Improving Data Quality for Title I Standards, Assessments, and Accountability Reporting

Guidelines For States, LEAs, and Schools

[Non-Regulatory Guidance]
April 2006

U.S. Department of Education

Office of Elementary and Secondary Education

Table of Contents

This paper was produced by DTI Associates, A Haverstick Company, under U.S. Department of Education Contract No. ED-01-CO-0066/0009.

FOREWORD: ABOUT THESE GUIDELINES

Spurred by the *No Child Left Behind Act* of 2001, virtually every educational reform program now includes an accountability component that requires sound data collection and reporting. Improving data quality has thus emerged as a high priority for educators and policymakers across the country. The list of programs for which data quality is relevant is extensive, and the scope of the issues involved is vast. This document does not attempt to cover the entire range of data quality issues. Small rural districts, for example, have a unique set of challenges that are not fully addressed in these general guidelines. The document also does not intend to provide comprehensive guidance for all Federal education programs. As its title – *Improving Data Quality for Title I Standards, Assessments, and Accountability Reporting* – suggests, this set of guidelines addresses only one discrete slice of the universe of data quality issues: those associated with the annual Report Card required of all States, local educational agencies (LEAs) and schools receiving Title I, Part A funds under Sec. 1111(h) of *No Child Left Behind* (NCLB).

Recommendations from the U.S. Department of Education's Office of the Inspector General (OIG) and the Government Accountability Office (GAO) provided the impetus for these guidelines. OIG noted in February 2004 that "assessment scoring errors could potentially jeopardize the successful implementation of NCLB," and recommended that the U.S. Department of Education develop "best practices for management controls over scoring of state assessments." In September 2004, GAO identified numerous data quality problems in the States related to NCLB accountability and reiterated OIG's call for guidelines. This document, focused on State, LEA, and school Title I Report Cards, tailors its guidance specifically to address the OIG/GAO recommendations.

Because these guidelines focus on NCLB Report Cards for Title I standards, assessments, and accountability, important data quality issues relevant to a number of other Federal and State programs are not addressed. For example, data reporting requirements related to NCLB supplemental educational services eligibility and participation, public school choice, persistently dangerous schools, limited English proficient students, and special education are beyond the scope of this document. The U.S. Department of Education has a number of data quality initiatives underway that cover a broader array of issues:

- **ED*Facts*** is a collaborative effort among the U.S. Department of Education, State Education Agencies, and industry partners to centralize State-reported data into one Federally coordinated, K-12 education data repository located in the U.S. Department of Education. The purpose of ED*Facts* is to increase the focus on outcomes and accountability rather than process; provide robust K-12 performance data by integrating student achievement and Federal program performance data; reduce data collection burden for the Department and the States; and provide data for planning, policy, and program management. **ED*Facts*** builds upon the foundation of the Education Data Exchange Network (EDEN), along with internal ED data sources, to create an education focused business intelligence tool for the 21st Century.

- The **National Center for Education Statistics** (NCES) is charged with the responsibility of working with other components of the U.S. Department of Education and with State and local educational institutions to improve the quality of education data. NCES is responsible for a grant program that provides funding to States for the development of high-quality longitudinal student information systems needed to compute a true cohort graduation rate. At the elementary/ secondary level, NCES recently released a *Cooperative System Guide to Building a Culture of Data Quality*, aimed at schools and school districts. At the postsecondary level, NCES has redesigned the Integrated Postsecondary Education Data System from a paper system to an online data collection, helping improve the quality of these data, while at the same time increasing their utility.

- The **Office of Vocational and Adult Education** (OVAE) and States have significantly improved the quality of state adult education performance data over the last several years, as States have implemented the National Reporting System for adult education. OVAE has enhanced States' capacity to provide high-quality assessment data by developing state data quality standards that identify the policies, processes, and materials that states and local programs should have in place to collect valid and reliable data.

- The **Office of Special Education Programs** (OSEP), within the Office of Special Education and Rehabilitative Services, has implemented a data-dependent accountability system, the Continuous Improvement and Focused Monitoring System (CIFMS), that has focused on State performance on a number of performance measures and regulation-based compliance requirements. In support of CIFMS, the office has provided ongoing technical assistance and data reviews to support States' efforts to provide valid, reliable, and accurate data related to the implementation of the Individuals with Disabilities Education Act.

- The **Office of Safe and Drug-Free Schools** (OSDFS) is improving State data systems and linking those improvement activities to other U.S. Department of Education initiatives. The *No Child Left Behind Act* requires that each State collect and report to the public certain school crime and safety data elements, such as truancy and the incidence of violence and drug-related offenses. OSDFS is currently implementing two initiatives designed to support improvement in the quality of data related to youth drug and violence prevention programs. Grants have been awarded to 17 States to provide support for enhancing efforts to collect data required by the Uniform Management Information and Reporting Systems (UMIRS) provisions in Title IV of NCLB (Safe and Drug-Free Schools and Communities Act). A second initiative involves the development of a uniform data set that includes the UMIRS data elements.

Readers interested in learning more about Federal data quality initiatives can consult the U.S. Department of Education's website at http://www.ed.gov.

1. Introduction

1.1 The Education Data Quality Guidelines: Purpose and Scope

The accountability provisions included in the *No Child Left Behind Act* of 2001 (NCLB) significantly increased the urgency for States, local educational agencies (LEAs), and local schools to produce accurate, reliable, high-quality educational data. With determinations of whether or not schools and LEAs make "adequate yearly progress" (AYP) dependent upon their student achievement data, it has never been more important for State and local data systems and reporting processes to produce accurate, reliable information. To assist in this effort, the U.S. Department of Education's Office of Elementary and Secondary Education has developed this set of education data quality guidelines.

Data Matters

The Government Accountability Office (GAO) cautioned, in a September 2004 report on NCLB, that "measuring achievement with faulty data can lead to inaccurate information on schools meeting proficiency goals." Both GAO and the U.S. Department of Education's Office of the Inspector General (OIG) have pointed out that the consequences of poor-quality NCLB data can be serious, including the possibility that schools and districts could be erroneously identified as being in need of improvement or corrective action.

GAO's 2004 report, "Improvements Needed in Education's Process for Tracking States' Implementation of Key Provisions," is available on the GAO website (www.gao.gov). Search for report GAO-04-734, or go to http://www.gao.gov/new.items/d04734.pdf.

A number of high-profile efforts are currently underway to improve the quality of the data reported to the U.S. Department of Education. Initiatives such as ED*Facts*, the Schools Interoperability Framework (SIF), the Data Quality Campaign, and numerous other Federal and State education data reform projects have begun the process of implementing systemic, long-term change in the way data are collected, analyzed, and reported. These efforts to reshape the foundations of current data management structures will take a considerable amount of time and resources to achieve. Until these systemic changes are complete, it is vitally important for States and localities to implement the best enhanced management controls possible over the data that are being used to make key judgments about AYP, funding, NCLB accountability, and other State and local education policies.

These guidelines do not impose any additional legal requirements beyond what is in the law, but rather are intended to provide information regarding good practices in data collection. The guidelines are intended to provide shorter-term, relatively inexpensive, interim procedures that States and localities can use *now* to improve data quality while more systemic restructuring is in progress. In some cases, such as in developing infrastructure and training staff, "short-term" measures will have a long-term impact on data quality procedures. The guidelines are built around the basic data elements required for NCLB Report Card reporting, but are designed to be applicable to other K-12 data that States, LEAs, and schools collect as well. Most States have had accountability systems long before NCLB, and States, LEAs, and schools collect data for a wide variety of purposes beyond Federal NCLB Report Card reporting. What these guidelines term "NCLB data" (from the Federal perspective) are in many cases data elements that States, LEAs, and schools have collected and analyzed since long before NCLB.

1.2 Federal Data Requirements for Report Cards Under No Child Left Behind

Broadly speaking, all States and LEAs must collect and report information on their academic assessments in reading/language arts and math (and science beginning in 2007-08), AYP results, and teachers' qualifications. Many of these data elements must be disaggregated by Federally-defined subgroups, necessitating the collection of student demographic information. A full discussion of Federal NCLB Report Card requirements can be found in non-regulatory guidance issued on September 12, 2003 by the U.S. Department of Education (at http://www.ed.gov/programs/titleiparta/reportcardsguidance.doc). A summary overview table of annual Federal NCLB Report Card reporting requirements for Title I, Part A recipients is below.

Federal Data Requirements for Report Cards for Title I, Part A Recipients Under *No Child Left Behind*	**Level of Reporting**			**Disaggregation Subgroups**								
	State-Level	LEA-Level	School-Level	All Students	Major Racial & Ethnic Groups	Students With Disabilities	Limited English Proficient	Economically Disadvantaged	Migrant[1]	Gender[1]	High Poverty Schools[2]	Low Poverty Schools[2]
Reading and Mathematics Assessment Data[3]												
Percentage of students tested	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔		
Percentage of students achieving at each proficiency level	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔		
Most recent 2-year trend data in student achievement for each subject and grade assessed	✔	✔	✔	✔								
LEA achievement compared to State achievement		✔		✔	✔	✔	✔	✔	✔	✔		
School achievement compared to LEA and State achievement			✔	✔	✔	✔	✔	✔	✔	✔		
Accountability Data												
Comparison between actual achievement and State's annual measurable objectives	✔	✔	✔	✔	✔	✔	✔	✔				
Student achievement on other academic indicators used for AYP (e.g., high school graduation rate)	✔	✔	✔	✔	✔	✔	✔	✔				
Number and names of LEAs and schools identified for improvement, corrective action, and restructuring	✔	✔	✔									
Percentage of schools identified for school improvement, corrective action, or restructuring		✔										
Teacher Quality Data												
Professional qualifications of all public elementary and secondary school teachers (e.g., bachelors and advanced degrees, licensure)	✔	✔	✔									
Percentage of all public elementary and secondary school teachers with emergency or provisional credentials	✔	✔	✔									
Percentage of core academic subject classes not taught by highly qualified teachers	✔	✔	✔								✔	✔

[1]Subgroups of migrant and gender are required subgroups for reporting purposes, but are not among the required subgroups for AYP determinations.
[2]High poverty refers to top quartile and low poverty refers to bottom quartile.
[3]Beginning in 2007-08, science assessment data will be included in this requirement.

Source: *Report Cards: Title I, Part A Non-Regulatory Guidance,* U.S. Department of Education, September 12, 2003, http://www.ed.gov/programs/titleiparta/reportcardsguidance.doc

It is important to note that the flexibility of NCLB allows States to require LEAs and schools to report on additional data elements beyond the Federal requirements, and any State or LEA may choose to report on as many optional data elements as it sees fit.

1.3 Current Issues in Data Quality

NCLB's greatly enhanced focus on data-driven accountability has brought with it a number of challenges for States, LEAs, and schools. Federal NCLB reporting requires that States have the capability to transmit standard statewide information on demographics, achievement, and teacher quality for all public school students and all public school teachers they serve. This can be a daunting task in a data collection and reporting environment often characterized by aging, "stovepiped" systems that may not be able to share data within a single school, much less across schools, LEAs, or an entire State.

The State of the Data

In its September 2004 study on "Improvements Needed in Education's Process for Tracking States' Implementation of Key Provisions," GAO found that "more than half of the state and school district officials we interviewed reported being hampered by poor and unreliable student data." (p. 3)

GAO's report, GAO 04-734, is available on the GAO website (www.gao.gov) at http://www.gao.gov/new.items/d04734.pdf.

