Behind the Numbers

Interviews in 22 States about Achievement Data and No Child Left Behind Act Policies

Phase II Report

Center on Education Policy Study of Student Achievement Since the Enactment of No Child Left Behind

Prepared by the
Human Resources Research Organization (HumRRO)
Alexandria, VA

For the
Center on Education Policy
Washington, DC

October 2007
# Table of Contents

Introduction .................................................................................................................. 1
Summary of Key Findings from Phase II ................................................................. 2
Confirmation of Phase I Data .................................................................................. 3
Post-NCLB Changes to State Assessment Systems ............................................... 4
How States Monitor Instructional Impact ............................................................... 10
How the Federal Government Could Help States Meet the Challenges of NCLB ..........12
Impact of Major Provisions of NCLB on Achievement .......................................... 15
Methodology for Phase II ....................................................................................... 22

Appendix A: Sample of States .............................................................................. 25
Appendix B: State Comments on Phase I Data ......................................................... 26
Appendix C: Sample Interview Protocol ................................................................. 28
Credits and Acknowledgments ............................................................................... 31
About the Center on Education Policy .................................................................... 31
Introduction

This report describes findings from a series of interviews with education leaders in 22 states, conducted and analyzed by researchers from the Human Resources Research Organization (HumRRO). The 22 participating states included North Carolina and Michigan (the two pilot states), as well as Arizona, Delaware, Florida, Idaho, Illinois, Kentucky, Louisiana, Massachusetts, Minnesota, Missouri, New Mexico, Pennsylvania, Oregon, Rhode Island, South Carolina, Virginia, Washington, West Virginia, Wisconsin, and Wyoming. These interviews, which were commissioned by the Center on Education Policy (CEP), constitute phase II of a comprehensive, CEP-sponsored study of student achievement since enactment of the No Child Left Behind Act (NCLB) in 2002.

Phase I of the study, carried out by CEP with support from HumRRO, looked at trends in state test scores from all 50 states before (where possible) and after NCLB took effect. Phase I included analyses of overall student achievement and achievement gaps between student subgroups for 1999 through 2006 or whichever of these years had comparable test data. Phase I findings were published in a June 2007 report, Answering the Question That Matters Most: Has Student Achievement Increased Since No Child Left Behind? That report, along with detailed profiles of achievement data for each state, is available on CEP’s Web site at www.cep-dc.org.

Phase II, the focus of this report, used interviews to gather qualitative information that could provide a context for better understanding the achievement trends identified during phase I. The 22 states selected for phase II included those with the longest periods of comparable test data before and after NCLB—that is, data that had not been rendered incompatible due to changes in test content or instruments, in cut scores for proficient performance, or in other significant aspects of the state’s testing program. These states were selected because of the possibility that states with longer trend lines would have more evidence on which to base their perceptions about the impact of NCLB. Information about the methods used to select states, conduct interviews, and analyze interview data can be found in the Methodology section near the end of this report. Dates for the interviews can be found in appendix A.

For phase II, HumRRO researchers conducted mostly on-site, in-person interviews from January through May 2007. Interviewees typically included the state superintendent and/or deputy superintendent and the state assessment director, and sometimes included other state officials. The interviews had two main purposes: 1) to collect additional evidence from state experts about the test score trends identified during phase I of the study; and 2) to gather qualitative information from state officials about the impact of various NCLB provisions on achievement and other outcomes and to listen to states’ suggestions for changing the law. Throughout the interviews, state officials were asked to provide evidence, in addition to the data collected for phase I, to support their claims about the impact of NCLB.

The information in this report is taken directly from interviews with state officials. Appendix C lists the questions state officials were asked. The interview questions were largely open-ended—in other words, states were not given a list of responses from which to choose. In cases where multiple states gave similar responses, the report includes a count of states reporting a particular response. Readers should not assume, however, that other states did not take the action implied
by the response; rather, states may be omitted if their interviewees did not raise a particular point during the interview.

**Summary of Key Findings from Phase II**

Several key findings emerged from the phase II interviews:

- **General confirmation of phase I data.** In general, states interviewed confirmed the accuracy of the data and achievement trends developed for phase I of the CEP achievement study. Several states provided more information, explanations, or clarifications of their phase I data, which CEP and HumRRO took into account before publishing the phase I report. Three states offered to provide missing data that had not been made available by the deadline for verifying phase I data, but these requests could not be accommodated within the phase I publication schedule.

- **Limited research-based evidence that corroborates or explains achievement trends or monitors instructional impact.** Most states had not done their own evidence-based studies of long-term achievement trends or factors that might explain these trends. According to state officials, this is largely because of a lack of staff capacity and the rising demands placed on state staff. On a related note, seven states reported that they monitor or have done studies of instructional impact, through such means as examining the impact of specific instructional practices on achievement or reviewing the alignment of instruction to standards. The remainder of the 22 states had less formal procedures to encourage certain instructional practices or did not monitor instructional impact.

- **Changes in state testing systems.** Many states have made changes in their testing systems since NCLB took effect in 2002, often in ways that have caused “breaks” in the compatibility of test data and made it invalid to track trends over several years. The most common changes were done to comply with NCLB requirements, such as testing more grades, disaggregating data for the mandated subgroups, or changing policies for assessing English language learners. States have also made other types of changes. Between one-third and two-thirds of the 22 states reported making one or more of the following changes: revising or adding to the content standards on which tests are based; adding new tests or modifying existing tests; changing the cut scores defining proficient performance on their tests; and changing test contractors.

- **Future breaks in test data and more publicly available data.** Many states are actively pursuing additional changes in their assessment systems that will continue to affect the comparability of test data in the future. Some of the most common changes underway include revisions of content standards, changes in high school exams as a means of strengthening graduation requirements, or adoptions of new or additional assessments. In a positive move, seven states indicated that they intend to increase the amount of test data available to the public.
• **Need for federal assistance on data and assessments.** Close to half of the states interviewed emphasized the need for federal technical assistance (and often funding) to help states and school districts develop the capacity to collect and analyze the complex assessment and other data required by NCLB. Some states called on the federal government to provide better information, guidance, or research about such topics as best practices for improving achievement or closing achievement gaps. Many states made other suggestions for broader types of federal assistance, such as better federal guidance about NCLB in general or more realistic deadlines for carrying out the law’s requirements.

• **Anecdotal evidence about positive impact of disaggregating data, little positive effects from other provisions.** When asked about the impact of the major provisions of NCLB on student achievement and achievement gaps, state officials usually cited anecdotal evidence rather than research studies to support their conclusions. Eight states concurred that disaggregating data for subgroups had helped to improve achievement or brought related benefits, although some states referred to negative consequences of disaggregation. However, most states interviewed did not view the current NCLB accountability requirements as a particularly effective means to identify low-performing schools or raise student achievement. States reported few or no impacts of other major NCLB provisions on achievement. In the case of supplemental educational services, several states spoke specifically about the difficulty of evaluating the effectiveness of these services in improving achievement for participating students.

**Confirmation of Phase I Data**

A purpose of the phase II interviews was to confirm, challenge, or elaborate on the trends in overall achievement test scores and achievement gaps identified by HumRRO and CEP during phase I of the comprehensive study. As explained in more detail in the Methodology section, interviewees were given a summary of test data and trends developed for their state during phase I of the study; with only one exception, these data had already been verified by state assessment personnel.

HumRRO researchers opened the interview by describing the phase I data that had been collected by HumRRO and CEP and verified by the state. Then the researchers reviewed the phase I findings and conclusions drawn by CEP and HumRRO from the data provided by the state. The researchers asked state officials whether they agreed with the conclusions regarding overall achievement and achievement gaps for their state, and whether the phase I analyses had missed any key components that might lead to substantially different conclusions about overall achievement or achievement gaps.1

Generally, states agreed with the phase I findings. Some states provided clarifications of data or additional information about their assessment policies. Other states expressed concerns about

---

1 One pilot state was not asked this set of questions, which was added to later interviews.
such issues as the impact of changing assessment policies for students with disabilities or English language learners (ELLs), the small size of certain subgroups, or the different picture of achievement gaps shown by assessments other than those used for NCLB. Before publishing the phase I report, CEP and HumRRO took into account these clarifications and concerns and addressed them in the report or accompanying state data profiles. Appendix B summarizes the main issues about the phase I data raised by states and describes how they were addressed.

