.