Among the key data quality problems associated with NCLB and other reporting are:

- *System non-interoperability*. Data collected in one system are not electronically transmittable to other systems. Re-entering the same data in multiple systems consumes resources and increases the potential for data entry errors.

- *Non-standardized data definitions.* Various data providers use different definitions for the same elements. Passed on to the district or State level, non-comparable data are aggregated inappropriately to produce inaccurate results.

- *Unavailability of data.* Data required do not exist or are not readily accessible. In some cases, data providers may take an approach of "just fill something in" to satisfy distant data collectors, thus creating errors.

- *Inconsistent item response.* Not all data providers report the same data elements. Idiosyncratic reporting of different types of information from different sources creates gaps and errors in macro-level data aggregation.

- *Inconsistency over time.* The same data element is calculated, defined, and/or reported differently from year to year. Longitudinal inconsistency creates the potential for inaccurate analysis of trends over time.

- *Data entry errors.* Inaccurate data are entered into a data collection instrument. Errors in reporting information can occur at any point in the process – from the student's assessment answer sheet to the State's report to the Federal government.

- *Lack of timeliness.* Data are reported too late. Late reporting can jeopardize the completeness of macro-level reporting and the thoroughness of review. Tight NCLB deadlines, for example, can lead to late reporting, poor data quality, and delayed implementation of program improvement efforts. Rushed reporting can often lead to poor data quality, while reporting that is delayed months or even years can often limit data utility.

Sections 2 through 4 of this document will address each of these problems, expanding on the issues that they raise and providing guidelines for overcoming them.

1.4 Document Overview and Organization

This document focuses on the processes and mechanisms of data collection and reporting – not the substance or content of particular data elements. The purpose of these guidelines is to help States, LEAs, and schools establish sound systems conducive to producing accurate, reliable data. The purpose is not to identify the types of data that should be collected or the uses for the data at the local level. These guidelines assume that all States, LEAs, and schools collect the data statutorily required at the Federal level by *No Child Left Behind*. However, as discussed here, the principles involved in building strong systems will be applicable to other data elements as well.

The target audience for the guidelines is two distinct but complementary groups:

- State and local accountability and assessment officers and staff
- State and local Management Information Systems (MIS) and data personnel

A key purpose of this document is to ensure that these two groups can work with and speak to each other using a common language and guided by a common set of understandings. The main body of the guidelines is written in a language and depth designed to be accessible to accountability and assessment professionals, but also credible to MIS professionals and data technicians. Throughout the document, vignettes and insets are used to provide specific technical information to the MIS audience and specific administrative applications and examples to program staff.

Following this introduction, the guidelines are organized into three main sections:

- Establishing a Solid Foundation
- Managing Consistent Data Collection
- Confirming Accurate Results

These sections are intended to track the data collection and reporting process through its basic phases, and capture the categories of management control structures that the U.S. Department of Education's Office of the Inspector General (OIG) identified in its February 2004 Management Information Report: monitoring, receipt and control, scoring, data quality, analysis, and reporting. Each section includes a brief text

introduction, followed by bulleted guidelines based on key distinct data quality issues. Insets containing real-world examples and technical information are also included to supplement the general guidance. Finally, each section finishes with a summary checklist that States and localities can use to gauge the status of their data quality-assurance efforts. These checklists are designed to provide a rundown of the overarching concepts from the guidelines in an easy-to-use format.

The guidelines conclude with a discussion of data quality roles and responsibilities at the State, LEA, and school levels and a list of resources for further information. Finally, the appendix includes reprints of all of the summary checklists found throughout the document, providing users with a convenient compilation of overarching principles.

2. Establishing a Solid Foundation

In February 2002, the U.S. Office of Management and Budget (OMB) published a set of Federal Information Quality Guidelines. These Guidelines, developed in response to a Congressional mandate, established a basic definition of data quality that included three overarching elements: utility, objectivity, and integrity. OMB also directed each Federal agency to develop its own Department-specific standards. The U.S. Department of Education published its Information Quality Guidelines in February 2003. However, as the Department's Inspector General noted, the Guidelines "addressed high level standards …and did not require management controls over scoring of State assessments" and other key NCLB data elements. This section will lay out the basic, underlying processes and systems that set a foundation for quality data.

Key Federal Information Quality Documents

OMB: Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies, February 22, 2002
http://www.whitehouse.gov/omb/fedreg/reproducible2.pdf

U.S. Dept. of Ed.: "U.S. Department of Education Information Quality Guidelines," February 2003
http://www.ed.gov/policy/gen/guid/infoqualguide.html

U.S. Dept. of Ed. OIG: Management Information Report, "Best Practices for Management Controls Over Scoring of the State Assessments Required Under the *No Child Left Behind Act* of 2001," February 3, 2004

GAO: "Improvements Needed in Education's Process for Tracking States' Implementation of Key Provisions," September 2004
http://www.gao.gov/new.items/d04734.pdf

2.1 Overview: Data Quality and No Child Left Behind

Within the confines of this document, the definition of "data quality" encompasses two of the three components of OMB's overarching definition: objectivity and integrity. These guidelines assume that the data elements required by NCLB and by States are, by definition, useful in measuring progress toward predefined Federal and State accountability standards.

The U.S. Department of Education's Information Quality Guidelines describe data "integrity" as the security or protection of information from unauthorized access or revision. "Objectivity" is the presentation of information "in an accurate, clear, complete, and unbiased manner." For statistical data, achieving this standard entails:

- Using clearly defined, broadly understood data definitions;
- Using clearly documented, well thought-out methodologies for data collection;
- Using reliable data sources;
- Processing data in a manner to ensure that data are "cleaned" and edited;
- Properly documenting and storing data collections and results;
- Producing data that can be reproduced or replicated;
- Conducting data collections and releasing data reports in a timely manner; and
- Establishing procedures to correct any identified errors.

2.2 Technical Infrastructure

Automated Systems. Having an adequate technical infrastructure in place is one key element in producing quality data. At a minimum, data collection, processing, and reporting should be automated and transmittable in an electronic format. Even in small States, LEAs, and schools, pen-and-paper systems for managing data will be overwhelmed by the emphasis that accountability systems such as those established under NCLB place on accurate, comprehensive, and timely data reporting. In addition to creating delays and consuming excessive resources, a system that relies on manual or outdated technology exacerbates all of the data quality problems discussed above in Section 1.3. Many of the data quality solutions included in these guidelines are difficult or impossible to implement without an automated data system.

Current Initiatives. The range of technology options available to States, LEAs, and schools in automating data collection processes is vast – from inexpensive desktop spreadsheets to fully integrated State data warehouses linked to every school. Driven largely by NCLB's requirements, numerous ambitious Federal and State initiatives are currently underway to implement comprehensive, state-of-the-art data collection, storage, and reporting networks. These networks are typically being built around a system of unique statewide student identifiers and individual student records, and are potentially capable of delivering real-time educational data to individual teachers at the classroom level. These systems also integrate automated data quality checks.

Interim Processes. Of course, these systems are complicated to develop and take time to complete. However, data, assessment, and accountability professionals at the State and local levels should not postpone steps to improve data quality while they wait for a fully automated, fully integrated statewide data system to be implemented. Several technical infrastructure practices that would improve data quality should be possible under current conditions. The key, regardless of which technology is used, is to establish technical processes that allow data to be checked as they are entered into the system and transmitted to other users.

In the Field: Types of Data Systems

Three different types of automated student data systems exist that promote data quality by State Educational Agencies (SEAs).

- West Virginia and Delaware host student information systems that are used by LEAs and schools on a day-to-day basis. When data are needed for reporting, the SEA can download what is needed from the real-time systems and receive up-to-date, comparable data.
- North Carolina and South Carolina promote data quality by providing the same software to all LEAs. Extracts for reporting purposes can be written once and used by all to promote timely and complete data collections.
- In Texas and Florida, two States that began collecting individual student records many years ago, data standards have been established that make it very clear what is expected to be reported by LEAs and in what format. In Texas, regional service centers check the LEAs' data before the data are submitted to the SEA.

<<<<<<<<<<<<<<< Technical Infrastructure Guidelines >>>>>>>>>>>>>>

General Principles

- *Automation:* All data collection and reporting systems should be automated, and should include automated system backups.

- *Interoperability:* All schools and LEAs should use compatible hardware and software – both within schools and LEAs and with the State system. No "translation" or re-entry of data should be necessary as data are transmitted from providers to collectors. People creating NCLB reports should be able to mine data from existing student information systems and other databases seamlessly, without requiring separate data collections for each NCLB element.

- *Connectivity:* All schools and LEAs in a State should be electronically connected through a network or a common web portal through which all data collection and reporting occurs.

- *Capacity:* Infrastructure established to link interoperable data systems – whether web portals or networks – should have sufficient capacity to accommodate, at a minimum, collection and reporting of all required NCLB data elements by all users at specific times. Infrastructure should also have sufficient capacity to include redundant (backup) data storage.

- *Utility:* The system should be structured around the needs of its users. Processes for gaining access, entering data, generating reports, and transmitting information should be transparent and cause the least possible burden to users.

- *Reliability:* Before they are deemed ready for operation, all data systems should be fully tested. System performance should be monitored on a continuing basis and an IT contingency plan should be in place to ensure the continuity of the system in the case of unforeseen disruptions (such as natural disasters).

Technical Focus: Building Infrastructure

What technical specifications should we build into our infrastructure to produce quality data?

- ✓ File descriptions for data elements, record format, and file design
- ✓ Actual collection instruments and data sources to be used
- ✓ Time period for collection and reporting deadlines
- ✓ Data conversion and processing
- ✓ Storage requirements
- ✓ Preferred formats for output reports
- ✓ Data security and confidentiality checks
- ✓ Controls on the accuracy and completeness of each input, process and output (audit trails, control totals, status flags, and system interrupt/restart procedures)
- ✓ Method for evaluating how well the system is performing, including percentage of forms verified and accuracy rates for coding and key entry

(Adapted from National Center for Education Statistics Cooperative Education Data Collection and Reporting Standards Project Task Force, *Standards for Education Data Collection and Reporting* , 4-4, 4-5, 4-22.)

2.3 Data Definitions

Data Dictionaries. A fundamental piece of any data quality infrastructure is a standardized set of precise data definitions that all providers use. A "data dictionary," which identifies all data elements and describes their content, coding options, and format, is essential to establishing consistent collection and reporting. Adhering to a standard data dictionary improves data quality by fostering interoperability of different reporting systems and promoting the use of comparable data across the entire State. Staff who understand the definitions of the data they are collecting, entering, and reporting will be less likely to commit errors. Data dictionaries can be useful even where systems remain un-integrated and un-connected to a wider network. They should be the foundation for staff training (see Section 2.4) and a resource for staff to use during the data quality review process (see Section 4.2).

Business Rules. A collection and reporting system that is linked directly to a data dictionary can greatly improve data quality as it funnels – or, in some cases, forces – data into a pre-defined configuration. This integration is achieved through the creation of systematic "business rules" that define acceptable values, character formats, and options for handling missing or unavailable data. In the absence of an integrated statewide network, another option is a web portal-based collection system, in which the central portal enforces data dictionary business rules as data are entered.

In the Field: New Hampshire's Data Dictionary

As part of the U.S. Department of Education's Data Quality and Standards Project, New Hampshire has begun to implement the "i.4.see" system, an automated education information database. Working with the Center for Data Quality (C4DQ), New Hampshire has established an online data dictionary that lists the definitions, data rules, and validation requirements for every data element reported.