As explained in appendix B, three states offered during the interviews to provide missing data sets related to mean scale scores and standard deviations that had not been made available within the extended deadline for verifying phase I data. CEP could not accommodate these requests due to the publication schedule for the phase I report, which was developed to ensure that policymakers would have the phase I results in ample time to inform NCLB reauthorization.

**Post- NCLB Changes to State Assessment Systems**

Since states began implementing NCLB in 2002, many have made changes to their assessment systems—often in response to the requirements of NCLB itself. Some of these changes have created breaks in the comparability of test data and have made it inadvisable to compare test data before and after the change. Examples include a switch to a different test instrument, a change in the cut score that defines the “proficient” level of performance on the test, a change in test content or in the standards to which the test is aligned, or alterations in other major aspects of a state’s testing system. The phase I study took great care not to compare test data that had been rendered incompatible by these types of changes.

The phase II interviews yielded additional information about changes in state testing systems. One interview question asked specifically whether states had changed their assessment systems after NCLB was enacted in 2002. In addition, state officials sometimes volunteered information about changes in testing in response to other interview questions. Interviewees indicated that state assessments have changed for a variety of reasons—sometimes to meet NCLB requirements and sometimes for other reasons.

**CHANGES REQUIRED BY NCLB**

One of the most common changes was the addition of more grades to the testing program to comply with the NCLB requirement for testing students in grades 3 through 8 and once during high school. Previously, some states tested at fewer grades or tested by grade span—for example, by testing students once in the primary grades, once in the upper elementary grades, once in middle school, and once in high school. Twelve states reported during our interview that they had added grades to their testing systems. Idaho, for example, did not begin grade-level assessments until 2003, and Massachusetts had to add tests at some grade levels. In addition, Michigan and Rhode Island noted that they had developed new grade-level expectations to clarify what would be assessed at each grade level. Rhode Island also developed achievement-level descriptions for each grade level.
Disaggregating test data by subgroups was another change required by NCLB; specifically, the law required states to break out test results for major racial/ethnic groups, for low-income students, and for students with disabilities and English language learners. Five states reported conducting at least some data disaggregation before NCLB, but many states did not. In the interviews, eight states reported making changes related to subgroup data. In addition, several other states noted that they had begun paying more attention to the performance of subgroups as a result of receiving yearly data on subgroup performance. The following information illustrates the types of changes made:

- **Oregon** changed its reporting to include disaggregated data for more student subgroups; from 1997 through 2001 only students who had been tested under standard conditions with assessments geared to their own grade level or above had been included. Beginning in 2002, data covered all students tested, including English language learners who were recent arrivals to the United States. However, 2005 and 2006 data did not include ELLs who were recent arrivals, in accordance with revised policy guidance from the U.S. Department of Education (ED).

- **West Virginia** began disaggregating test data by federally defined subgroups as well as by district and school.

- **Missouri** has become much stricter about the types of tests and accommodations permitted for students with disabilities to ensure the state is in compliance with NCLB.

- In **North Carolina**, English language learners had previously been allowed two years of classroom instruction before being tested, but ED policy guidance allows ELLs to be exempted from testing in reading only for the first year they attend a school in the U.S. North Carolina revised its policy to be consistent with this guidance.

**OTHER CHANGES**

Other changes were also discussed. Fifteen states reported adopting new content standards or revising their current standards since 2002. In some cases, these revisions were in response to the NCLB requirement to test grades 3 through 8. States that had tested only a few grades within that span had to develop standards for the additional grades. In other cases, states revised standards for different reasons, such as the decision to implement a new assessment. An assessment official in one such state explained the state’s changes in content standards in this way:

*We have changed the content we’re testing. One of the biggest changes in the [state] assessment system is that under the guidance of [a state assessment commission] we reconfigured how we think about content areas. So in reading, for instance, instead of having the 80 or 117 benchmarks, we have eight skills that are assessed . . . Those eight skills really cover the entire standard but they do it in a different way, so we’ve dramatically reduced the number of assessment targets but [have] done that in a way that we really think that we’re assessing the same breadth and depth of the content.*
In a change that is not directly attributable to NCLB but may affect adequate yearly progress (AYP) determinations under the law, 12 states reported changing the cut scores that define the proficient level and other performance levels on their tests or modifying performance standards. While it was not always clear whether these states had lowered or raised their cut scores, some states did provide information about the direction of the shift in cut scores:

- **Delaware**, for example, set cut scores in 2006 that were slightly higher in reading and slightly lower in math than in 1999.

- **Oregon** lowered the performance standards for high school but raised them for elementary school to address differences between grades in the percentage of students meeting standards. According to state officials, the magnitude of the changes roughly equaled about a third of a standard deviation at the elementary and high school levels. As part of this process, a representative group of Oregonians determined that the new high school cut score was the most appropriate for 10th graders.

- **North Carolina** set higher performance standards for its math end-of-grade exams in 2006. But to adjust for the impact of the changing performance standard, the state also revised its NCLB annual measurable objectives (the percentage of students that must score at the proficient level for a school or district to make AYP), resulting in a steeper path to 100% proficiency in 2014. Reading standards in grades 3-8 will be reviewed in 2008.

- **Illinois** reset its cut scores for 8th grade math in 2006 to bring them more in line with the cut scores for other grades. Previously, the 8th grade math cut scores had been set much higher, according to state officials—most likely because cut scores for different grades were set by different committees.

- Two states, **Louisiana** and **Massachusetts**, reported making changes in performance descriptors that they said had no effect on scoring or performance standards. Louisiana had changed its performance label descriptor from “proficient” to “mastery” in 2003 to eliminate confusion with terminology associated with AYP. Massachusetts in 2002 went from having three cut scores to five, to add greater precision to the ends of the state’s scoring scale.

Ten states spoke about adding tests or modifying existing tests. Examples include the following:

- **North Carolina** added science tests at grades 5 and 8 and also changed some alternative assessments in reading and mathematics in grades 3-8.

- **South Carolina** added an alternate assessment for students with disabilities, as well as a high school exit exam.

- **Florida** added tests in science and writing.
Eight states reported changing test contractors since NCLB took effect. Typically, this was done because the contract with the current contractor had expired and the state was required by law to reopen competitive bids, or because the test requirements had changed—for example, by adding tests in more grades—and the state required competitive bids in this situation. Virginia changed testing contractors; the new contractor assumed complete responsibility for the testing program beginning with the fall 2006 administration. Washington officials said that the contract with its current test contractor was due to end soon, and the open bid process would be used to select a new test contractor. For some states, changes in contractors contributed to problems in maintaining a system that could be compared over time. New Mexico, for example, had changed test contractors and tests multiple times since 2002, most recently in 2004. As a result, said one state official, “there was no continuity in the past.”

Three states reported changing their time windows for administering tests. Michigan, for instance, changed its test window to obtain test results more quickly, a move intended to facilitate planning for school improvement.

Some states have switched, or are planning to switch, to a vertical scale for scoring tests—a type of scoring that puts scores from different grades on a common scale, enabling students’ scores to be compared across grades. The following examples emerged from state interviews:

- After NCLB was enacted, Arizona developed a vertical scale for grades 3-8. This was done to address gaps in the assessment and allow the state to track student growth.

- Illinois developed its vertical scale in 2005, a move that state officials said made it easier for the public to understand test results.

- New Mexico established its vertical scale two years ago in hopes of moving in the future toward a growth or value-added model for gauging student performance.

- Officials in Delaware reported that the state test uses a “vertically articulated” scale, in which cut scores for 8th grade students are set higher than those for 5th grade students. (A true vertical scale requires that students in a particular grade be given test items from adjacent grades, at the least.) Delaware also assigns identifier numbers to each student, enabling the state to track students longitudinally and determine whether subgroups are making more or less than a year’s worth of growth, officials said.