Schools are the linchpin of the "i.4.see" system. Because schools are the ultimate "owners" of student data, and because they know best when the data are accurate, they are responsible for submitting and revising NCLB data. Automated validation routines, based on customized business rules, allow data to be validated at multiple levels: first when the school submits its data, then when the LEA and State review the information for anomalies and final reporting to the Federal government. A key feature of the system is automatic, real-time feedback on the status of data for every submission. Based on the validation rules in the data dictionary, the system labels each piece of data "rejected" or "accepted" and flags rejected data for correction. Rejected data files are accompanied by error messages that refer automatically to the relevant data validation rules.

For further information on i.4.see, access the NH DOE website at http://www.ed.state.nh.us/education/datacollection/i4see.htm

Data Definitions. In some cases, the U.S. Department of Education (through the National Center for Education Statistics), the U.S. Office of Management and Budget, or the *No Child Left Behind* Act maintains a definition of a required data element. Where Federal definitions do not exist, a standard definition should be used for all LEAs and schools in the State. For example, the U.S. Department of Education allows flexibility among States on the definition and parameters of a "full academic year." Once States define data elements such as these, it is important that the definition be adopted uniformly across all data systems in all LEAs. This information should be maintained in an accountability workbook that is readily

available to staff in schools and districts. Hardware and software should be configured around standard definitions, and the accountability guide should provide a clear description of how data collection, entry, and reporting processes work.

Data Granularity. To the maximum extent possible, all data elements should be collected and stored in their most "granular" form. In other words, each component of a calculated data element should be collected separately and stored separately in the database. For instance, when collecting graduation rate data, it is better to store a total number of students graduating and a total number of students eligible to graduate than to store only a computed percentage. To ensure that data reported by all LEAs and schools are comparable, percentages, ratios and other computed data should not be computed until final calculations are made at the State level. If LEAs are completing forms (rather than sending in individual student or staff records), they should report the component parts of the formula and the SEA should compute the percentages.

<<<<<<<<<<<<<<<<< Data Definitions Guidelines >>>>>>>>>>>>>>>>>

General Principles

- *Unique Identifiers:* To the maximum extent possible, unique statewide identifiers should attach to every student and teacher for whom NCLB data are required.

- *Indivisibility:* Every data element should be defined and collected in as "granular" a format as possible. For example, the data dictionary should separate total days in membership and total days in attendance and indicate how they can be used to compute an attendance rate.

- *Comprehensiveness:* Data dictionaries should include all relevant information for each data element, including its definition, unique code, dates of collection, and technical business rules (e.g., "three-digit number" or "ten non-numerical characters, all caps").

- *Accessibility:* The data dictionary should be easily available to all staff at the State, LEA, and school levels. The dictionary should be posted on-line, available for download into databases and applications, and distributed in hard copy format.

- *Permanence:* Never delete codes or definitions from the data dictionary. Codes or definitions that change or go out of date should be de-activated so that staff will not use them inadvertently, but they are important to maintain so that historical comparisons and longitudinal analysis can occur.

- *Validity:* Business rules should not be the final arbiter of valid data. Data should be checked by a staff member who will know if an anomaly captured by a business rule is, in fact, an error. For instance, business rules may identify counts that are out of range based on previous years' data, but are, in fact, accurate because a significant change has occurred in the reporting unit.

Guidelines for Specific NCLB Data Elements

- *NCLB Demographic Data*
  - All schools and LEAs in the State should use a single standard definition and set of codes for each Federally required NCLB subgroup.
  - If existing State and local subgroup definitions differ from NCLB definitions, the dictionary should clearly identify which description is to be used for Federal NCLB reporting purposes or which groups should be combined for Federal reporting.

- *NCLB Assessment Data*
  - The State data dictionary should include information on links between specific assessments or assessment items and NCLB academic standards.
  - The State accountability guide should include information about performance levels on State standards-based assessments and describe how they relate to the computation of AYP for schools and districts.
  - Where a number of standardized assessments are given in a single school or LEA, the data dictionary and accountability guide should clearly identify which assessment is used for Federal NCLB AYP reporting purposes.

- *NCLB Accountability Data*
  - Because different NCLB elements are required at different reporting levels, data dictionary definitions should distinguish between school-, LEA-, and State-level data elements.
  - The State accountability workbook should contain information on all key accountability elements, including minimum subgroup "N" size, proficiency levels equal to "advanced," "proficient," and "basic," graduation rate calculations, and any other academic indicators used by the State.
  - In cases where an accountability indicator may have a variety of possible definitions, component data for that indicator should be maintained separately in the database. For example, "graduation rate" can be defined differently in different reporting contexts. Rather than storing a single aggregated rate, all potential component pieces of graduation rate should be available, to allow various rates to be calculated as needed.

Technical Focus: Defining Data Elements

What types of information should I include in the my data dictionary?

A typical data dictionary entry from the National Center for Education Statistics (NCES):

- Name: Adequate Yearly Progress Status
- Number: 0028
- Definition: An indication as to whether the education institution meets AYP standards.
- Element Type: ID
- Field Length: 4
- Domains: IEU [Intermediate Educational Unit], LEA, School, SEA
- Related Entities: Accountability Report
- Related Options: 0911 – Does not meet AYP standards, 0910 – Meets AYP standards

The NCES Handbooks Online website contains data dictionary entries for a comprehensive list of U.S. Department of Education data elements. The Handbook can be accessed at http://nces.ed.gov/programs/handbook/index.asp.

2.4 Staff Organization and Training

The Data Quality Team. As important as a solid technical infrastructure and a data dictionary are to producing quality data, it is people who determine whether or not NCLB and other data reporting meets a high standard of accuracy. Automation, interoperability, and connectivity of information technology can provide a framework for producing good data, but such tools are only as powerful as their users make them. While creating staff time for training, implementation, and monitoring of sound collection and reporting practices can pose real challenges (particularly for smaller LEAs), investing in the creation of a data quality team can deliver large returns. Staff involvement at *all* levels – school, LEA, and State - is essential.

In the Field: Empowering Data Stewards

When Virginia's Fairfax County Public School System began its push for improved data quality through the Education Decision Support Library, a key element in its approach was that the consumers of the data would drive the system, rather than the technology staff. Fairfax designated "data stewards" at every school, who assumed ownership over specific data elements. For example, one data steward oversees enrollment data, another monitors course grades, and a third tracks assessments. A school-district-level data workgroup, composed of a group of stewards and district data personnel, meets every month to discuss data issues proactively and do strategic planning.

Crucially, Fairfax County's data stewards were given not only responsibility, but also strong authority to make decisions. While overarching data quality policies and technical systems are developed at the district level to ensure a standard data framework, data stewards work within those policies to develop the business rules, data definitions, and quality check processes that will be used for each element. Data stewards monitor and review all data collected in their domain, and when they identify a data anomaly, they are empowered to resolve it. Data stewards at the school level, who know their data best, have the final say on the "right" number that will be reported to the district and the State.

For further information on Fairfax Co., Virginia's Education Decision Support Library (EDSL), visit http://www.fcps.edu/DIT/edsl/index.html.

School-level Ownership. Ultimately, most of the required data elements in any education data report, including those on a State or LEA annual NCLB report card, "belong" to schools. Schools are where the students and teachers actually perform the activities that are being measured, and school-level personnel are the initial input point for much of the most important student outcome information. Because the most effective method of improving data quality is to prevent errors from occurring, staff organization and training at the school level are critical to producing reliable reports at the LEA, State, and Federal levels. School staff have a strong interest in producing accurate data, and should be given the responsibility – and time – for developing a proprietary interest in maintaining the quality of data they collect, report, and use. Through regular oversight, engagement, and feedback, LEAs and States can train school-level staff in the relevance of not only the micro-level student data with which they are most familiar, but also of macro-level information. Bad data at the school level that result in an erroneous NCLB report card can be significant for a school, LEA, or State – not just in terms of resources, but also in terms of prestige, morale, and a host of other effects.

State Educational Agency Leadership. The SEA plays an essential role in ensuring that data obtained from schools and LEAs will meet reporting requirements – both at the State and Federal level. Since nearly all required data originate in schools and districts, the SEA must provide leadership and guidance to ensure the highest quality data possible. SEAs must develop data systems that ensure all LEAs and schools can report data in a timely manner and with the least amount of burden, while giving the SEA the information and flexibility to meet State and Federal reporting requirements. Data systems in many States and LEAs are undergoing changes and enhancements, in part in response to Federal data needs, but also because there is a greater perceived need for useful and timely data for decision-making. This magnifies the importance to SEAs of developing the technical and operational skills of data personnel at the LEA and school levels. Because a shared understanding of key data systems and procedures is a major factor in creating effective feedback loops among SEAs, LEAs, and schools, improving these skills can improve efficiency and reduce tensions during the collection and review process.

This document contains staff development guidance for practices that are already in place in some organizations, but not in others. It is hoped that all States and LEAs may find these guidelines of use to evaluate the status of existing systems and plan for improving the procedures and systems used to collect data for both decision-making and reporting. The guidelines that follow focus on general principles for organizing and training staff to facilitate what the National Forum on Education Statistics calls "a culture of high quality data." That culture should pervade all levels of the data organization – from schools to the State Departments of Education and the U.S. Department of Education.

In the Field:
Meeting the Data Quality Challenge in a Small LEA

The process of training and organizing staff to establish an efficient, effective data quality team can be a daunting task for any LEA or school. For small and/or rural LEAs with limited resources and administrative staff of only a few people, the challenge is magnified. As a result, some small schools and LEAs take a "we just can't do it" approach, relying on vendors, a State data office, or other outside agencies to process and validate their data. This hands-off method carries grave risks: potentially erroneous information, which could have serious consequences for funding and accountability, may not be caught by distant data personnel who are not familiar with the particular school in question. While the general guidelines in this document do not fully address the unique set of challenges that many small LEAs face, useful examples exist for implementing critical building-level data review despite limited staff and resources.

In Charles County, Maryland, for instance, data collection and processing are administered primarily from the LEA data office – with key data review and validation functions carried out by a designated data steward at the building level. The LEA, with a staff of two to three people working on data part-time, is responsible for assembling all of its data in its data warehouse. As data files are created for information such as student demographics or assessment results, district staff transmit individual school data to a test coordinator in each building for review and verification. Errors or other data questions are filtered back to the LEA level before the data are finalized for transmittal to the State and the Federal government. At least once per year, the LEA data staff provide training in data quality procedures for all test coordinators. Test coordinators, who may be principals, teachers, or other administrative staff, receive a small stipend for this extra duty – which generally occurs over a 24-hour turnaround period twice per year.

<<<<<<<<<<< Staff Organization and Training Guidelines >>>>>>>>>>>

General Principles

- *Organization:* Designate dedicated staff at the State, LEA, and school levels with specific responsibility for and authority over monitoring data quality.
  - **State level:** establish a chief information officer, a data quality office, and a data policy advisory committee.
  - **LEA level:** establish a chief information officer, a data quality coordinator, and an NCLB data quality specialist. In small LEAs with few students and a small administrative staff, all functions might be performed by a single person or shared among all staff.
  - **School level:** establish a data quality oversight team including data stewards responsible for each key data element. In small schools, the "team" might be a single staff member responsible for reviewing and verifying data collected at the LEA or State level and serving as a conduit for data quality information among the school, the LEA, and the State.