Other states are implementing systems of student identifiers to allow them to track student growth. Idaho officials said that the state board would like to expand growth measures to include fall and spring testing in the same school year. Next year’s assessment will have unique student ID numbers to help link performance to a particular student, and a growth model may be put in place in the next few years. Washington interviewees said that vertical scales will be implemented in about two years, but that it would likely be another few years before a trend can be analyzed. The state plans to conduct studies as the vertical scale is implemented. Minnesota uses an index system for its AYP calculations in which a proficient student receives 1 point, a partially proficient student 0.5 point, and a student who is not proficient 0 points.
Finally, North Carolina and Florida reported that they have already received permission from the federal government to use a growth model to measure student achievement for making AYP determinations.

**FUTURE CHANGES AFFECTING STATE ASSESSMENTS AND TEST DATA**

States were asked what, if any, changes they planned to make in their state assessments. Respondents discussed a variety of changes, some of which were being actively pursued and some of which were still in the early discussion stages. The analysis that follows was limited to changes that were being actively pursued. Taken together, state responses suggest that the year-to-year comparability of assessment data within states is likely to be affected by significant changes in state testing systems for years to come.

Seven states were in the process of revising at least some of their academic content standards—a step that usually necessitates a redesign of state assessments to align them with the new standards. Examples of state responses include the following:

- **Florida**’s language arts standards had already been revised and adopted by the state board, and revisions to the mathematics standards were expected to be adopted by summer 2007. The math revisions are expected to be major; state officials explained that the math standards are being revised around the focal points developed by the National Council of Teachers of Mathematics, which restrict the content taught at each grade while demanding greater depth and rigor within the more narrowly defined content.

- **Minnesota** is reviewing and revising its mathematics standards as part of its regular review cycle for each subject area.

- **South Carolina** expected to adopt revisions to its academic content standards in English/language arts and mathematics standards. This process will lead to new standard setting and revamping of state tests.

- **Washington** was reexamining math standards and also planned to review science standards.

Officials in other states described a variety of changes planned for their assessment systems. Five states reported adding or changing their high school assessments as part of an effort to strengthen graduation requirements. The following state responses illustrate this trend:

- **North Carolina** is implementing new high school exit standards that require students in the graduating class of 2010 to pass end-of-grade exams in five subjects and successfully complete a graduation project.

- Legislation pending in Arizona calls for end-of-course testing in a few high school subjects.
In Massachusetts, students in the class of 2010 must pass a statewide science test, in addition to exit exams already required in other subjects, in order to receive a diploma.

At the time of the interviews, three states were introducing or planning to introduce new assessments. Minnesota will begin implementing a new mathematics assessment for all students in 2011; assessments in other subjects will be revised later, after content standards are revised. Kentucky will add the ACT college admissions test to its statewide assessment, along with Kentucky Core Content Tests in every grade level, to replace the augmented norm-referenced test used previously. As part of its testing system redesign, the state will reduce the number of open-response questions in favor of multiple-choice questions. Delaware had planned to implement a new assessment in 2009, but these plans have changed due to budget constraints.

Virginia plans to develop assessments that are aligned to modified achievement standards for students with disabilities, consistent with final regulations on this issue recently released by the U.S. Department of Education. Virginia officials also reported working with ED on assessment issues for English language learners, including criteria for designating students as limited English proficient.

Most of the other states interviewed also anticipated making changes to their assessment systems, but these plans were still in the early discussion stages, so they have not been included.

**AVAILABILITY OF TEST DATA**

State officials were also asked several questions about their plans to release test data in future years; these questions were an outgrowth of the difficulty that CEP and HumRRO staff experienced during phase I of the study in obtaining certain types of data from some states that one would expect all testing systems to have. One question dealt with the future accessibility of data via the Internet and another with anticipated changes in public availability of data.

Seven states reported that they intended to increase the amount of data available to the public. Minnesota officials, for example, said that standard deviation data had been posted to the state’s Web site shortly after CEP and HumRRO had requested it for phase I; these data had not been posted previously due to a capacity issue. Rhode Island officials said that the state plans to make available sufficient data to answer questions about student achievement trends in a timely manner, through an online data warehouse that will be used along with the assessment data currently available on its Web site. Washington reported that by 2012 it will have a much friendlier state Web site that will include growth models and vertical scaling. Achievement trend lines may be disrupted, however, due to the possibility of new content standards and the move from a comprehensive test to end-of-course exams.

Seven states indicated that they had no plans to make more test data available to the public. Missouri officials, for example, said they could not guarantee that all data would be available, but that they were happy to provide it when asked. South Carolina noted that all of its technical reports are already posted online, although not in the same place as the test data. New Mexico officials noted that much data is already available through its Student Teacher Accountability Reporting System (STARS). This system, which uses a data warehouse model, has reported and
calculated AYP and achievement information from the 2007 assessment and allows districts to analyze their data without relying on Excel.

Five states reported that they already were posting necessary data. For example, Massachusetts and Delaware emphasized that all their test data, except for student-level data, were currently online.

Additional themes emerged from the interviews relating to other aspects of data collection and posting. First, four states pointed out that their ability to make data publicly available is constrained by limited time, funding, and staff. Interviewees in Arizona, to cite one such state, specifically referred to a lack of time to prepare and showcase assessment results as a major constraint, while Oregon officials mentioned limited resources. Second, three states commented on the difficulty of fulfilling myriad requests for data. For example, West Virginia officials said it was impossible to make data available in all the various formats that potential users requested. Third, interviewees in three states commented on the public’s lack of understanding about data and their uses.

Finally, Florida officials pointed out that the state’s longitudinal, integrated “education data warehouse” has piqued the interests of researchers inside and outside the state who want access to longitudinal data to support research proposals. Florida has refined an access process that permits external researchers to apply for data, be reviewed and coached, and be approved for access to defined data sets. The “coaching” part includes guidance about data definitions and data element connections, as well as the development of a “quid pro quo” arrangement whereby the state education department requests related topical research that will assist in evaluating programs. In this way, said state officials, several scientifically-based research studies have informed policy and evaluated program approaches.

**How States Monitor Instructional Impact**

States were asked if they monitor instructional impact—in other words, if they monitor the impact of different instructional practices on student achievement or if they ensure that their instruction aligns to state standards—and if so, whether they have seen substantial instructional changes since NCLB was enacted 2002. States responded in a variety of ways, ranging from states that conducted studies of the impact of different instructional methods to states that did not monitor instructional impact at all.

The strongest examples of instructional monitoring were in the eight states that have conducted or sponsored studies of instructional impact or have put in place a monitoring system that documents district or school efforts to improve alignment or instructional practices. In two of these states, interviewees described studies that had been conducted several years ago and reported little activity since. Officials in the remaining five states mentioned a variety of examples:

- **Idaho** has implemented a pilot study of 24 middle schools throughout the state that have not met their AYP goals under NCLB. Currently in its second year, the study is
examining instructional alignment with state standards and types of instructional practices in classrooms.

- **Missouri** described a long-term contract with the Assessment Resource Center at the University of Missouri-Columbia to conduct a consequential validity study that will compare the intended consequences of the assessment system on instruction with the actual effects in the classroom.

- **Oregon** has a system for monitoring school improvement and investigating strategies being used by schools that have made impressive gains in achievement.

- **Virginia** described an internal analysis of middle school mathematics achievement and teachers who hold mathematics endorsements. As expected, achievement was higher for teachers who have a mathematics concentration.

- **Kentucky** collects data from students on quality of instruction, which is part of the Kentucky performance report.

While not explicitly studying instructional impact, four other states described steps they took to encourage instructional practices that are tied to state standards and assessments. **Massachusetts** provides content frameworks and some lesson plans for teachers. In addition, Massachusetts seeks to tie instructional practice to state standards and assessments by giving schools and districts access to every student's response to every question on the state assessment, along with a map of which questions evaluate which state standards; this is done through data analysis tools such as TestWiz and the state data warehouse. Massachusetts also releases all test items from every assessment on its Web site.

**Louisiana** encourages, but does not require, districts to use state curricula. **Arizona** and **Florida** require districts and schools to document that they are using state standards.

**Arizona** and **Florida** were also among the four states that indicated they take additional steps to monitor schools that do not make AYP; **Kentucky** and **Idaho** were the other two. **Arizona** sends a “solution team” to the school to look at the curriculum and ensure it is aligned with the state’s academic standards. **Florida** monitors schools low-performing schools, as well as schools participating in the federal Reading First program, through its Office of School Improvement.

Four states described district or local monitoring of instructional impact instead of, or in addition to, state monitoring. For example, **Kentucky** provides electronic, hand-held classroom observation tools, but the state does not collect data from these observations, which are used at the local level. **Washington** reported that some districts have analyzed math by course, state scores, and class passing rates.

**Delaware** and **Florida** conduct audits or instructional reviews, either voluntary or required, that include an examination of instructional validity.
Washington officials said that Reading First was the only program for which instructional impact could be monitored. The state lacks resources to monitor math courses, and math has not been subject to a list of prescribed instructional approaches as reading has with Reading First. The state suggested that end-of-course exams might be one way to hold teachers and districts accountable.

**How the Federal Government Could Help States Meet the Challenges of NCLB**

States were asked what kinds of technical support, guidelines, auditing, or other resources the federal government could provide that would help address the challenges of NCLB. The interviewers did not attempt to limit states’ responses by asking, for example, for the three most important types of support states needed. Therefore, many states suggested several types of support.

**FEDERAL ASSISTANCE RELATED TO DATA AND ASSESSMENTS**

A key theme arising from state responses was the need for federal assistance on data and assessment issues related to NCLB.

Eight states mentioned the need for the federal government to help build state capacity to carry out the data requirements of NCLB. Officials in Idaho, for example, said their greatest obstacle has been the increased expectation for statistical and research knowledge among state and district staff. Oregon officials commented that they do not have enough time to analyze data. New Mexico interviewees described the difficulty involved in finding people who can analyze data. The state has lost several staff members and has had difficulty recruiting people with assessment and psychometric expertise, especially since salaries are relatively low and the cost of living is high. The university system does not train people in these areas, according to New Mexico interviewees, so it is difficult to build technical knowledge. The lack of people with knowledge of data and assessment issues is also a problem at the school district level, said one state official:

*In the small rural districts, we have a situation where the same person is answering the phones, doing the counseling, you know, maybe driving the bus—who knows what all—and they know absolutely nothing about Microsoft Excel, and they just throw their hands up in frustration because the data is unusable . . . They can’t even open the data file, let alone begin to break it down and analyze it.*

Florida, in response to an open-ended question about the impact of other requirements of NCLB on achievement, described the difficulty it had in retaining and training people skilled in large-scale assessment.

Eleven states mentioned the need for federal funding to implement NCLB in general; among them were several states that specifically cited the need for funding to build state capacity in data and assessment. In addition to other funding needs, Wisconsin officials cited the need for consistent funding to maintain state systems required by the federal government. Virginia
contacts described the difficulties the state has encountered in funding the data management activities required by NCLB, noting that sometimes the funds have been taken from the assessment system in order to support data management. Although not directly related to data and assessment, states also mentioned the need for funding to build capacity to help schools identified for improvement under NCLB.2

Four states requested better guidance or templates from the federal government related to assessments and data. Florida wanted guidance on how to make the transition from an old to a new assessment while still maintaining longitudinal data. Idaho suggested that the federal government create guidelines and standards for education research that would take the form of a technical check list and would explain more clearly to districts what information they should provide to the state. West Virginia proposed that the federal government create a Web site data template so all states would know what is needed and expected and would not be overwhelmed by requests for data in various forms. Wyoming interviewees maintained that better guidance on the state’s alternate assessment system in the earlier stages of NCLB implementation would have prevented some problems and saved time. One state official explained the issue in this way:

[W]hen I say better guidance . . . we had an older [alternate assessment] system. We started down a road of development; decided that wouldn’t work; started down another road of development; decided that wouldn’t work. Finally, we got a clear sense of guidance, and so where we are now, I think, is in a very good place. But we really spent—what?—a year and a half going down a developmental line that would have been clearly out of compliance.

Arizona and Florida raised an issue that, although not directly related to NCLB, has a bearing on state assessments. These states urged the federal government to provide information that would better explain how the National Assessment of Educational Progress (NAEP) sets performance levels on its tests and why these levels differ from states’ proficiency levels on state tests; they proposed that this type of information would help the public better understand these differences and prevent misuse of NAEP and state test data. Arizona officials expressed particular concern that the public may not understand why 60% of students taking the state assessment are considered to “meet standards” according to the state’s definition, while only 20% of the students taking the NAEP are considered proficient according to the NAEP definition of proficiency. Such a wide disparity in percentages between the two assessment systems causes confusion among stakeholders who think that “meet standards” and “proficient” describe the same level of performance. Arizona recommended that the federal government set guidelines for public use of NAEP data to prevent inappropriate use.

FEDERAL ASSISTANCE RELATED TO IMPROVING ACHIEVEMENT

Some states made suggestions about federal assistance related to student achievement.

2 Detailed findings and recommendations about state capacity issues can be found in the CEP report. Educational Architects: Do State Education Agencies Have the Tools Necessary to Implement NCLB?, available at www.cep-dc.org.
Five states suggested that the federal government provide information on research about achievement and best practices for raising achievement. One of these states, Rhode Island described a need for Web-based assistance in best practices, in addition to other suggestions. South Carolina officials expressed frustration at the lack of available research on what works from the U.S. Department of Education and observed that it would be helpful to know what successful states are doing to close achievement gaps, implement NCLB-mandated science assessments, or make other improvements. A Pennsylvania official proposed that a voluntary system of national standards and tests be developed, along with blueprints and pedagogy, noting that this approach is common in other industrialized nations.

Wisconsin officials expressed concern about what they saw as an overemphasis on large-scale assessments as the main data for measuring school success, and said the state would like to see increased flexibility in the measures that states may use for accountability.

OTHER TYPES OF FEDERAL ASSISTANCE

Officials in the 22 states cited several other kinds of assistance that the federal government could provide to help them address the challenges of NCLB. Although these other kinds of assistance are only tangentially related to achievement and assessment, they are summarized here because they may be of interest to federal policymakers and officials in other states. Interviewees’ responses can be grouped into four categories.

- **Better general guidance.** Thirteen states requested improved guidance on NCLB from the federal government. Some of these states, such as South Carolina, simply wanted more timely guidance, while others desired greater clarity. Kentucky, for example, wanted a more reasonable and thorough explanation of how the federal government makes decisions regarding recommended programs or models of instruction.

- **More appropriate subgroup policies.** Four states—Arizona, Delaware, Missouri, and Washington—wanted the federal government to develop more appropriate policies for assessing and providing instructional supports to students with disabilities and English language learners. Missouri, for example, noted the difficulty of complying with what state officials viewed as the strict federal policies regarding the use of test accommodations with special education students.

- **More realistic timelines and more timely information.** Seven states mentioned issues related to deadlines or timelines for implementing NCLB. Most states wanted more time to carry out requirements. Minnesota, for example, referred to the challenge of trying to meet strict timelines before final federal guidance is available and asked for enough time between the issuing of federal guidelines and the deadlines for meeting requirements.

- **Greater flexibility.** Three states—Kentucky, Wisconsin, and Washington—requested more flexibility in implementing NCLB.
Impact of Major Provisions of NCLB on Achievement

HumRRO researchers asked state officials about the impact of each of the following major provisions of NCLB on state efforts to improve student achievement and reduce achievement gaps among subgroups:

- Accountability requirements, including the requirement for districts and schools to demonstrate adequate yearly progress in improving student achievement
- Requirements to disaggregate data by subgroups
- Highly qualified teacher provisions
- Public school choice
- Supplemental educational services
- Requirements to use scientifically based research
- Other requirements

States were asked whether they have conducted any research or collected data to support their conclusions about the impact of these provisions. HumRRO researchers found little evidence that states had done so; typically, officials offered anecdotal evidence to support their conclusions. Those research studies that were cited are briefly described below.

The themes discussed below under each provision are those raised by more than one state. Several other points were raised by a single state. These single-state comments have not been included because they do not directly address the question of how much impact NCLB has had on student achievement trends; were not mentioned often enough to constitute a theme; and have been covered in other analyses of NCLB, including CEP’s past annual reports on NCLB implementation.