- *Communication and Feedback:* Establish clear and regular communication channels among data providers and data collectors at all levels – State, LEA, and school. When an LEA identifies erroneous or questionable data, for example, the school-level owner of the data should be alerted and given the opportunity to make corrections.

- *Ownership:* Create incentives for those who are closest to the data collection to take a proprietary interest in its quality. School staff, for example, should feel ownership of the data they collect on student demographics, program participation, enrollment, and achievement. Incentives should be both cautionary (e.g., the serious consequences of being erroneously identified as not making AYP) and motivational (e.g., rewards such as extra funding, higher accreditation status, or other awards for schools that do a good job of ensuring their own data quality).

- *Inclusion:* Include information technology (IT) staff, policy staff, and teachers in the data quality process, along with data stewards, report preparers, data entry personnel, and all data owners. If vendors play a significant continuing role in data collection, include their representatives in all phases of the process as well.

- *Coordination:* Ensure that regular communication, consultation, and cross-training occurs across all data systems personnel at the SEA, LEA, and school levels. Where multiple IT systems operate in a stove-piped fashion with differing business rules and separate personnel (e.g., separate financial, assessment, and special education databases), it is critical that staff operate under a standard set of data definitions, data review processes, and data validation rules. Ideally, coordination should lead to an integration of all stove-piped processes into a single, unified system.

- *Education:* In addition to training in data entry methods and procedures, educate staff on the larger context of the data collections. Where do NCLB and other data originate? Where do they end up? Why are they being collected? Staff who understand the purposes for a data collection and the possible consequences of errors are less likely to "just fill something in" to satisfy the government.

- *Technical Training:* Hold regular standard training sessions for all personnel involved in the NCLB reporting process. These sessions should describe and demonstrate the procedures for NCLB data collection, entry, and reporting.

- *Documentation:* Prepare a State-level data quality handbook, including information on coding, data entry, sample forms, and the larger context of reporting. The handbook should be available on-line, and a summary checklist should be posted prominently wherever data entry takes place.

- *Ongoing Assistance:* Establish a data quality help desk at the LEA or State level, or designate a "go to" person to be available to answer data questions from the field. Having a convenient, dependable resource for authoritative answers can be the difference between "I'll just fill something in" and getting the data right.

<u>Guidelines for Specific NCLB Data Elements</u>

- *NCLB Demographic Data*
  - Designate a single data steward in each school who is responsible for ensuring that data are entered according to standard definitions.
  - Train all school and LEA data staff on the *Federal* definitions of each of the required NCLB subgroup categories.
  - Train all data staff in the relationship between NCLB subgroup classifications and AYP determinations.
  - Disseminate a list of "translations" between Federal NCLB demographic definitions and State-, LEA-, and school-level demographic definitions.

- *NCLB Assessment Data*
  - Designate a single data steward in each school who is responsible for monitoring the correctness of identity information on the assessment forms.
  - Train all school and LEA data staff in the content and purpose of the NCLB assessment, and explain the difference between the assessment used to determine NCLB AYP and other Federal, State, and local assessments.
  - Train teachers, assessment proctors, and assessment scorers in the specific scoring procedures related to the State's or LEA's NCLB AYP assessment.

- *NCLB Accountability Data*
  - Designate a single data steward in each school who is responsible for ensuring that the data submitted for accountability purposes is correct.

- Train all data staff in NCLB accountability measures, including definitions of "advanced," "proficient," and "basic" in the State accountability plan.
- Train all data staff in the "other indicators" being used for NCLB, including the State's definitions of elements such as graduation rate and dropout rate.
- Educate all data staff in the uses of NCLB accountability data and the potential consequences of errors in reporting results.

- *NCLB Teacher Quality Data*
  - Designate a single data steward in each school who is responsible for ensuring that teacher assignment information is correctly submitted.
  - Train all data staff in the reporting requirements for NCLB teacher quality data, including the State's definition of "fully licensed/certified."

Technical Focus: Training Staff

What should I include in a comprehensive session to train all data staff on procedures for producing high quality data?

- ✓ Discussion of the goals and objectives for good data quality
- ✓ Description of data quality policies and definitions
- ✓ Identification of key personnel involved in the collection/reporting process
- ✓ Dates, times, and durations of collection activities
- ✓ Required procedures for administering specific collections
- ✓ Limits on acceptable deviation from specified procedures
- ✓ Ethical and legal responsibilities related to security/privacy
- ✓ Hands-on practice with the data entry system and the data collection instrument to be used (administering the test or transcribing the record)
- ✓ Samples of reports that will be produced
- ✓ The help desk number, or whom to call with a question during collection

(Adapted from National Center for Education Statistics Cooperative Education Data Collection and Reporting Standards Project Task Force, *Standards for Education Data Collection and Reporting,* 3-5, 3-6, 4-13.)

Summary Checklist: Establishing a Solid Foundation

Does your data infrastructure have the following characteristics?

- Data collection, processing, and reporting systems are ***automated*** and data can be transmitted in an electronic, interoperable format.

- Immediate ***interim processes*** for improving data quality are in place, as larger systemic initiatives are implemented.

- A ***data dictionary*** identifies all data elements used in collection and reporting, and describes their content and format.

- Systematic ***business rules*** define acceptable values, character formats, and options for handling missing or unavailable data.

- Hardware and software, along with staff training, are configured around standard ***data definitions*** and business rules.

- ***Data granularity*** is preserved by collecting data in their most basic forms and computing percentages, ratios, and other calculations at the State level.

- A ***data quality team*** has been established at the school and LEA levels to craft and implement data quality goals and procedures.

- Data collection, reporting, and review encourages ***school-level ownership*** of basic information, and staff training reflects this responsibility.

- LEA- and State-level personnel play leading roles in building a ***culture of data quality*** among staff at all levels.

3. Managing Consistent Collection

As Section 2 showed, preventing data errors before they occur is one crucial component to improving data quality. A solid technical infrastructure, a comprehensive data dictionary, clear documentation of accountability measures, and an organizational culture of data quality can greatly reduce the potential for data problems. However, there are also a number of steps that States, LEAs and schools can take during data collection to improve the quality of the results reported. In cases where basic infrastructure is inadequate or in the process of being upgraded, these management controls can help prevent inaccurate data from entering the system and being reported to subsequent levels. This section provides guidelines on designing sound collection instruments, collecting assessment data, and creating effective collection timelines.

3.1 Overview: Data Collection Processes

Mining Existing Data. Many of the data required for State- and LEA-level reporting and for the annual NCLB report cards are available from existing data sources (such as individual student record systems), and do not require a separate "collection." Rather, these data can be "mined" by examining current databases and extracting the relevant information for a new purpose. Demographic data, for example, should exist in most student information systems continuously, and can be updated on an ongoing basis without repeated collections "from scratch" for NCLB Report Cards and other reports. This creates a more stable bank of data that will be less prone to data entry errors and higher in quality over time. Attendance rates, graduation rates, teacher credential information, teachers' course assignments, and prior years' accountability data should also be maintained in databases on a continuous basis. Although these data are mined rather than collected, an ongoing data quality monitoring process will help to ensure their accuracy. In determining how many years worth of data will be maintained for a given data element, a key consideration is context: what timeframe will be useful for end-users of the data? For example, data on individual teachers' credentials may be useful for decades, but data on their course assignments might become obsolete after a few years. Statutory requirements at the State and Federal level also affect decisions on how many past years of data are kept. Federal Title I NCLB Report Cards, for instance, require that States, LEAs, and schools report 2-year trend data for accountability indicators.

Student Assessment Data Issues. Current year assessment data collection presents a special set of data quality issues. Distinct from most other data collection, this process involves the initial collection of new pieces of data. Data are transferred from a student answer sheet to a database and then aggregated for reporting to the LEA, the State, and the U.S. Department of Education. In some cases, students enter their assessment answers into an online assessment system directly. In others, students fill out paper-and-pencil answer sheets that are manually scored with scanners or by other means. Scoring is usually done by an assessment vendor hired by the SEA, which submits assessment results to the SEA, LEAs, and schools in electronic and/or paper formats. In a few instances, assessments are scored at the local level by local or regional staff, and results are provided to the SEA.

3.2 Collection Instruments

Instrument Design. Managing a consistent data collection process begins with well-designed collection instruments. Poorly-prepared collection tools can put data quality at risk before a single piece of data is collected or entered into a database. In the case of NCLB Report Card data, there are a number of different types of collection instruments that might be used. The most common is a basic data entry form, either paper-based or electronic. The form requests certain pieces of information, and the data provider enters the information. Designing data collection forms should not be taken lightly: the layout, labeling, or instructions (or lack thereof) on an instrument can produce most of the NCLB data quality problems listed in Section 1.3, from inconsistent item responses to data entry errors to lack of timeliness.

In the Field: Rhode Island's Data Portal

The Rhode Island Department of Elementary and Secondary Education has implemented a one-stop data collection, entry, and reporting portal. Accessible by password through the Department's public website, the portal contains not only data submission and reporting forms, but also a fully functional data dictionary, data validation rules, and State data policies. School and LEA NCLB report cards are generated from data collected via the website, including results on Rhode Island's statewide NCLB assessments.

Because Rhode Island's portal-based collection system links basic information in standard formats, it creates interoperability among separate data categories. All of Rhode Island's LEA data systems must link to the portal and must use the standard reporting formats it establishes. For example, by tapping into the student information database, the system can pre-label student assessment forms with the names and demographic information for all students taking NCLB assessments. Students thus do not report their own demographic data, greatly reducing the potential for error in reporting on subgroups – a critical NCLB data element.

Assessment Instruments. It is important to note that a student's assessment answer sheet is often a data collection tool in itself. Assessment systems in which students fill in "bubble sheets" or enter answers into an electronic database are, essentially, asking students to enter their own "achievement indicators" into the system. This makes the clarity of the answer sheet's instructions critical. Students' lack of knowledge of key data definitions – particularly in an area such as demographic information – can seriously damage data accuracy. Students unaware of the definition of "Native American," to use one actual example, might self-report in that ethnic/racial category because they were, indeed, born in the U.S. Whenever direct student data entry can be avoided, alternate collection tools should be used. For example, student identifiers – including name, ID number, subgroup category – can be pre-coded from student information systems onto answer sheets, expediting the reporting process and improving accuracy. To ensure confidentiality and avoid introducing extraneous factors into the administration of the assessment, this pre-coding should be accomplished by means of a confidential, unique identifier – not by means of information printed on the answer sheet in a way that would be readable by the test-taking student or others.

Working with Vendors. Data quality issues make it especially important that States, LEAs, and schools work closely with vendors in preparing and producing all data collection tools. States in particular need to work with their assessment and scoring contractors to ensure that assessment items align directly with the academic content

standards established for NCLB, scoring is done correctly, and student scores are correctly reported. Whenever possible, SEAs should require or facilitate pre-coding of answer sheets by assessment contractors. In addition, vendors should provide – or assist in providing – training to staff responsible for administering the data collection and completing any necessary data entry, particularly for students using answer sheets that are not pre-coded. Other time-saving procedures should be negotiated to decrease the time required for the vendor to provide assessment results to the SEA.