ACCOUNTABILITY REQUIREMENTS

Of the 22 states in the study, only Idaho had no response about the impact of the NCLB accountability requirements. Massachusetts officials noted that state already had a very similar testing system in place before NCLB, so it had a head start on accountability. Massachusetts did have to add tests in some grades but generally was able to maintain systems that were already in place rather than fundamentally changing its accountability system.

The remaining 20 phase II states discussed several issues regarding NCLB accountability requirements, ranging from studies some states were conducting to more anecdotal comments about how NCLB requirements affect the states. Three states reported collecting data or conducting studies:

- Missouri referred to a research study that is underway on the unintended and intended consequences of the state’s assessment program. Preliminary data suggests that there are many misconceptions about NCLB and that people are responding to the law’s requirements in many ways, including some ineffective ways.
• **Oregon** reported contracting with an outside firm that made suggestions to improve the state’s AYP data collection and reporting processes.

• **Pennsylvania** reported gathering data on the number of schools that failed to make AYP and their reasons for not passing. The state hopes to use this information to allocate resources more effectively.

Most other responses were anecdotal. Taken together, the responses suggest that in the view of many states, the current accountability requirements are not the most effective policies for identifying low-performing schools or increasing student achievement.

Three states, **Delaware, Illinois,** and **Wisconsin,** suggested that little evidence exists to demonstrate that the AYP requirements and accountability sanctions are actually helping schools raise achievement. **Delaware** noted that it had not seen research by the federal government on the effectiveness of consequences of not meeting NCLB. **Wisconsin** asked where the evidence exists showing that labeling schools as in need of improvement actually improves student achievement; the state also observed that NCLB points out deficiencies without really helping schools improve.

States cited what they viewed as flaws in the NCLB accountability provisions and sometimes made suggestions for changing these provisions:

• **Opposition to “all or nothing” rule and desire for differentiated consequences.** Five states expressed dissatisfaction with the “all or nothing” policy in which a school fails to make AYP if just one subgroup falls short of AYP targets. Eight states suggested that NCLB switch to a system of differentiated consequences to help distinguish between schools in which just one subgroup fails to meet AYP targets and those in which many subgroups fall short. **Minnesota,** for example, proposed that the law include levels of sanctions that would credit schools for progress or would focus on schools with chronic underperformance by one subgroup.

• **Lack of rewards for progress.** Two states, **Minnesota** and **South Carolina,** suggested that NCLB should reward schools for making substantial progress toward achievement goals, even if the goals are not met immediately.

• **Desire for flexible growth models.** Three states—**Delaware, Florida,** and **Washington**—expressed support for an accountability system based on growth models that track individual students’ achievement as they progress from grade to grade rather than setting fixed targets for achievement within a grade. (This idea is being piloted in some states with permission from ED; **Florida** is a state that has been approved to use a growth model.) **Delaware** said that states should be able to choose which growth model works for them—for example, vertically moderated or articulated standards.

• **Peer review process.** **Delaware** and **Kentucky** voiced a desire for changes to the process for conducting peer reviews of state assessment systems. **Delaware** pointed to the need
for better training and clearer criteria for those doing peer reviews, and Kentucky said it would like assurances that the review process is consistent across states.

- **Conflict with state accountability.** Kentucky and Washington cited the problems arising from inconsistencies between their state accountability and assessment system and the federal one. Kentucky reported difficulties implementing the federal AYP requirements because its state system bases decisions about schools on two years of data, which the state views as a sounder approach than basing decisions on one year’s data. Washington noted that the federal government did not consider the good models that already existed in some states but instead tried to fit every system into the same model.

**DISAGGREGATION OF DATA BY SUBGROUPS**

As noted earlier in this report, the No Child Left Behind Act requires states and school districts to break out test scores and other data for various subgroups of students. In the interviews, states identified both positive and negative consequences of this disaggregation.

Eight states said that the attention placed on subgroups as a result of disaggregation had helped to improve achievement or brought related benefits. Examples include the following:

- **North Carolina** indicated that disaggregation has led to improved student achievement by focusing on the performance of all students.

- **Michigan** and **South Carolina** suggested that disaggregation of data has drawn more attention to subgroups whose performance may have been overlooked in the past. **South Carolina** added that disaggregation has helped refute myths about some groups of students being unteachable.

- **Washington** explained that as a result of disaggregation, everyone in a school cares about the achievement of all subgroups.

Six states raised possible negative consequences from disaggregation.

- **South Carolina, Delaware,** and **Washington** expressed concern that a single student sometimes counts as a member of more than one subgroup—for example, the test scores of an English language learner from a low-income family who is also identified as disabled would count for each of these subgroups. As one state official observed, “you keep getting hit and hit and hit and hit for the same child.” **Delaware** further noted that a single student who counts in multiple subgroups could cause a school to miss its AYP target.

- In a related issue, **Pennsylvania** suggested implementing varying degrees of consequences for not making AYP. Under the current system, the consequences of missing AYP for a single subgroup are the same as missing AYP for all subgroups.
Missouri and New Mexico noted that some schools do not make AYP because of the low performance of the subgroup of students with disabilities. Missouri officials added that while English language learners can learn enough English to be reclassified out of that subgroup, most students with disabilities are not reclassified as no longer needing special education.

Four states brought up research issues related to data disaggregation. Massachusetts, for instance, questioned what research had been done related to subgroups and noted that most states do not do much research because of cost. Oregon officials remarked that although the state was working on closing achievement gaps, the state was concerned about a lack of good research and best practices. West Virginia pointed to studies conducted on students with disabilities that were available on the state’s Web site.

HIGHLY QUALIFIED TEACHERS

States are required by NCLB to ensure that all teachers of core academic subjects are “highly qualified” according to the law’s definition, which includes having a bachelor’s degree, being fully certified, and demonstrating knowledge and skills in the subjects they teach by having sufficient subject-matter coursework, passing a state test, or meeting other state criteria.

Two states maintained that they see no evidence that highly qualified teachers have an impact on student performance. Arizona said that there were no empirical data to determine a connection between highly qualified teachers and student performance. Michigan similarly said that there were no relevant data to substantiate the effect of highly qualified teachers, noting that the research literature does not show a strong link between a teacher’s educational level and student outcomes.

Four states expressed general support for the highly qualified teacher requirements, with some reservations. Kentucky said that reporting of teacher qualifications has stimulated public dialogue in a beneficial way, although questions about the definition of highly qualified remain. Delaware also said that the state now tracks teachers and students and is producing “real and accurate” reports on highly qualified teachers rather than estimates. South Carolina said that the teacher requirements have been positive and noted that the state has tried to show leadership in helping teachers become highly qualified.

Following are the most common concerns or suggestions raised by phase II states about the NCLB teacher requirements:

- **Challenges for special education and rural schools.** Three states described difficulties finding enough highly qualified teachers for rural districts or schools. Four states noted the difficulties of finding highly qualified special education teachers or ensuring that special education teachers who teach multiple subjects are highly qualified in all the subjects taught.

- **State versus federal requirements.** Three states noted that the NCLB highly qualified teacher requirements sometimes conflict with state certification requirements.
• **Highly effective vs. highly qualified.** Several states commented that the NCLB definition of highly qualified does not mean that teachers are effective in the classroom. Delaware, for example, noted that other criteria, such as the ability to relate to students, should be considered, and that a better term would be highly effective, rather than highly qualified. Not all of these states wanted to see the NCLB requirements shift toward a definition of teacher effectiveness, however. Virginia said that attempts to label teachers as highly effective would be seen as an intrusive form of teacher evaluation and recommended that this issue be left to the state or school districts. South Carolina noted that shifting the federal definition in a way that raised the bar even further would exacerbate difficulties of finding highly qualified teachers in all core subjects and in shortage areas like special education.

**PUBLIC SCHOOL CHOICE**

Schools that have not made AYP for two consecutive years must allow parents the option to transfer their children to another public school that has not been identified for improvement. Many of the states interviewed expressed general dissatisfaction with the choice provisions. Issues raised by multiple states included the following:

• **Not a viable or needed option.** Ten states said that choice was not a viable or necessary option. In some cases this was because districts had just one school per grade span and could not offer another school to provide choice. In other cases, this was because parents already had school choice through charter school or state open enrollment policies (the latter was mentioned by five states).