<<<<<<<<<<<<<<< Collection Instruments Guidelines >>>>>>>>>>>>>>>

General Principles

- *Relevance:* Collection instruments should request only information directly appropriate to the specific reporting requirement or data need under consideration. Before converting a paper form to electronic format, do a data audit to determine what is necessary and what is outdated or redundant. The key question is "what do we need, both now and in the future?"

- *Uniqueness:* Collection instruments should request only information that is not available from any other existing source. Student demographic information, for example, is typically available permanently in student information systems.

- *Clarity:* Instructions for collection instruments should state clearly and directly – in jargon-free language – what is expected for each data entry. Data definitions should be provided for each instance of data input, and procedures for recording missing or null data ("N/A's" and "0's") should be clear. "N/A," for example, can represent "Not Applicable" or "Not Available" – inputs with very different meanings. Instructions should specify which is applicable. In addition, instructions should make clear which elements are required (i.e., do not allow a non-response), and electronic forms should prevent leaving such elements blank.

- *Validity:* Business rules (see Section 2.3) identifying expected responses should be built into all data collection forms, either automatically (electronically) or in the instructions to the form. Forms should be designed to discourage non-responses, and only one discrete data element should be requested at a time. Required elements (those where "N/A" or other non-responses are not acceptable) should be flagged for validity checks if the response rate is less than 100 percent.

- *Ease of Use:* Make the collection process as easy as possible for the provider. Where possible, the collection tool should track with the format and sequence in which data staff typically maintain their records. Instruments should describe the mechanics of entering each data element (e.g., enter an "X" or a " ✔ ").

Guidelines for Specific NCLB Data Elements

- *NCLB Demographic Data*

- Wherever possible, use existing data sources to collect information on students' subgroup affiliation. New collection instruments should not be necessary if existing instruments can be automatically mined for these data.
- Ensure that data in the student demographic information system are current prior to assessment. Pay particular attention to items such as poverty status or limited English proficiency status, which may change from year to year.
- If separate definitions exist for identifying subgroups for Federal-, State-, or LEA-level reports, ensure that collection instruments state clearly that NCLB reports should use the definition for Federal reporting. In addition, ensure that the Federal definition is available to data entry personnel.
- Do not ask students to self-report NCLB demographic data. Use information mined from the student information system for this purpose.

- *NCLB Assessment Data*
  - Wherever possible, avoid using assessment instruments to collect non-assessment data, particularly when students are entering data directly to the collection system (e.g., through a scanned form or electronic answer sheet).
  - Ensure that assessment instruments clearly associate each item with one State academic content standard.
  - Ensure that assessment instruments are capable of being measured objectively. Data quality is more easily assessed for multiple choice, directed-response instruments than for constructed-response or essay instruments.
  - Determine all acceptable responses to each closed-ended item and assign numerical values or codes to each response. Also assign codes or values for blanks or missing data.
  - Field test all assessment instruments before they are used "live," including a review of reliability of answer keys, scoring rubrics, and scoring equipment.

- *NCLB Accountability Data*
  - Wherever possible, use existing data sources to collect information on accountability indicators such as prior year assessment scores, graduation rates, and attendance. New collection instruments should not be necessary if existing data systems can be mined for these data.
  - Align accountability data collection instruments with the State's complete list of academic indicators, including optional or "other" NCLB elements.

- *NCLB Teacher Quality Data*
  - Wherever possible, use existing data sources to collect information on teachers' credentials, qualifications, and assignments. New collections are unnecessary if existing data systems can be mined for these data.
  - Administrative documentation of teacher qualifications for NCLB purposes should include clear and detailed definitions of "fully licensed/certified," "emergency/ provisional certification," "core academic subject," and other NCLB terms.

Technical Focus: Preparing for Assessments

What can I do before testing to help ensure that the assessment data we report for NCLB accurately reflect our students' actual achievement against State standards?

- ✓ Ensure that test items and test keys align with State academic content standards.
- ✓ Create a "test map" documenting each item's correct response, score points, related standard, and other key attributes.
- ✓ Test scanning or other electronic scoring equipment to verify accuracy.
- ✓ For hand-scored assessments, provide pre-test training in scoring procedures.
- ✓ Train test proctors on assessment procedures.
- ✓ For scaled-score assessments, verify the accuracy of raw score/scaled score conversion tables.

(Source: Council of Chief State School Officers. *Quality Control Checklist for Processing, Scoring, and Reporting.* Technical Issues in Large-Scale Assessment – A State Collaborative on Assessment and Student Standards, January 2003.)

3.3 Student Assessment Data Collection

NCLB assessment programs are generally run by the SEA, and statewide assessments are generally processed and scored by contracted specialists. However, schools and LEAs also play critical roles in ensuring the quality of assessment data that are reported. To avoid coding errors and time-consuming data entry, LEAs should work with SEAs and their contractors to pre-code answer sheets with identifying information and demographics. As noted in Section 3.2, such pre-coding should use confidential, unique identifiers rather than information that would be readable by test-taking students or others. If pre-coding is not possible, teachers and counselors should take responsibility for entering accurate demographic data onto answer sheets. To avoid errors, students should not be involved in entering this information onto answer sheets.

In the Field: Big Sky StAR

Great Falls Public Schools, a system of more than 10,000 students in northwestern Montana, has implemented a fully automated student assessment record system - StAR. The system, developed through a contract with a commercial vendor, allows teachers and administrators to access test results for dozens of standardized assessments given in their LEA. Data are uploaded to the system from a variety of sources. For example, student demographic data are imported from Great Falls' student information system, while assessment data are uploaded to a website from CD-ROMs sent by the district to the vendor.

A key feature of the database is its automatic cross-reference links among assessments, answer keys, and Montana's academic content standards. For any student – or any subset of students, including NCLB subgroups – the system can display results for specific test items and the particular State standards associated with those test items. This provides an accurate and reliable storehouse of assessment information to be used for reporting on Federal NCLB requirements, as well as for instruction at the classroom level targeted to individual students' needs.

Local staff and assessment proctors should monitor students as they complete assessments to ensure that students make clear entries on the answer sheets and erase stray marks and changes. In addition, data stewards and other school assessment staff should perform checks proactively on a representative sample of assessment instruments and answer sheets before shipping them to the vendor for scoring. This will promote more accurate scoring by identifying potential problems early in the process, and will reinforce the critical principle that it is the school – not the vendor – that ultimately "owns" the information to be reported.

Working with Vendors on Assessment Quality Control: The TILSA SCASS Checklist

Whether or not a vendor has been contracted to process standardized assessments, schools, LEAs, and States are ultimately responsible for ensuring accurate reporting of results. The Council of Chief State School Officers (CCSSO) has developed a detailed checklist to assist school, district, and State staff in working with vendors to implement an assessment quality control system. The *Quality Control Checklist for Processing, Scoring, and Reporting* is available from the CCSSO's State Collaborative on Assessment and Student Standards, Technical Issues in Large-Scale Assessment website at http://www.ccsso.org/content/pdfs/scorereportQCchklst.pdf. The Checklist contains 33 items in 8 categories, including descriptions of tasks, recommended expertise, and recommended materials.

<<<<<<<<< Student Assessment Data Collection Guidelines >>>>>>>>>

Guidelines for Specific NCLB Data Elements

- *NCLB Demographic Data*
  - Do not use student assessment answer sheets to collect demographic information.
  - Collect NCLB demographic data as a part of the regular record-keeping in the school and LEA. Update it as needed, especially prior to the assessment period. Ensure that the database containing NCLB demographic data can be matched to the scores for the students using a unique identifier.

- *NCLB Assessment Data*
  - With electronic scoring systems, verify the accuracy of the scoring mechanism before beginning data entry, using a sample of actual student answer sheets.
  - Test administrators – including third-party vendors, where applicable – should be able to track the status of all assessments during the scoring process. Both physical location and stage in the scoring/reporting process should be available for each individual assessment answer sheet.
  - At the end of the scoring process, re-verify the accuracy of the scoring mechanism, using a sample of actual student answer sheets and a separate process.
  - For constructed response items, ensure that raters' scores are consistent both across different raters and within the same rater over time. This should occur

during the scoring process – not as a later check after the process has finished.

- Check individual responses and total scores for inconsistencies, out-of-range errors (e.g., a score of "110" where the maximum possible is 100), and missing responses.
- When raw scores are converted to scale scores, verify manually that an actual sample of student raw scores has been converted accurately into scale scores.
- An answer sheet or, at a minimum, a database entry should be prepared for every student that does not take an assessment, so that all students are included in the final reporting.

- *NCLB Accountability Data*
  - Maintain prior years' assessment results permanently to be able to report longitudinal trends and comparisons for NCLB accountability purposes.
  - To track participation rates, maintain records on students that do not take assessments.

Technical Focus: Scoring Assessments

How can I prevent reporting errors by monitoring the location, status, and scoring accuracy of student assessments in real time?

The U.S. Department of Education's OIG reports that contractors working on scoring the National Assessment of Educational Progress implemented the following controls:

- ✓ Account for all test booklets using bar codes and scanners.
- ✓ For reports that must be shipped for scoring, use the shipping company's tracking system.
- ✓ Establish sample "calibration" questions to set the appropriate scoring parameters for all scorers.
- ✓ Periodically perform calibration scoring to ensure questions are scored consistently across scorers. Scorers re-calibrated after any break longer than 15 minutes.
- ✓ Have supervisors "backread" scored assessments periodically to ensure individual scorers perform consistently over time. Approximately 10 percent of each scorer's work is backread during the course of scoring.
- ✓ Have backup scorers periodically re-score a sample of questions already scored to check on interrater reliability (done daily and reviewed with scorers in real time).
- ✓ Use trend scoring to compare this year's scores with previous years' scores and identify significant inconsistencies.
- ✓ Prepare frequency distributions to show ranges of typical scores, and monitor how the actual distribution compares in real time as scoring progresses.

(U.S. Department of Education, Office of the Inspector General. Final Audit Report, *Review of Management Controls Over Scoring of the National Assessment of Educational Progress 2000,* Report #ED-OIG/A05-C0010, June 2003.)

3.4 Schedules and Deadlines

Firm, Clear Schedules. Producing quality data takes time. Hurried or *ad hoc* collection and reporting, done on the spur of the moment in response to an unexpected request, greatly increases the potential for virtually all of the data quality problems listed in Section 1.3. Under time pressure, data entry is more likely to contain errors, information is more likely to be reported as unavailable or missing, the "just fill something in" temptation is more likely to arise, and thorough review and validation of data is less likely to occur. A sound infrastructure (as described in Section 2) and well-designed collection instruments can mitigate many of these problems. However, even the best data system requires a clearly established, firmly enforced collection, validation, and reporting schedule to ensure good data quality.

Federal NCLB Reporting Timelines

For many States, the Federal reporting requirements in NCLB call for data submissions on a timetable that is more accelerated than they are accustomed to meeting. To meet the needs of students and parents who may have the option to exercise public school choice or take advantage of supplemental educational services, the NCLB statute mandates that AYP determinations "shall take place before the beginning of the school year following such failure to make adequate yearly progress." (Sec. 1116(b)(1)(B)) In most States, this means that the results of State NCLB assessments administered in the spring must be collected, entered, validated, and available to school districts by the end of the summer - a span of three to four months.

A Continuous, Inclusive Process. The State data quality schedule should include a continuous management process that updates all elements of the data system on a regular basis. In addition to collection, validation, and reporting processes, schedules should include regular updates to the data dictionary, regular technical system performance tests, regular staff training, and regular data quality policy reviews. All schedules should take into account both policy *and* technical considerations. Neither MIS staff nor policy staff should dictate deadlines alone. Both should be involved in ensuring that timelines are workable and meet Federal and State requirements.