• **Reluctance of parents to transfer.** Seven states indicated that parents are reluctant to switch their children’s schools, often due to convenience and familiarity.

• **Interdistrict choice.** Michigan and Rhode Island reported that some districts were reluctant to accept choice transfer students from other districts, particularly students from underperforming schools in urban districts.

• **Choice for higher-achieving students.** New Mexico, Wisconsin, and Virginia expressed concern about school choice being used by parents whose children are not struggling. Virginia officials pointed out that this would drain schools of funds available for students in the subgroup having trouble.

• **Switch order of sanctions.** Five states said they believe that supplemental services should be offered before school choice is permitted.

**SUPPLEMENTAL EDUCATIONAL SERVICES**

Schools that have not made AYP for three consecutive years must make available free tutoring or remedial services, known as supplemental educational services, to children from low-income
families. Parents may choose a public, private nonprofit, or private for-profit provider to deliver these services from a list of approved providers for their district.

A common theme about supplemental services, raised by nine states, is the difficulty of evaluating whether these services are effective in improving achievement among participating students. One such state, Wisconsin said there is little evidence of school improvement as a result of the supplemental services requirement, while Michigan noted that there is no way to evaluate effectiveness. New Mexico and Arizona observed that the sheer volume of data related to supplemental services from many different public and private providers has hindered evaluation efforts:

- New Mexico said that it is trying to make supplemental service providers more responsible. The state has completed three years of state-funded evaluations of supplemental services and also requires providers to use a state-approved assessment as part of the approval process. Still, state officials noted, there is no way to efficiently monitor providers. The services provided are so diverse in terms of type, delivery method, and duration that it is difficult to gauge success. Each provider has pre- and post-measures, but it is risky to use those for evaluations.

- Arizona indicated that it is difficult to track every student throughout their program participation. The state has a student management system, and districts and service providers have to comply. Program participation data, which are recorded annually, have not been especially reliable, but the state hopes to obtain some longer-term data this year.

Three states cited lack of funding as a factor contributing to problems with evaluation. Delaware emphasized that the federal government provides no money to evaluate supplemental services. Washington similarly noted that states need funding to measure the effectiveness of these services. Rhode Island said that despite the requirement for districts to set aside 20% of Title I money to implement choice and supplemental services, districts lacked the capacity to implement and monitor the necessary services.

Six states discussed the issue of research evidence—or lack thereof—for the effectiveness of supplemental services. Three of these states—Kentucky, Michigan, and Minnesota—mentioned efforts at the state level to evaluate these services. Kentucky has let a contract for an outside evaluation of supplemental services that will look at the quality of providers and the extent to which achievement has improved for participating students. Michigan reported that it had a request-for-proposal (RFP) pending to study the effectiveness of supplemental services, but it has taken 14 months for the RFP to be released, which has slowed down research on this issue. Minnesota reported that there is interest in the state legislature about looking at the quality of supplemental services and whether they align with the state’s standards.

Three other states were dissatisfied with the research aspect of supplemental services. Oregon officials contended that choice and supplemental services are not based on good research. Wisconsin made the point that supplemental services have taken funding from classroom strategies to assist low-performing children. Washington said that community supplemental
service providers should have to adhere to the same standards as schools and should be required to prove they can add something to the curriculum.

Several other comments about supplemental services were raised by more than one state:

- **Parent decision making.** Arizona and Delaware officials pointed out that parents often consider several factors in addition to the quality of services when selecting a supplemental service provider for their child; examples include convenience of the location, transportation distances, and familiarity. Many non-school providers are unfamiliar to parents.

- **Rural service delivery problems.** Three states with large numbers of rural schools—Arizona, Idaho, and Wyoming—described the difficulties involved in delivering supplemental services to rural schools. Even when online providers are used to serve students in isolated areas, it is difficult to find local people to supervise students and monitor internet connections, or provide technology support.

- **Funding concerns.** Eight states had comments about funding and supplemental services. Michigan, for example, stated that more students are eligible to receive services than can be funded to receive them.

- **Delays.** Idaho and Kentucky noted the difficulty of delivering services in a timely manner. Idaho is implementing most supplemental services through online providers, but schools face technology support challenges. The state reported being a year behind on its supplemental service features.

**SCIENTIFICALLY BASED RESEARCH**

NCLB requires the use of programs and instructional practices that are based on scientific research. States expressed concern about various aspects of this provision; what follows are comments raised by multiple states.

Three states said they believed aspects of NCLB itself are not based on scientific research. Michigan, for instance, had asked officials at the U.S. Department of Education for research on the impact of NCLB sanctions on achievement, but said it was told that no research on that topic exists. Wisconsin noted that the consequences of failing to make NCLB requirements were not grounded in research.

Four states said that not enough programs had been endorsed by research or that scientifically based research was lacking. Wyoming, for example, said that the state is using the scientifically based research approach but that little such research has been done. Noting that very few studies meet the gold standard, Virginia proposed that the federal government recognize studies that are less than gold standard but still adequate research.
Three states expressed concern about what they saw as only partial implementation of scientifically based research. Idaho, for example, asserted that while research on reading instruction and curriculum is available, research on math and science is limited.

A clearer definition of scientifically based research is needed, according to three states. For example, South Carolina made the point that educational vendors can label their programs, textbooks, and instructional strategies as scientifically based without appropriate evidence. Wisconsin stated that innovation seemed to be limited by the current definition of scientifically based research.

Three states described efforts they have undertaken at the state level to implement the NCLB provisions for scientifically based research. Pennsylvania listed a Web site that the state uses to find scientifically based research. Arizona explained how its participation in the Comprehensive School Reform grant program requires use of scientifically based research. Massachusetts described how it was reviving its office of research, planning, and development after its dissolution at the beginning of NCLB to focus on accountability.

**OTHER NCLB REQUIREMENTS**

At least six states did not answer the question about other requirements of NCLB that had an impact on achievement, probably because it was not asked due to time constraints. States that did answer the question often reiterated or clarified their responses about topics discussed elsewhere in this report. New issues were often raised by just one state and are not reported here.

Two states expressed concern about the appropriateness of assessing certain groups. Kentucky was concerned about the appropriateness of testing students in early grades, and Rhode Island with assessing students with disabilities against grade-level expectations without some modifications.

Four states pointed to what they saw as positive requirements of NCLB. These included the focus on standards for what students should learn (Kentucky and Rhode Island), and restructuring and transparency in the curriculum and increased parental involvement (Michigan).

**Methodology for Phase II**

**PROTOCOL DEVELOPMENT AND PILOT STATES**

HumRRO staff developed the interview protocol and revised it after consultations with CEP and an expert panel assembled by CEP to advise on phases I and II of the comprehensive achievement study. (Information about the members and roles of the expert panel can be found in the phase I report, *Answering the Question That Matters Most.*) To further refine the protocol and the interview process, HumRRO conducted interviews in two pilot states, North Carolina and Michigan, on January 18 and February 1, 2007. The final protocol contained a series of questions about the impact of various NCLB provisions that were common to all 22 phase II
states, as well as specific questions about achievement trends tailored to each particular state. Appendix B contains a sample interview protocol without the state-specific trends.

SAMPLE SELECTION

After taking into account the study’s budget and schedule and preliminary findings from the early rounds of phase I data collection, CEP and HumRRO decided to conduct phase II interviews in a sample of 20 states, in addition to the two pilot states. After consulting with CEP and the expert panel, HumRRO selected states for phase II interviews based on the availability of comparable, relevant state assessment data. This policy was recommended by the expert panel after it reviewed results from the two phase II pilot states, and was intended to focus phase II on the states with the greatest potential for providing useful additional information. First priority for phase II interviews went to states with continuous comparable data on both the percentages of students scoring at the proficient level and the mean scale scores on state tests for a period that included both pre- and post-NCLB trends (1999-2006). Additional states were added to the phase II sample if they had at least some of these data available from 1999 to 2006. And finally, three states were included to gather state-level information from states in which CEP planned an upcoming district-level study of student achievement.

The 20 phase II states were selected before states had verified the test data collected for phase I of the study. During the phase I verification process, some states provided additional data that changed their category for data availability. Because the phase II sample had already been selected at this point and interviews were being planned or were underway, the phase II sample was not changed to reflect these changes in category. Appendix A lists all the phase II states.

INTERVIEW PROCESS

The HumRRO research team conducted interviews with officials in the 20 states (not counting the two pilot states) from February 22, 2007, through May 10, 2007.

Before each site visit the lead HumRRO interviewer e-mailed an information packet to the point of contact at the state department of education to distribute to interview participants. These packets included full copies of the protocol, which contained our major conclusions about achievement trends in the state, planned interview questions, and line graphs depicting math and reading proficiency data across time. The pilot states received less information prior to the site visits; North Carolina, the first pilot, received only a brief description of planned interview topics, and Michigan received only highlights of planned interview questions.

When HumRRO staff conducted site visits, they brought a full copy of the state’s profile of achievement data and trends, as well as additional information obtained during data verification. Packets included the following:

- Tables and graphs presenting data on the percentage of students scoring at the proficient level in math and reading
- Tables and graphs presenting math and reading scale score data
- Achievement gap tables showing percentages proficient in reading and math for various subgroups
• Achievement gap tables showing reading and math scale score data for various subgroups
• Test characteristics file describing state’s assessment
• Copy of signed data verification checklist

Generally, these results were simply discussed with interview participants. If state education staff asked to see a particular graph or table, however, interviewers shared that information. Typically, department of education staff did not ask to see additional results.

Typically, two HumRRO researchers traveled to state departments of education; however, one interview was conducted by telephone due to state scheduling constraints. The number of state officials present during interviews ranged from one to eight. During the interviews, one researcher served as the lead interviewer and the other researcher was responsible for recording responses by taking detailed notes and ensuring that a digital recorder was operating. Interviews usually lasted from two to three hours. Generally, state officials were not limited in their responses but were free to answer as much or as little as they chose. For example, instead of asking state officials to provide the three most important kinds of support the federal government might provide to help address challenges, interviewers asked officials what kinds of technical support, guidelines, auditing, or resources the federal government might provide.

QUALITATIVE DATA ANALYSIS

After an interview was completed, the researcher responsible for taking notes produced and compiled transcripts by computer from the digital recordings. This researcher produced summary notes for each state, using the verbatim transcripts as needed for clarification. The notes were shared with the lead interviewer to ensure accuracy.

Next, a HumRRO researcher who had taken part in several state interviews developed an outline of broad themes for the phase II report, based on feedback from CEP, the expert panel, and experience gained during the interview process. This outline became the framework for additional qualitative analysis.

Because the interviewees were not limited in their responses, each response to a single question might include several “answers,” which we called comments. It was not uncommon for a state to provide several comments for a particular question, so the number of total comments for any question usually exceeded 22, the number of phase II states. The researcher read comments, grouped them into themes, and used these themes to develop this report.

To keep the report to a manageable length, many comments made by state officials have been omitted from this text. Typically a few representative comments—quotations, paraphrases, or summaries—have been included to illustrate a particular theme. Often the comments omitted are those that express a singular view, particular to one state or one person and not repeated in other states. The direct quotations used in the report were taken from the verbatim interview transcripts. Paraphrased comments were taken from the interview summary.
# Appendix A: Sample of States

Personnel from the following states were interviewed for this study.

<table>
<thead>
<tr>
<th>State</th>
<th>Interview Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>North Carolina (pilot #1)</td>
<td>1/18/07</td>
</tr>
<tr>
<td>Michigan (pilot #2)</td>
<td>2/1/07</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>2/22/07</td>
</tr>
<tr>
<td>Louisiana</td>
<td>2/27/07</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>2/28/07</td>
</tr>
<tr>
<td>Kentucky</td>
<td>3/5/07</td>
</tr>
<tr>
<td>Oregon</td>
<td>3/7/07</td>
</tr>
<tr>
<td>Illinois</td>
<td>3/8/07</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>3/9/07</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>3/13/07</td>
</tr>
<tr>
<td>New Mexico</td>
<td>3/19/07</td>
</tr>
<tr>
<td>Idaho</td>
<td>3/19/07</td>
</tr>
<tr>
<td>Delaware</td>
<td>3/20/07</td>
</tr>
<tr>
<td>South Carolina</td>
<td>3/21/07</td>
</tr>
<tr>
<td>Minnesota</td>
<td>3/26/07</td>
</tr>
<tr>
<td>Arizona</td>
<td>3/27/07</td>
</tr>
<tr>
<td>West Virginia</td>
<td>4/9/07</td>
</tr>
<tr>
<td>Missouri</td>
<td>4/12/07</td>
</tr>
<tr>
<td>Wyoming</td>
<td>4/16/07</td>
</tr>
<tr>
<td>Washington</td>
<td>4/18/07</td>
</tr>
<tr>
<td>Florida</td>
<td>5/1/07</td>
</tr>
<tr>
<td>Virginia</td>
<td>5/10/07</td>
</tr>
</tbody>
</table>
Appendix B: State Comments on Phase I Data

During the phase II interviews, states were asked whether they agreed with the findings about overall achievement and achievement gaps drawn from their phase I data, and whether the initial phase I analyses were missing any key components that might lead to substantially different conclusions. Generally, states agreed with the phase I findings. Some states, however, made clarifications or raised concerns, described below. CEP addressed the issues raised by states in various ways, also described below.

STUDENTS WITH DISABILITIES AND ELLS

Three states raised issues related to changing testing policies for students with disabilities and English language learners. South Carolina and Virginia reported making changes in testing policies for students with disabilities, such as phasing out “out-of-level” testing and introducing alternate assessments. In addition, Virginia reported changes in the assessments given to ELLs. Wisconsin interviewees pointed out that the data on which phase I conclusions had been based had left out ELLs who were tested with an alternate assessment, and that this had likely impacted the apparent gap reduction.

In several places throughout the phase I report, CEP urged readers to be cautious in drawing conclusions about trends for students with disabilities and ELLs because of changing test policies for these subgroups. In addition, the phase I report refrained from highlighting trends for these subgroups in the summary bullets for all states. Finally, data for these two subgroups were shaded in the state data profiles to show that they should be viewed differently from data for other subgroups.

OTHER CLARIFICATIONS

Several states clarified points relating to their phase I data.

- West Virginia and Wyoming reiterated that their minority populations were very small. Recognizing that inferring meaningful trends for small groups is unreliable, CEP’s phase I report and state profiles flagged with a footnote those subgroups that were small or had changed significantly in size. CEP also included cautions throughout the report about interpreting results for these flagged subgroups. In addition, the phase I report did not highlight trends for these small or changing subgroups.

- Arizona officials noted that only grade 10 data should be used to analyze high school assessment results; students who take the high school assessment in grades 11 and 12 are retaking the test because they did not pass it the first time. State officials also described pre-2005 data as unreliable because state standards changed several times and demographic characteristics had been entered by students. Consistent with this clarification, CEP’s phase I state profile of Arizona included only grade 10 students in the high school assessment results and did not report pre-2005 data.
• **Oregon** officials said that data on achievement gaps reported for NCLB purposes and reflected in the state’s interview packet do not fully represent the state's efforts and accomplishments in reducing gaps. Oregon participates in the State Action for Education Leadership Program (SAELP) through the Wallace Foundation, which emphasizes the cultural competence of education staff and promotes closing the achievement gap in literacy. Oregon uses data other than NCLB data to analyze achievement gaps, plan gap reduction efforts, and identify schools that are closing gaps. Although this additional information was valuable for phase II report of the study, it did not fit the phase I data framework, which focused on assessments and data used for NCLB.

**DATA MISSING AFTER PHASE I VERIFICATION**

During the interviews, three states offered to provide missing data related to mean scale scores and standard deviations that had not been made available within the extended deadline for verifying phase I data. In particular, the following three states made this type of request:

• **Florida** and **Illinois** officials said that the state could make available mean scale scores and standard deviation data, which had not been provided during the phase I verification process.