Statewide Deadlines. The guidelines that follow provide information on implementing a statewide data reporting schedule that will meet Federal timelines and still maintain safeguards for data quality. It is critical that these efforts be directed from the State level, because the SEA is ultimately responsible for ensuring that Federal reporting timelines are met. A key overarching principle is to prepare ahead. While it is true that the turnaround time for assessment results may be tight because of testing schedules, most other data required for NCLB Report Cards should be available earlier in the year. Teacher quality data, subgroup demographics, past years' accountability information, graduation rates, and attendance rates, for example, should not need to wait until the last moment for collection and validation. The NCLB timelines are transparent in the statute, and States – in consultation with data providers at the school and LEA levels – should build their data reporting schedules with those timelines in mind.

<<<<<<<<<<<<<<< Schedules and Deadlines Guidelines >>>>>>>>>>>>>>>

General Principles

- *Standardization*. Standard statewide reporting deadlines should be established, based on Federal NCLB timelines. LEA- and school-level deadlines should be set at the local level to meet the State schedule.

- *Separation.* Deadlines should be set separately and distinctly for collecting, validating, and reporting each required data element. Wherever possible, timelines for different elements should be staggered to avoid overwhelming data collection and validation processes at any single point in the calendar.

- *Feasibility.* Schedules and deadlines should take into account the technical capabilities of all data providers. While technical issues should not dictate timelines for reporting, local systems must be physically capable of meeting collection and validation schedules set at the State level.

- *Follow-up Capability.* State-, LEA-, and school-level schedules should build in substantial time for following up with data providers on data anomalies, missing items, and other data quality issues. Reporting to the next level should not occur until all anomalies have been resolved.

- *Transparency.* All schedules and deadlines should be set in consultation with key personnel responsible for providing data and validating data quality. Final timelines should be disseminated well in advance of deadlines and periodic reminders should be relayed to key data staff.

- *Firmness.* State deadlines should be firm and *include consequences* for non-compliance. Specific procedures should be established for permitting and processing late data reporting.

<u>Guidelines for Specific NCLB Data Elements</u>

- *NCLB Demographic Data*
  - Collect demographic data prior to conducting the State assessment so that it will include information on all students in the school district. Combine the demographic data database with assessment data to produce subgroup results after assessment data have been collected and reviewed, in a separate process.

- *NCLB Assessment Data*
  - Reserve extensive time in the schedule exclusively for assessment data collection and validation, as soon as possible after the assessments are administered. Detailed timelines at all levels – school, LEA, and State – are important to ensure timely reporting according to Federal requirements.

- *NCLB Accountability Data*
  - Monitor, validate, and record attendance, dropout, and course completion data continuously throughout the year so that data can be submitted quickly at year's end without having to locate missing students.

- When possible, allow State data systems to "pull" past years' accountability data directly from LEA or school systems for Federal reports. As long as past years' data were properly validated before they were first reported, this process saves time and promotes consistent reporting across years.

- *NCLB Teacher Quality Data*
  - Collect teacher assignment data twice during the school year and compare these data to the certification/licensure records for employed teachers. Collect all of each teacher's assignments; do not just collect the primary assignment, since the teacher may be teaching several different courses.
  - Monitor data on numbers of non-highly qualified teachers teaching core subjects throughout the year. Backtracking at the end of the year to collect these data will be time-consuming and prone to errors, inconsistencies, missing data, and inaccuracies.

Technical Focus: Scheduling Data Collection and Validation

How would I put together a data collection schedule at the State level?

One example of a State data collection reporting schedule appears below.

Massachusetts Department of Education 2004-2005 Data Collection Reporting Schedule

Form	What does it collect?	Level of Data	Data as of	Due Date	Collection Method
SIMS* October 1 Collection	Individual student data	Student	Oct. 1	Nov. 4	File Transmission
District and School Staffing Report	District and School staff by certification area, with licensure and highly-qualified information	District/School	Oct. 1	Nov. 1	On-line Form
Individual Private School Report	Student enrollment data for private schools	School	Oct. 1	Nov. 1	Paper
Tech Plan Update	District Technology Plan	District/School	Oct. 1	Nov. 19	On-line Form
SIMS Special Education Federal Child Count	Individual Student Data (all 52 data elements for special education students only)	Student	Dec. 1	Jan. 6	File Transmission
School Attending Children	Number of school age children by municipality attending public and private schools, or home-schooled	City/Town	Jan.1	March 1	On-line Form
SIMS March 1 collection	Individual student data	Student	Mar. 1	Mar. 28	File Transmission
Title 1 Performance & Achievement Report	Title I programming information	School/District	school year	May 15	Paper
SIMS End of Year Collection	Individual student data	Student	End of Year	July 14	File Transmission
School Safety and Discipline Report (SSDR)	Individual data for each exclusion or disciplinary action resulting from a violent act, drug or weapon offense.	Incident	school year	July 30	File Transmission

(Source: Massachusetts Dept. of Education website - http://www.doe.mass.edu/infoservices/data/schedule.html#)
*Note that SIMS is Massachusetts' abbreviation for its Student Information Management System.

Summary Checklist: Managing Consistent Collection

Does your data collection process have the following characteristics?

- Information collection for most data elements is achieved through ***mining existing data***.

- Layout, labeling, and instructions for data forms and systems contribute to ***clear, straightforward collection instruments***.

- Assessment instruments ***avoid direct student entry*** of information available from other sources.

- States and LEAs ***work with vendors*** to ensure that assessments and other data collection instruments align with definitions and standards.

- The State has established ***firm, clear schedules*** and deadlines for collecting, validating, and reporting required data elements.

- Data quality validation is a ***continuous, inclusive process*** that updates all elements of the data system on a regular basis and takes into account both policy *and* technical considerations.

- ***Statewide deadlines*** spread collection and reporting responsibilities over the full school year.

4. Confirming Accurate Results

An ongoing system of checks on data accuracy, along with mechanisms for mid-course corrections of information already in databases, allows States, LEAs, and schools to catch errors that the basic infrastructure (see Section 2) or data collection process (see Section 3) may have missed. Despite the best efforts of instrument designers, data quality teams, and vendors, data errors and inaccuracies will inevitably occur. Even well-designed instruments and well-trained, conscientious staff can produce data errors if post-collection management controls do not contain thorough data quality validation checks and strong security and privacy safeguards.

4.1 Overview: Data Quality Management Controls

Ongoing Validation. The data quality process does not end with a successful data collection. Having an ongoing set of management controls over data gathering is important because, as was discussed in Section 3, much data "collection" is in fact a process of pulling information from existing databases ("mining"). In most cases, information on key elements such as student demographics, teacher qualifications, and past years' accountability results has already gone through an initial collection and is simply waiting to be mined. (Collecting current-year student assessment data is an important exception – see Section 3.3.) For these data, a set of business rules and validation checks can help ensure that the final report does not reproduce and transmit errors that occurred during the original collection. Wherever possible, ongoing data quality checks should be automated, performed on a regular schedule, and linked automatically to the data dictionary. In most established data systems, for example, regular quality control sampling seeks 100 percent verification of a sample of records from either the entire system or from a specific system level. Such procedures can be effective in finding correctable errors.

Data Quality Validation Flags

Non-correspondence. Some members of the population for which data are collected are not in the corresponding database, or some members represented in the database have no (or incomplete) corresponding data.

Invalid value. The value entered is not possible, given the data element's definition. For example, "-1" as an assessment score or "?R" as a teacher's length of service.

Invalid code. The code entered does not exist in the data dictionary.

Out-of-range. The value entered, while theoretically possible, is outside the expected range of responses (for example, a dropout rate of 100% or a teacher salary of $1 million). Out-of-range errors should rely on analysis of historical trends and should generate system flags rather than outright error messages.

Embedded Security Safeguards. Good privacy and security practices should be embedded in every phase of the data quality process. From creating a technical infrastructure to training staff and creating collection and validation instruments, security considerations must be incorporated into every decision that policymakers and MIS professionals make.

4.2 Data Review and Validation

Formatting Checks. Data review and validation occurs in a number of different forms and at a number of different levels throughout the data collection and reporting process. During the initial collection and entry phase, as noted in Section 2.4, most data are "owned" at the school level. Graduation and attendance counts and student demographics are likely to originate with building staff. Initially, validation consists of automated data checks that ensure that data entered into the database for the first time are in the proper format: they conform to the limits on field size, character type, and value restrictions built into the data dictionary. In systems that are not automated, this stage requires great attention to detail and, preferably, at least one manual back-up reader to confirm.

In the Field: "Show Me the Data"

Working with the Center for Data Quality (C4DQ), the state of Missouri undertook an initiative to reduce the number of errors in its educational data. Among the problems that Missouri experienced were bad AYP subgroup data (caused by students' inconsistency in self-reporting on their assessment answer sheets) and incorrect figures for average staff salaries and numbers of enrolled students.

The key to Missouri's reform effort is an automated system that reduces both the complexity and the time required to ensure that accurate data reaches the State and Federal governments. The web-based, interoperable system developed with C4DQ contains a set of business rules – set by the state's data policy team – that check all data against an acceptable set of possible responses. Data are cleaned for range errors (such as the staff members whose salaries were entered as $1 million), invalid codes (such as using P4 to refer to four-year-olds instead of PK), and other formatting anomalies.

For more from C4DQ on Missouri's data quality initiative, see c4dq.com/customers/cp-missouri.asp

Cleaning the Aggregate. After initial collection and entry, data must be cleaned to flag out-of-range errors, non-correspondence problems (missing or extra data), and aggregation and rounding errors. This level of validation requires more sophisticated analysis, and should be conducted by the data steward responsible for the data at the school level. Before data are transmitted from the school to the LEA, it is imperative that the school confirm that they make sense. Data that conform to all data dictionary definitions of proper formatting and size may nonetheless be inaccurate. Aggregation of data into a full school-level report can also reveal systematic errors that were not evident to any single staff member during initial entry of individual records (e.g., improbably large numbers of "Native American" demographic codes or attendance rate skewed downward by an errant entry of a "9" instead of a "90"). Errors due to incorrect rounding, or the aggregation of percentages that are themselves rounded figures, can produce seemingly small inaccuracies that may lead to significant errors in AYP calculations.

Follow-up Process. School-level data should not be transmitted to the LEA until the designated school official has signed off on a clean set. Data stewards should be responsible for following up on any questionable data before they are reported upward. The farther data move from their source, the more difficult it is for data validation processes to detect the types of out-of-range errors that are obvious to school personnel who work with the data on a daily basis. This process then repeats itself at the LEA and State levels: responsible data validation staff perform quality checks, return any erroneous or questionable data to the provider, and sign off on a final data set.

<<<<<<<<<<<<< Data Review and Validation Guidelines >>>>>>>>>>>>

General Principles

- *Regularity.* Monitoring during data collection should occur on a regular basis according to a pre-determined schedule. Wherever possible, control processes should be automated to ensure adherence to the schedule.

- *Consistency.* Validation checks should be performed automatically each time data are entered into the system. The data dictionary should define business rules that will consistently identify entries as out of range, missing, incorrectly formatted, or having invalid codes.

- *Interoperability.* Validation rules should be consistent across various databases and systems sharing information. Each time data are transmitted from one system to another (e.g., from the school to the LEA or from the LEA to the State), data should be re-checked.