• **Idaho** interviewees noted that the effect size analyses for phase I did not include data for students who were *not* eligible for free or reduced-price lunch or students who were *not* in special education. This was because the raw data necessary to make effect size calculations (mean scale scores and standard deviations) for these two groups had not been made available within the phase I deadline. In Idaho and other states where these data were not available in time, the phase I analyses determined gaps for low-income students (those eligible for free or reduced-price lunch) and students with disabilities (those served by special education) by comparing these two subgroups with *all* students in the state, rather than with non-low-income or non-disabled students.

Since the phase II interviews were conducted after states had already been given considerable time to verify their phase I data and after the initial data analyses had been completed, CEP could not accommodate these requests to add new data and still meet the publication schedule for the phase I report, which was developed to ensure that policymakers would have the phase I results in time to inform NCLB reauthorization.
Appendix C: Sample Interview Protocol

The following text served as a shell for each state’s interview protocol. Bracketed sections indicate where specific state information was inserted.

Interviewers requested a group interview that included the state superintendent, deputy superintendent, and assessment director, lasting three hours. Due to time and availability constraints of state personnel, some interviews were scheduled for two hours, and the personnel interviewed varied.

INTRODUCTION

Before we begin, I’d like to clarify how the information we obtain during this interview will be used. Because we are tying these site visits so closely with state assessment results, we plan to make state-specific references regarding site visit content in our report. We will NOT, however, quote any of you specifically by name without obtaining your personal permission to include that direct quote in the report. Within two weeks of this site visit, we will send you the transcript of this interview so you can correct any errors of transcription. When our report has been prepared, sometime in the spring, we will send you the portion of the report that pertains to your state so you can correct any factual errors and review any quotations included.

We are recording these site visits so we can obtain a transcript of the interview and make sure that we accurately note and report everything from the visit. It would be helpful for the transcriptionist if we could all go around and state our names. [Start this process].

Now, please let me give you a little bit of additional background about this study and the purpose of this site visit. This study is intended to provide an independent data-based analysis to inform the reauthorization discussion for NCLB.

The study is designed to address two of the most important questions about NCLB:

- Has student achievement increased since the passage of No Child Left Behind?
- Have achievement gaps decreased since the passage of No Child Left Behind?

We want to compare our impressions of your state’s assessment data in regards to these two questions with your impressions. If our conclusions about achievement trends and narrowing achievement gaps differ from yours based on the data we have, we’d like to explore those differences. In addition, this study will also address the availability of assessment data for conducting such research. One of the purposes of these visits, in addition to addressing those two major research questions, is to learn more about the types of data that states make available and to understand why states make changes to their assessment systems.

The study is being conducted in two phases. The first phase has been completed. It involved gathering NCLB information about student achievement before and after NCLB for all 50 states. We are now conducting the second phase in which we are gathering more detailed information on achievement and gap trends and identifying possible links between those findings and state
activities to satisfy NCLB requirements. The reports generated from this study have considerable potential to inform policy related to the reauthorization of NCLB.

CONDUCTING INTERVIEW

I’m going to go over some of the conclusions we reached through analyzing the [state name] assessment data we obtained during phase I of this study. [Review conclusions about state.]

We are now interested in getting your take on these general patterns and in learning anything more you can tell us to help us understand the patterns of student achievement in your state.

Question #1: Do you agree with our conclusions about the overall achievement data provided by your state?

A. How would you characterize overall student achievement in [state] since 1999?
B. When we were conducting our analyses, were we missing any key component that might have led to substantially different conclusions?
C. [Specific state questions]
D. Do you collect any additional information to help track state-level achievement trends? Are those data publicly available?

The second major research question for our study involves achievement gaps. [Review conclusions about state.]

Question #2: Do you agree with our conclusions about the gap data provided by your state?

A. Are we missing some key component that might lead to a substantially different conclusion for any subgroup?
B. [State-specific questions]
C. Are there subgroups for which your state has made special efforts to reduce achievement gaps? If so, which group(s)? What has been done? Have you collected any additional data that might indicate that those gaps are narrowing?
D. Do you have any evidence to speak to whether achievement gaps were closing prior to 2002? What was that evidence based on (percent proficient, mean scale scores, etc.)?
E. Does your state collect additional data related to the performance of subgroups that might inform our existing analyses?
F. Does your state collect data from programs implemented to reduce achievement gaps that are conducted at the district level? Have districts implemented programs that have led to substantial gap reductions? What evidence is available regarding district-level gap reduction?

Question #3: A secondary purpose of this site visit is to address some questions about data availability and data quality under NCLB. We are also trying to gain understanding of the

---

Interview questions should be read more or less verbatim; however, the interviewer should provide clarification as needed for the interviewee to respond.
changes that state assessment systems make, why those changes are implemented, and whether they are effective.

A. [State-specific questions]
B. What changes did you need to make to your assessment system when NCLB was implemented in 2002?
C. Were any of the changes made to the assessment system since the implementation of NCLB directly related to student achievement, either in terms of statewide achievement trends or in terms of achievement gaps?
D. What future changes are planned for state assessment in [state]?
E. We were able to obtain most data for [state] on the web but had to acquire standard deviations (SDs) during the data verification process. Are there any plans to make all data including SDs more easily accessible via the internet?
F. How does your state plan to maintain or change the level of publicly available data in upcoming years? Does [state] plan to make additional data related to achievement trends and achievement gaps available to researchers or education consumers in the future? If we were to repeat this study in 2012, would [state] have sufficient trend data (including means, SDs) to clearly answer our research questions? Please explain.
G. Does your state monitor instructional impact? If so, have there been substantial instructional changes (either statewide or for individual subgroups) since 2002?
H. What kinds of technical support, guidelines, auditing, resources, etc. might the federal government provide that would help address these challenges?

Question #4: As you know, NCLB has several component features. Let’s discuss the impact—positive or negative—of these features as they relate to your efforts to increase student achievement and reduce gaps.

I’m specifically interested in studies you have conducted or data you have collected that provide evidence for any conclusions. If those studies point toward specific changes that might lead to improvements to NCLB during reauthorization, we’d be particularly interested in hearing them.

A. Requirement for adequate yearly progress and accountability
B. Highly qualified teachers
C. Disaggregation of data
D. Supplemental services
E. Choice
F. Scientifically-based research
G. Requirements I haven’t mentioned

CLOSING INTERVIEW

That was my last question. I want to thank you very much for participating in this interview. Your input will be very valuable. We genuinely appreciate the time, thought, and energy you’ve put into providing us feedback about your state’s student achievement and achievement gaps. Is there anything that you want to share with us that we have not previously discussed (what didn’t we ask that we should have)?
Credits and Acknowledgments

This report was researched and written by Lisa E. Koger, Hilary L. Campbell, Arthur A. Thacker, D. E. “Sunny” Becker, and Lauress L. Wise, all of the Human Resources Research Organization. The report was edited by Nancy Kober, a consultant to the Center on Education Policy.

The Center on Education Policy and HumRRO extend their deepest gratitude to the state superintendents, deputy superintendents, assessment directors, and other state officials who agreed to be interviewed for phase II of this study. Their comments were most enlightening and will help build a wider understanding of state implementation of NCLB. CEP would also like to thank the members of the expert panel that provided invaluable advice on the design of phases I and II of the achievement study; they include Laura Hamilton, Eric Hanushek, Frederick Hess, Robert Linn, and W. James Popham.

The Center on Education Policy is grateful to The Ewing Marion Kauffman Foundation, The Carnegie Corporation, The William and Flora Hewlett Foundation, and The Ford Foundation for their support of the Center’s study of achievement and the No Child Left Behind Act. The George Gund Foundation, The MacArthur Foundation, and the Phi Delta Kappa International Foundation also provide the Center with general support funding that assisted in this endeavor.

About the Center on Education Policy

Based in Washington, D.C., and founded in January 1995 by Jack Jennings, the Center on Education Policy is a national independent advocate for public education and for more effective public schools. The Center works to help Americans better understand the role of public education in a democracy and the need to improve the academic quality of public schools. We do not represent any special interests. Instead, we help citizens make sense of the conflicting opinions and perceptions about public education and create the conditions that will lead to better public schools.