- *Reliability*. Ongoing monitoring should include periodic review of a sample of data for accuracy and completeness. Wherever possible, reliability reviews should use independent verification processes rather than the regular quality check system (e.g., manual comparisons with other databases).

- *Accuracy.* Data checks should include confirmation that calculations are sound. Rules for rounding should be clear and consistently observed and, until final aggregation at the State level, data should be reported as raw numbers rather than pre-calculated percentages.

- *Feedback Capability.* Ongoing monitoring should include the capability to record and respond to data providers' and requestors' concerns about the collection and reporting system.

- *Flexibility.* Data collection systems should be able to be updated or changed as data quality issues emerge. Dynamic data dictionaries should allow validation rules to be changed as provider and requestor needs evolve, and as data definitions change.

- *Transparency.* Information on all ongoing data quality monitoring procedures should be collected as a staff resource and archived as a continuing reference. Handbooks on management controls should include descriptions of valid data elements from the data dictionary and processes for correcting errors.

- *Documentation.* Data collection systems should include a user-friendly capability to document data quality problems in real time. Users should be able to document intentional deviations from the regular collection processes and business rules

immediately, including known instances of non-responses. All data entry should include coding that identifies the person responsible for the data.

Guidelines for Specific NCLB Data Elements

- *NCLB Demographic Data*
  - Data on subgroup affiliation should be maintained as part of schools' student information system and should be validated and updated at least once a year.
  - Demographic data should be maintained permanently and stored in a single database. New collections of data "from scratch" should not be conducted for each annual NCLB report.
  - Validation rules should flag for further review any subsets of student data (e.g., grade levels, classes, special populations, free or reduced lunch, etc.) whose aggregate demographic profile deviates significantly from the historical or expected result.

- *NCLB Assessment Data*
  - See Section 3.3 for guidelines on controls for student assessment data.

- *NCLB Accountability Data*
  - Past years' assessment records should be restricted to "read-only" status for most staff. Once entered and verified as accurate, updates and changes should not be necessary for these data.
  - Quality checks on all indicators should include automatic flags for values outside a pre-defined range of expected results.

- *NCLB Teacher Quality Data*
  - Data on teacher qualifications, credentials, and course assignments should be maintained as part of schools' and LEAs' personnel databases, and should be validated and updated at least once a year.
  - Teachers should be given "read-only" access to the NCLB-related teacher quality information included in their file for data validation purposes. Only the designated owner of the data (e.g., the data steward) should have the authority to change records.

Technical Focus: Automated Data Validation

What processes can I put into place to validate assessment data?

- ✓ Conduct test runs of the reporting database using preliminary data.
- ✓ Review sample booklets/answer sheets to check accuracy of automated scoring.
- ✓ Conduct diagnostic tests on scanning machines prior to each production run.
- ✓ Create an edit program that checks each response for valid value ranges, consistency with other data values, and acceptable data type.
- ✓ Activate real-time reporting of scanning errors.
- ✓ Limit access to making scoring changes and designate an "official score" that will be final.
- ✓ Constantly update interrater reliability table during scoring process to allow real-time tracking.

(Adapted from National Center for Education Statistics Cooperative Education Data Collection and Reporting Standards Project Task Force, *Standards for Education Data Collection and Reporting,* 3-6, 4-10.)

4.3 Data Privacy and Security Issues

Student Privacy. None of the Federal NCLB Report Card reporting requirements necessitates the identification of individual students, and NCLB has several built-in safeguards (based on the requirements for reporting de-identified information under the Family Educational Rights and Privacy Act) to ensure that disaggregated data used to report achievement results for subgroups cannot be traced back to an identifiable individual. Therefore, it is important to establish a data collection, entry, and reporting system that protects individual students' privacy to the maximum extent possible. For example, unless there is a compelling reason (e.g., State or Federal law) to use them as student identifiers, social security numbers should not be attached to NCLB records. Instead, the crucial step of implementing unique statewide identifiers can be accomplished by using automated number generators or coded algorithms. In addition to the obvious advantages of this approach in protecting the security of students' privacy, using system-generated identifiers rather than social security numbers has a practical, data quality dimension as well. Using social security numbers as identifiers requires a data collection in itself as parents are asked to provide their children's numbers. This consumes precious resources and necessitates a complex, tedious, and costly data checking and validation process to confirm that the nine-digit numbers have indeed been accurately reported for every student. (Note that Federal law requires that parents be informed that social security numbers will be collected, and allows them to refuse to provide the numbers if they so choose.)

Further Information: Data Security and Privacy

Data security and privacy covers an enormous and complex range of issues, both IT and policy-related. The guidelines in this section are meant to be a basic overview of overarching principles. For more in-depth information on protecting the privacy of student records, including issues such as parental permission, readers can contact the U.S. Department of Education's Family Policy Compliance Office (FPCO). FPCO administers the Family Educational Rights and Privacy Act (FERPA), and school officials who need technical assistance or have questions on FERPA can contact FPCO by e-mailing FERPA@ED.gov. For more detailed, formal inquiries, FPCO's mailing address and telephone number are:

Family Policy Compliance Office
Office of Innovation and Improvement
U.S. Department of Education
400 Maryland Ave, SW
Washington, DC 20202-5920
(202) 260-3887

Requests for FERPA compliance training for school officials may be sent to FERPA.Client@ED.gov.

System Safeguards. Maintaining a secure data system requires a combination of technical and human safeguards. On the technical side, it is critical that all hardware, software, and network infrastructure be firewall-secure from unauthorized external access and password-protected to control internal access. Periodic system tests should be run to ensure that technical security protocols remain effective. On the human side, it is important that the data quality team develop specific policies on who will have access to what data and how that access will be controlled. Ideally, data stewards will be owners of all data in their domain, and will be responsible for monitoring and maintaining the security of those data. Staff training for all school-, LEA-, and State-level personnel, including ethical and legal responsibilities for maintaining security, is essential.

<<<<<<<<<<<<<< Data Privacy and Security Guidelines >>>>>>>>>>>>>

General Principles

- *Minimalism.* Records and notes created during the data collection process – whether electronic or paper – should contain only the minimum necessary personally identifiable information.

- *Exclusivity.* Access to data should be strictly limited to personnel with specific responsibility over each data element or a "legitimate educational interest" in viewing it (as defined by State or local policy). Electronic databases should be password and log-in protected, and personally identifiable information should be accessible only when necessary for a specific reporting purpose.

- *Awareness.* Staff training should include building an understanding of Federal, State, and local privacy laws and their application to ongoing data collections. Privacy experts should conduct sessions specifically addressing data security issues, and should be readily accessible during data collection periods to answer questions from the field.

- *Documentation.* Develop a written list of policies and practices related to data security and privacy and ensure that it is disseminated to all personnel involved in data collection, entry, and reporting.

- *Comprehensiveness*. Statewide system-generated identifiers should be created for all individual student records. Using a statewide system will allow tracking students as they move between schools and districts. Social security numbers should not be used as student identifiers.

Technical Focus: System Security

What can I do to improve data security?

The National Forum on Education Statistics recommends the following steps:

- ✓ Document the date and reason for collecting each data element, to ensure appropriate use and access.
- ✓ Identify all record files and data elements within the files as "restricted" or "unrestricted."
- ✓ Develop a filing system for records that minimizes the possibility of misplacing confidential information and enabling unauthorized access.
- ✓ Document all changes and additions to files, including who made them and when they were made.
- ✓ Have systems operators monitor system access through a recordkeeping system that requires passwords to be changed every three months.
- ✓ Include a 10 second warning message before access to restricted records is granted. Message would notify user of restricted status and penalty for unauthorized access, and contain a prompt asking for affirmation that the user understands privacy policies before gaining access.
- ✓ Standardize security protocols, encryption technologies, and digital signatures where data transfer occurs between different systems.
- ✓ Avoid excessive copies of back-up records, and label documents as "original" or "back-up."

(Source: National Forum on Education Statistics. *Forum Guide to Protecting the Privacy of Student Information: State and Local Education Agencies*, NCES 2004–330. Washington, DC: 2004, p. 55. The Forum Guide is available on-line through the National Center of Education Statistics website at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2004330.)

Summary Checklist: Confirming Accurate Results

Does your data validation process have the following characteristics?

- The data quality process includes regular ***ongoing validation*** of new *and* existing data.

- Data systems include ***embedded security safeguards*** throughout all collection, entry, and reporting processes.

- Initial data validation consists of automated quality checks that ensure data are in the ***proper format***.

- Data stewards ***clean aggregated data*** to flag out-of-range errors and confirm reported data "make sense."

- A systematic ***follow-up process*** is in place to correct questionable data before reporting to the next level.

- ***Student privacy*** is enhanced by using system-generated identifiers rather than social security numbers.

- ***System security*** includes a combination of technical and human safeguards involving access to records.

5. Conclusion

Ensuring good data quality is not easy. Needless to say, the practices outlined in these guidelines should not be taken as guarantees of attaining perfect accuracy for all information. Even the best data quality system can – and inevitably will – fall victim to simple human error. Despite the best technical systems, staff training, and data monitoring, "plausible errors," once introduced into a data system, are almost impossible to catch. These guidelines, intended to serve as short-term, easily implemented steps to improve NCLB Report Card data while more systematic State and national efforts move forward, aim to avert the most readily preventable errors.

Ultimately, whatever the limitations, working to improve data quality is well worth the effort. Ensuring that decisions on AYP and funding are based on the most accurate information possible is clearly in the interest of educators, parents, and students at every level. However, there are even more fundamental reasons to invest in a well-functioning, user-friendly system of data management controls. Data-driven decision-making is a reality of sound educational practice that is here to stay. To meet the needs of students in an increasingly competitive national – and global – educational environment, it is essential that teachers and administrators have information that they can trust as they implement instructional programs.

5.1 Roles and Responsibilities

Achieving good data quality is a responsibility that resides at all levels of the educational system: Federal, State, LEA, and school. The question of "which level is responsible for which specific data quality task?" is a difficult one to answer, given the wide range of structures, systems, and policies that exist in each State. However, a few general principles may be useful. The following lists summarize action steps that schools, LEAs, States, and the Federal Government can take to help improve data quality. As with all of the guidelines in this document, these action steps are not presented as additional requirements beyond what is in the law. Instead, they are intended to provide information regarding good practices in data collection at each level.

<<<<<<<<<<<<<<<<<<<<<< School Level >>>>>>>>>>>>>>>>>>>>>>>>

- Assume ownership over all data originating at the school level. This could include:
  - Student assessment data;
  - Student demographic data (including NCLB subgroup information);
  - Graduation rate data – both aggregated and in its component parts, to allow calculation of different variations of the rate; and
  - Highly Qualified Teacher data, including the number of core classes taught by teachers who are not highly qualified.

- Integrate school-level data systems with the district-wide system to enable electronic sharing and promote the use of standard definitions.

- Designate data stewards responsible for school data quality and data entry personnel responsible for inputting new data into the system. Data stewards should visually review all data for face validity and correctness.

- Establish a school-level schedule for data collection and reporting, based on LEA- and State-level deadlines.

- Validate and certify the accuracy of all school-level data before transmittal to the LEA.

- Provide regular feedback to the LEA and State level on data quality issues and concerns that arise during collection and reporting.

<<<<<<<<<<<<<<<<<<<<<<<<< LEA Level >>>>>>>>>>>>>>>>>>>>>>>>>

- Assume ownership over all data originating at the LEA level. This could include:
  - Aggregated LEA student assessment data;
  - LEA graduation rate data – both aggregated and in its component parts, to allow calculation of different variations of the rate;
  - Aggregated LEA demographic data (numbers in each NCLB subgroup);
  - Aggregated Highly Qualified Teacher data, including numbers and percentages of core courses taught by teachers who are not highly qualified; and
  - LEA-level data trends over time.

- Integrate the LEA data system with the statewide data system to enable electronic sharing and promote the use of standard definitions.

- Establish a data quality team, including LEA-level chief information officer.

- Establish an LEA-level schedule for data collection and reporting, based on State-level deadlines.

- Coordinate staff training on data quality across all schools in the LEA.

- Establish a process for error remediation between schools and the LEA.

- Validate and certify the accuracy of all LEA-level data before transmittal to the State.

- Provide regular feedback to the State on data quality issues and concerns that arise.

<<<<<<<<<<<<<<<<<<<<<<<<< State Level >>>>>>>>>>>>>>>>>>>>>>>>>

- Assume ownership over all data originating at the State level. This could include:
  - Aggregated statewide student assessment results;
  - Statewide graduation rate data – both aggregated and in its component parts, to allow calculation of different variations of the rate;

- Aggregated statewide demographic data (numbers in each NCLB subgroup);
- Teacher credentialing/licensing data; and
- Accountability data comparisons among LEAs.

- Establish a data quality office led by a chief information officer responsible for setting statewide data quality policies and technical standards.

- Establish standard statewide definitions and procedures for all reporting elements.

- Develop and disseminate standard statewide forms for reporting data.

- Integrate the State-level data system with the Federal reporting interface.

- Develop a State handbook – including training materials – for implementing data quality processes at the local level.

- Establish a statewide schedule for data collection and reporting.

- Provide continuously available assistance to districts and schools on data quality issues. Ensure that training is given to those responsible for providing NCLB data.

- Hold LEAs and schools accountable for accurate, complete, and timely reporting. Establish a process for error remediation between LEAs and the State.

- Validate and certify the accuracy of all LEA-level data before transmittal to the Federal government.

- Provide regular feedback to the Federal government on data quality issues and concerns.

- Align State assessments to State academic achievement standards and performance standards. Ensure that State assessments are objective and fair.

- Adhere to any requirements that are clearly defined in law or regulation in State data collection activities and reporting.

- Clearly define requirements that are left to the State, and base these definitions on valid, nationally recognized standards.

<<<<<<<<<<<<<<<<<<<<<<<<< Federal Level >>>>>>>>>>>>>>>>>>>>>>>>>

- Continue to set nationwide Consolidated State Performance Report reporting deadlines which are established through *Federal Register* notices.

- Continue to identify required NCLB reporting elements as collected in the Consolidated State Performance Report.

- Continue to develop a common interface for collecting State-level reports through ED*Facts*.

- Establish a process for error remediation between States and the Federal government, through ED*Facts*.

- Continue to develop general nationwide guidelines for improving NCLB data quality.

5.2 Resources for Further Information

As a document outlining practical steps for improving data quality, these guidelines are informed in large part by the recent work of numerous Federal, State, local, private, and non-profit entities. Many organizations are involved in implementing fundamental, systemic long-term changes to the nation's data collection and validation practices, and the overarching principles they have developed are an invaluable resource. In focusing on NCLB Report Card data quality issues, this document leans heavily on non-regulatory guidance documents disseminated by the U.S. Department of Education, the Federal implementing regulations for NCLB, and the *No Child Left Behind* Act of 2001 itself. Listed below is an extensive set of resources that can be tapped for further information on both data quality and NCLB.

<<<<<<<<<<<<<<<<<<<< Data Quality Resources >>>>>>>>>>>>>>>>>>>>

Federal

Government Accountability Office. *Improvements Needed in Education's Process for Tracking States' Implementation of Key Provisions*, Report #GAO-04-734, September 2004. http://www.gao.gov/new.items/d04734.pdf

National Center for Education Statistics. *NCES Handbooks Online* (access to NCES data dictionary). http://nces.ed.gov/programs/handbook/index.asp

National Center for Education Statistics Cooperative Education Data Collection and Reporting Standards Project Task Force. *Standards for Education Data Collection and Reporting*, December 1991. http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=92022

U.S. Department of Education. *U.S. Department of Education Information Quality Guidelines*, February 2003. http://www.ed.gov/policy/gen/guid/infoqualguide.html

U.S. Department of Education, Office of the Inspector General. Final Audit Report, *Review of Management Controls Over Scoring of the National Assessment of Educational Progress 2000,* Report #ED-OIG/A05-C0010, June 2003.

U.S. Department of Education, Office of the Inspector General. Management Information Report, *Best Practices for Management Controls Over Scoring of the State Assessments Required Under the No Child Left Behind Act of 2001*, State and Local Report #04-01, February 3, 2004.

U.S. Department of Education, Office of Vocational and Adult Education. *Peer Evaluation Resource Guide*, prepared for Data Quality Initiative Meetings, May, 2001.
U.S. Office of Management and Budget. *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies*, February 22, 2002. http://www.whitehouse.gov/omb/fedreg/reproducible2.pdf

State and Local

Council of Chief State School Officers. *Quality Control Checklist for Processing, Scoring, and Reporting.* Technical Issues in Large-Scale Assessment – A State Collaborative on Assessment and Student Standards, January 2003. http://www.ccsso.org/content/pdfs/scorereportQCchklst.pdf

Council of Chief State School Officers. *Data Quality and Standards Project* [website]. http://www.ccsso.org/projects/Data_Quality_and_Standards_Project/

Fairfax County Public Schools. *Education Decision Support Library Website*. http://www.fcps.edu/DIT/edsl/

New Hampshire Department of Education. *Initiative for School Empowerment and Excellence* [i.4.see Website]. http://www.ed.state.nh.us/education/datacollection/i4see.htm

Private and Non-Profit

Accountability Works. *Model Contractor Standards and State Responsibilities for State Testing Programs*, Education Leaders Council. educationleaders.org/elc/events/model%20contractor.pdf

Ligon, Glynn D. *White Paper: A Technology Framework to Support Accountability and Assessment: How States Can Evaluate Their Status for No Child Left Behind*. ESP Solutions Group, 2004. nclbtechsummits.org/summit1/GlynnLigonWhitePaper.pdf

National Center for Educational Accountability. *9 Essential Elements of Statewide Data-Collection Systems.* Austin, TX, 2004. http://www.nc4ea.org/files/9%20elements%20Brochure.pdf

National Forum on Education Statistics. *Forum Guide to Protecting the Privacy of Student Information: State and Local Education Agencies*, NCES 2004–330. Washington, DC: 2004. http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2004330

National Forum on Education Statistics. *Forum Guide to Building a Culture of Quality Data: A School & District* Resource (NFES 2005–801). U.S. Department of Education. Washington, DC: National Center for Education Statistics, 2004. http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2005801

<<<<<<<<<<<<< *No Child Left Behind* Resources (Federal) >>>>>>>>>>>>>

The full text of the *No Child Left Behind Act of 2001* (PL 107-110) can be accessed through the U.S. Department of Education's www.ed.gov website at http://www.ed.gov/policy/elsec/leg/esea02/index.html.

Implementing regulations for *No Child Left Behind* were published in the *Federal Register* by the U.S. Department of Education on December 2, 2002. The full text of the regulations can be found at http://a257.g.akamaitech.net/7/257/2422/14mar20010800/edocket.access.gpo.gov/2002/pdf/02-30294.pdf

The U.S. Department of Education has issued a number of non-regulatory guidance documents on NCLB implementation. A full list of guidance documents, including links to those documents, is at http://www.ed.gov/policy/elsec/guid/edpicks.jhtml?src=fp

References for selected non-regulatory guidance documents relating to data collection requirements are below. All of the following documents were prepared by the U.S. Department of Education:

Report Cards: Title I, Part A Non-Regulatory Guidance, September 12, 2003. http://www.ed.gov/programs/titleiparta/reportcardsguidance.doc

Standards and Assessments Non-Regulatory Guidance, March 10, 2003. http://www.ed.gov/policy/elsec/guid/saaguidance03.doc

LEA and School Improvement Non-Regulatory Guidance, January 7, 2004. http://www.ed.gov/policy/elsec/guid/schoolimprovementguid.doc

Improving Teacher Quality State Grants, Title II, Part A Non-Regulatory Guidance, Revised January 16, 2004. http://www.ed.gov/programs/teacherqual/guidance.doc

Title I Paraprofessionals Non-Regulatory Guidance, March 1, 2004. http://www.ed.gov/policy/elsec/guid/paraguidance.doc

Final Non-Regulatory Guidance on the Title III State Formula Grant Program – Standards, Assessments and Accountability, February 2003. http://www.ed.gov/programs/nfdp/NRG1.2.25.03.doc

APPENDIX: SUMMARY CHECKLISTS

For users' convenience, the summary checklists from each of the three main sections of the guidelines are reprinted together over the next three pages.

Establishing a Solid Foundation

Does your data infrastructure have the following characteristics?

- Data collection, processing, and reporting systems are ***automated*** and data can be transmitted in an electronic, interoperable format.
- Immediate ***interim processes*** for improving data quality are in place, as larger systemic initiatives are implemented.
- A ***data dictionary*** identifies all data elements used in collection and reporting, and describes their content and format.
- Systematic ***business rules*** define acceptable values, character formats, and options for handling missing or unavailable data.
- Hardware and software, along with staff training, are configured around standard ***data definitions*** and business rules.
- ***Data granularity*** is preserved by collecting data in their component forms and computing percentages, ratios, and other calculations at the State level.
- A ***data quality team*** has been established at the school and LEA levels to craft and implement data quality goals and procedures.
- Data collection, reporting, and review encourages ***school-level ownership*** of basic information, and staff training reflects this responsibility.
- LEA- and State-level personnel play leading roles in building a ***culture of data quality*** among staff at all levels.

Managing Consistent Collection

Does your data collection process have the following characteristics?

- Information collection for most data elements is achieved through ***mining existing data***.
- Layout, labeling, and instructions for data forms and systems contribute to ***clear, straightforward collection instruments***.
- Assessment instruments ***avoid direct student entry*** of information available from other sources.
- States and LEAs ***work with vendors*** to ensure that assessments and other data collection instruments align with definitions and standards.
- The State has established ***firm, clear schedules*** and deadlines for collecting, validating, and reporting required data elements.
- Data quality validation is a ***continuous, inclusive process*** that updates all elements of the data system on a regular basis and takes into account both policy *and* technical considerations.
- ***Statewide deadlines*** spread collection and reporting responsibilities over the full school year.

Confirming Accurate Results

Does your data validation process have the following characteristics?

- The data quality process includes regular ***ongoing validation*** of new *and* existing data.
- Data systems include ***embedded security safeguards*** throughout all collection, entry, and reporting processes.
- Initial data validation consists of automated quality checks that ensure data are in the ***proper format***.
- Data stewards ***clean aggregated data*** to flag out-of-range errors and confirm reported data "make sense."
- A systematic ***follow-up process*** is in place to correct questionable data before reporting to the next level.
- ***Student privacy*** is enhanced by using system-generated identifiers rather than social security numbers.
- ***System security*** includes a combination of technical and human safeguards involving access to records.